Simultaneous Gross Error Detection and Data Reconciliation Using Gaussian Mixture Distribution

by

Hashem Alighardashi

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
PROCESS CONTROL

Department of Chemical and Materials Engineering

University of Alberta

Abstract

The intensive competitive nature of the world market, the growing significance of quality products, and the increasing importance and the number of safety and environmental issues and regulations, respectively, have increased the need for fast and low-cost changes in chemical processes to enhance their performance. Any possible changes and modifications in a system in order to control, optimize, evaluate the behavior of the process, or achieve the maximal performance of the system require clear understanding and knowledge of its actual state. This information is obtained by processing a data set - collecting it, ameliorating its accuracy, and storing/using it for further analysis. It should be emphasized that in today's highly competitive world market, increasing the accuracy of measurements by resolving even small errors can result in substantial improvements in plant efficiency and economy.

Industrial process measurements play a significant role in online optimization, process monitoring, identification, and control. These measurements are used to make decisions which potentially influence product quality, plant safety, and profitability. Nonetheless, they are inherently contaminated by errors, which may be random and/or systematic/gross errors, due to sensor accuracy, improper instrumentation, poor calibration, process leak, and so on. The objective of data reconciliation and gross error detection is the estimation of the true states and the detection of any faults in the instruments which could seriously degrade the performance of the system. Data reconciliation techniques deal with the problem of improving the accuracy of raw process measurements and their application allows optimal adjustment of measurement values to satisfy material and energy constraints. These methods

also make possible estimation of the unmeasured variables. However, data reconciliation approaches do not always provide valid estimates of the actual states, and the presence of gross errors in the measurements significantly affect the accuracy levels that can be accomplished using reconciliation. Therefore, the main focus of this work is to develop a framework to obtain the accurate estimates of reconciled values while reducing the impact of gross errors.

In reality, operating conditions under which a process works change with different circumstances. Therefore, it is vital to develop a model that is capable of identifying and switching between operating regions. To this end, a method is proposed for simultaneous gross error detection and rectification of a data set which contains different operating regions. First, the data set is divided into several clusters based on the number of operating regions. Then, the same operation, i.e., data rectification is performed on each operating region. It must be noted that all of the proposed approaches in this thesis do not require to preset the parameters of the error distribution model, rather they are determined as part of the solution. They are also applicable to problems with both linear and nonlinear constraints, in addition to the ability to determine the magnitude of gross errors. Furthermore, these methods/approaches detect partial gross errors, so it is not required to assume that gross errors exist in the entire data set. Finally, the performance of the proposed methods is verified through various simulation studies and realistic examples.

Preface

All of the work presented henceforth is an original work by Hashem Alighardashi in the Computer Process Control (CPC) group at University of Alberta. The materials presented in the thesis are part of the research project under the supervision of Professor Biao Huang and are funded by Alberta Innovates Technology Futures (AITF) and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

A version of Chapter 2 of the thesis is submitted to Industrial and Engineering Chemistry research journal, *I&EC research*, ACS publications. I was the lead investigator, responsible for all major areas of concept formation, formulation, data analysis, as well as manuscript composition. Magbool Jan N was involved in the early stages of concept formation and contributed to manuscript edits.

To my lovely parents, family, and friends for their support and encouragement.

Hashem Alighardashi

August 2017

Acknowledgments

I wish to express my gratitude to Professor Biao Huang, my supervisor, who provided me with the opportunity to work and research in his group and broaden my knowledge and professional skills. I am deeply grateful for his countless support, guidance, and encouragement throughout my Master's program at the University of Alberta. I am thankful to him for his supportive attitude and inspiration during my studies. It was a great pleasure for me to explore my graduate studies under his supervision.

It is my honor to be a member of the Computer Process Control (CPC) group where we broaden our knowledge in various research aspects and enjoy participating in discussions. This work would not have been possible without the help of several past and present member of CPC group. I would like to thank my colleagues Mohammad Rashedi, Shabnam Sedghi, Nabil Magbool Jan, and many others for their help and support.

I would like to acknowledge the Department of Chemical and Materials Engineering, University of Alberta, for giving me the opportunity to pursue my Master's study in a pleasant environment. I would like to thankfully acknowledge the Alberta Innovates Technology Futures (AITF) and the Natural Sciences and Engineering Research Council of Canada for their financial support.

Last but not least, I want to express my love and deepest gratitude to my dearest who were always there no matter what: my mother for dreaming big for me, my father for believing in me, and my brother and sister for motivating me. Without you and your support, none of this would be possible. I would like to extend my gratitude onto my friends as well, for having my back and always being there in different stages of my life.

Respectfully,

Hashem Alighardashi

Contents

List of Tables

List of Figures

List of Abbreviations

EM	*Expectation Maximization*
E-step	*Expectation Step*
M-step	*Maximization Step*
MLE	*Maximum Likelihood Estimation*
MAP	*Maximum A Posteriori*
OP	*Overall Power*
AVTI	*Aerage Number of Type I Error*
CR	*Correct Rate*

Chapter 1

Introduction

1.1 Motivation

Raw measurements which are collected from industrial plant operations contain substantial information and play a significant role in process identification, process control, and online optimization. However, these measurements have to be processed prior to further analysis, since they are often corrupted by errors, including random errors and possibly gross errors, which may affect the subsequent applications, i.e., biased results for estimation. The presence of gross errors in the measurements affects the reliability of optimization and control solutions. Therefore, in this work, in order to remove both kinds of measurement errors, we characterize the measurement noise model using a Gaussian mixture distribution with two modes, where each mixture component denotes the error distribution corresponding to random error and gross error, respectively. Based on this assumption, a maximum likelihood framework for simultaneous steady state data reconciliation and gross error detection is proposed. Therefore, the problem of data rectification is solved based on principles of probability, and all the estimated parameters and rectified values are obtained based on the principle of the highest probability. The identification of normal noise versus gross error is based on the highest probability, similar to the likelihood ratio tests. Since the proposed framework involves noise mode as a hidden variable denoting the existence of gross errors in the data, it should be solved using the Expectation Maximization (EM) algorithm. However, maximum likelihood estimation does not extend coverage to the actual states where there are "equivalent sets of gross errors", and prior information has to be incorporated to handle these special cases. Therefore, in order to cover all different situations in data rectification, the Maximum a Posteriori (MAP) framework is proposed. Nonetheless, in reality, operating conditions under which a process works change with different circumstances. Therefore, it is necessary to develop a model which is capable of identifying and switching between operating regions. Accordingly, a method is proposed for simultaneous gross error detection and rectification of a data set which contains different operating regions. To this end, the data set must be divided into several clusters based on the number of operating regions. Then, the same operation, i.e., data rectification has to be performed on each operating region. It must be noted that all proposed approaches in this thesis do not require to preset the parameters of the error distribution model, rather they are determined as part of the solution. The approaches presented here are also applicable to problems with both linear and nonlinear constraints, in addition to the ability to provide the magnitude of gross errors. Furthermore, they detect partial gross errors, so it is not required to assume that gross errors exist in the entire data set. Finally, the performance of the proposed methods is verified through various simulation studies and realistic examples.

1.2 Thesis Outline

The rest of the thesis is organized as follows:

In Chapter 2, an expectation maximization approach is developed in order to eliminate both random and gross errors. This approach is based on the Maximum Likelihood Estimation (MLE), and Gaussian mixture distribution with two modes is used to describe the noise model. In this approach, the parameters of the error distribution model are determined as part of the solution. In addition, problems with both linear and nonlinear constraints are covered in the proposed approach. Finally, the performance of the method is demonstrated through several simulation case studies and realistic examples.

Since the proposed method in Chapter 2 is based on the MLE framework, it cannot handle different gross errors which have the same effect on the objective function. Therefore, in Chapter 3, prior information is considered in the objective function as well. Incorporating the prior information provides not only more accurate estimates of the true states, but also allows some other possible situations of gross error sets in measurements to be rectified. It is worth to mention that if there is a gross error in a measurement it is not necessary to assume all of the sample points of that variable are contaminated by the gross error, i.e., it could partially occur in a data set. Several simulation studies and examples are provided to evaluate the effectiveness of the proposed method.

Both Chapter 2 and Chapter 3 deal with data rectification for cases with only a single operating region. In Chapter 4, the proposed method is extended such that it is capable of identifying different operating regions, estimating the distribution parameters of each region, and then returning the rectified estimates of each region. Finally, different case studies of an example are presented to illustrate the efficiency of the proposed method.

Conclusions and recommendations for future work are provided in Chapter 5.

1.3 Main Contributions

The main contributions of the thesis can be summarized as follows:

1. A new method is proposed for simultaneous gross error detection and data reconciliation using a contaminated Gaussian distribution based on Expectation Maximization (EM) algorithm.

2. The method is further extended by incorporating the prior information in order to obtain the more accurate estimate of the true states and to be able to rectify various

kinds of raw measurements.

3. The proposed robust data rectification approach is developed when there exist multiple operating regions in a data set.

1.4 Submitted Publications

Materials of the thesis have been submitted in the following publications:

1. H.Alighardashi, N. Magbool Jan, B. Huang, "Expectation Maximization Approach for Simultaneous Gross Error Detection and Data Reconciliation Using Gaussian Mixture Distribution", *I&EC Journal*. (Chapter 2)

Chapter 2

Expectation Maximization Approach for Simultaneous Gross Error Detection and Data Reconciliation Using Gaussian Mixtrue Distribution

Process measurements play a significant role in process identification, control, and optimization. However, they are often corrupted by two types of errors, random and gross errors. The presence of gross errors in the measurements affects the reliability of optimization and control solutions. Therefore, in this work, we characterize the measurement noise model using a Gaussian mixture distribution, where each mixture component denotes the error distribution corresponding to random error and gross error, respectively. Based on this assumption, we propose a maximum likelihood framework for simultaneous steady state data reconciliation and gross error detection. Since the proposed framework involves noise mode as a hidden variable denoting the existence of gross errors in the data, it can be solved using the Expectation Maximization (EM) algorithm. This approach does not require the parameters of the error distribution model to be preset, rather they are determined as part of the solution. Several case studies are presented to demonstrate the effectiveness of the proposed approach.

2.1 Introduction

Raw measurements collected from processes are of vital importance for online optimization, process monitoring, and control. These measurements are used to make decisions which might influence product quality, plant safety, and profitability. Nonetheless, they are contaminated by measurement errors - random errors and possibly gross errors - due to sensor inaccuracy, improper instrumentation, poor calibration, and so on[1,2]. The accuracy of process measurements is often improved using data reconciliation techniques. However, the presence of gross errors in the measurements significantly affect the improvement in accuracy that could be accomplished using reconciliation[3]. Therefore, the focus of this work is to develop a framework for obtaining the accurate estimates of reconciled values while reducing the impact of gross errors.

There exist three main approaches to handle gross errors in the measurements: sequential gross error detection and data reconciliation, simultaneous gross error detection and data reconciliation, and robust data reconciliation that provides reconciled estimates by ignoring gross errors without actually identifying them. In the first approach, gross error detection is performed using statistical tests; then the measurement containing a gross error is either eliminated or compensated with the average value and finally, data reconciliation is performed[2]. In general, the data has to be processed for the detection and identification of gross errors that have to be eliminated prior to performing data reconciliation. Various statistical tests or methodologies have been proposed to detect gross errors – the global test[4,5], nodal test[5–7], measurement test[5,8,9], generalized likelihood ratio test[10], the Bonferroni test[11], and principal component analysis[12]. However, it is desirable to identify not only the presence of gross errors, but also the location and magnitude of these errors[13]. One of the approaches that can eliminate gross errors is the serial elimination strategy[14,15]. In this strategy, each measurement is tested one by one using the above mentioned statistical tests, and finally, possible gross errors are eliminated. Alternatively, a serial compensation strategy can estimate the magnitude of gross errors in which gross errors are estimated and the measurements are compensated in turn[10]. However, the results of this kind of gross error elimination rely on the accuracy of the estimated magnitude of gross errors[11]. After employing one of the gross error detection and elimination strategies discussed above, a conventional steady state data reconciliation technique which minimizes the weighted least square criterion satisfying the steady state model of the process is performed[16]. It should be noted that the conventional reconciliation problem assumes there is no gross error present in the data and hence, gross error detection should be performed prior to data reconciliation.

In the second approach, gross error detection and data reconciliation are performed simultaneously. In this regard, Soderstrom et al.[17] proposed a mixed integer linear programming problem to estimate the magnitude of the gross errors and simultaneously obtain the reconciled estimates. On the other hand, Arora and Biegler[18] presented a mixed-integer nonlinear programming formulation that can systematically identify the magnitude of the faulty sensor in the redescending estimator framework and consider bias magnitude as decision variables in the optimization problem. Both of the mentioned methods employed branch and bound schemes to solve the optimization problem and hence, are not suitable for large systems. Alternately, Yuan et al.[19] proposed a novel hierarchical Bayesian framework, in which the problem is divided into three layers, and in each layer one or two parameters are obtained using Bayes rule.

In the third approach, the ideas from robust statistics are used to obtain the reconciled estimates by minimizing the influence of gross errors whenever the magnitude of the residuals exceed the threshold values. This approach involves defining robust objective functions for data reconciliation depending on the assumption of noise distribution. In this regard, Tjoa and Biegler[20] utilized the contaminated Gaussian to characterize the noise model such that each mode of the contaminated Gaussian signifies the occurrence of random errors and gross errors. A hybrid successive quadratic programming (SQP) method was proposed to solve the resulting non-convex data reconciliation problem. Following a similar noise model and using the prior distribution of states, a maximum likelihood rectification problem was formulated[21]. On the other hand, Alhaj-Dibo et al.[1] presented an analytical form for the reconciled estimates in terms of the weighting matrix. The weight matrix is shown to be the function of residuals and hence, an iterative solution algorithm was proposed. Different choice of noise distribution would result in different objective functions for robust data reconciliation. For a comparative study of different objective functions, the reader is referred to the works of Prata et al.[22] and Özyurt[23]. It is important to note that these objective functions require the parameters of the noise model to be tuned out in order to successfully mitigate the influence of gross errors[18,20].

In this chapter, we follow the work of Tjoa and Biegler[20] to characterize the measurement noise distribution as a Gaussian mixture distribution with two sensor modes, where each mode represents the occurrence of random errors and gross errors, respectively. The main contributions of this work are three-fold: (1) an Expectation-Maximization (EM) algorithm is proposed to solve the simultaneous gross error detection and data reconciliation problem; (2) there are no tuning parameters in our proposed approach in comparison to the work of Tjoa and Biegler[20] and Alhaj-Dibo et al.[1], and (3) this approach can be directly used to handle multiple gross errors in measurements, and can be easily extended to the partially measured case, and also for nonlinear systems.

The rest of the chapter is organized as follows. Section 2.2 contains the background and a general introduction to data reconciliation problem. In section 2.3, the definition of data reconciliation is presented using the Gaussian mixture distribution as a measurement noise model. In section 2.4, an Expectation Maximization algorithm is presented such that the parameters are estimated along with the reconciled estimates. In section 2.5, several examples are presented to evaluate the performance of the proposed approach and finally, conclusions are presented in section 2.6.

2.2 Data Reconciliation

Data reconciliation is a technique to improve the accuracy of raw measurements using the mathematical model of the process. There are three main components in a reconciliation problem: process model, measurement model, and assumptions on measurement noise distribution. Obtaining the accurate estimates of reconciled values relies heavily on the knowledge of these components. Let $\mathbf{y} \in \mathbb{R}^n$ be denoted as a vector of raw measurements, $\mathbf{x} \in \mathbb{R}^n$ as true values of the process variables, and $\epsilon \in \mathbb{R}^n$ as the measurement errors. Assuming that the sensor noise is additive, the measurement model is given by,

$$y_i = x_i + \epsilon_i \tag{2.1}$$

where y_i and x_i denote the measured and true value of the i^{th} variable, respectively.

Given the process model of the form, $\mathbf{f}(\mathbf{x}) = 0$, and assuming the measurement error ϵ_i to be independent of x_i, the general formulation of a steady state data reconciliation problem can be expressed as[2,21]:

$$\arg\max_{\mathbf{x}} \quad P\{\mathbf{y}|\mathbf{x}\} \tag{2.2}$$

$$\text{s.t.} \quad \mathbf{f}(\mathbf{x}) = 0 \tag{2.3}$$

$$\mathbf{g}(\mathbf{x}) \leq 0 \tag{2.4}$$

where the objective function signifies the probability of measurements, and $\mathbf{g}(\mathbf{x})$ denotes the inequality constraints signifying the safety limits or variable bounds. Due to the presence of errors, the raw measurements do not satisfy the process model and hence, solving the above optimization problem determines the reconciled values by maximizing the probability of measurements.

For linear processes, assuming a zero-mean Gaussian distribution of measurement errors and no inequality constraints, the reconciliation problem can be stated as a weighted least squares problem:

$$\min_{\mathbf{x}} \quad (\mathbf{y}-\mathbf{x})^T V^{-1} (\mathbf{y}-\mathbf{x}) \tag{2.5}$$

$$\text{s.t.} \quad A\mathbf{x} = 0 \tag{2.6}$$

where A is the model coefficient matrix, and V is the covariance matrix of measurements and it is often assumed to be a diagonal matrix. The reconciled estimates are given by the following closed-form solution[2]:

$$\hat{\mathbf{x}} = [I - VA^T(AVA^T)^{-1}A]\mathbf{y} \tag{2.7}$$

The above data reconciliation problem, Equation (2.5), is based on the assumption that measurements contain only random errors. Therefore, the presence of gross errors invalidates the statistical basis of the reconciliation problem and hence, results in a smearing effect[2]. In other words, in the presence of gross errors, the conventional approach adjusts the measurements such that they follow the process model constraints but the adjustments might result in erroneous reconciled values. In order to account for the presence of gross errors in the measurements, it is often assumed that ϵ_i follows a non-zero mean Gaussian distribution (i.e., $\epsilon_i \sim N(b_i, \sigma_i^2)$) for the measurement model given by Equation (2.1). Equivalently, ϵ_i can be characterized using the zero-mean Gaussian distribution with the measurement model given by $y_i = x_i + b_i + \epsilon_i$. Using these noise characteristics and the form of the measurement model, the simultaneous gross error detection and data reconciliation can be formulated as a mixed-integer optimization problem[17,18]. Recently, Yuan et al.[19] proposed a hierarchical Bayesian framework that iteratively solves for reconciled values, gross error detection, and hyperparameters of the noise model.

Alternately, the concepts from robust statistics can be utilized to define robust objective functions, $\rho(r_i)$, which are expressed as functions of standardized residuals, $r_i = (y_i - x_i)/\sigma_i$, and the corresponding robust data reconciliation problem is stated as[1,24,25]:

$$\min_{\mathbf{x}} \quad \sum_{i=1}^{n} \rho(r_i) \tag{2.8}$$

$$\text{s.t.} \quad A\mathbf{x} = 0 \tag{2.9}$$

where ρ can be any monotone function of standardized residuals such that the effect of gross errors is reduced on estimation of reconciled values. The robust objective function ρ can be derived from the maximum likelihood functions for contaminated Gaussian, Cauchy and Logistic distributions. Further, the Fair function, Lorentzian function and the redescending M-estimator can also be used with the proper tuning parameters to eliminate the influence of gross errors and perform robust data reconciliation[23].

2.3 Problem Statement

Let Y denotes the data matrix of raw measurements as follows:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix} = \begin{bmatrix} \mathbf{y_1^v} \\ \mathbf{y_2^v} \\ \vdots \\ \mathbf{y_n^v} \end{bmatrix} = \begin{bmatrix} \mathbf{y_1^s} & \mathbf{y_2^s} & \cdots & \mathbf{y_m^s} \end{bmatrix} \tag{2.10}$$

where superscripts s and v refer to sample point and variable, respectively. The i^{th} row of the Y matrix represents the variable i measured at different sampling instants, and the j^{th} column represents the set of n variables measured at j^{th} sampling instant. Hence, the measurement model for all variables is given by,

$$\mathbf{y_j^s} = \mathbf{x} + \epsilon \tag{2.11}$$

where $\mathbf{x}$ is defined as,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{2.12}$$

It should be noted that the measurement model given by Equation (2.11) is the vector form of the measurement model presented in Equation (2.1) at the j^{th} time instant.

In this work, following the works of Tjoa and Biegler[20] and Alhaj-Dibo et al.[1], a Gaussian mixture distribution with two modes is assumed for the measurement error vector, ϵ, to account for the presence of gross errors. The sensor mode with small variance corresponds to random noise, and the other abnormal one with large variance corresponds to gross errors. Therefore, the measurement error distribution can be written for the i^{th} variable as,

$$\epsilon_i \sim \delta_{i1} N(0, \sigma_{i1}^2) + \delta_{i2} N(0, \sigma_{i2}^2) \tag{2.13}$$

such that $\delta_{i1} + \delta_{i2} = 1$. Representing gross errors with the probability density function with zero mean might appear as a contradiction to the definition of gross error. However, the main idea is that by assigning a large variance for the second mode of the error distribution, $\pm 3\sigma_{i2}$ could cover the magnitude of the gross error. Therefore, if there is an error between $-3\sigma_{i1}$ and $-3\sigma_{i2}$ or $+3\sigma_{i1}$ and $+3\sigma_{i2}$, it would be considered as a gross error. While the gross errors can be described by various probability distributions such as Gaussian location mixture or a uniform distribution, representing gross errors by zero-mean Gaussian distribution is a mathematically more tractable solution and has also been used in the literature[1]. Since the two modes are mutually exclusive, the sensor model for a single measurement set can be expressed as,

$$P\{\mathbf{y}|\mathbf{x}\} = P\{\mathbf{y_j^s}|\mathbf{x}\} = \sum_{k=1}^{2} P\{\mathbf{y_j^s}|\mathbf{x}, I_{ij} = k\} P\{I_{ij} = k\} \tag{2.14}$$

where I_{ij} denotes the hidden sensor mode of variable i at the j^{th} sample point. This can be either in the normal mode or in the abnormal mode. Using the above sensor model, we can define the log-likelihood of m measurements set as:

$$\arg\max_{\mathbf{x}} \quad lnP\{Y|\mathbf{x}\} = ln(\sum_{k=1}^{2} P\{Y|\mathbf{x}, I_{ij} = k\}P\{I_{ij} = k\}) \tag{2.15}$$

Assuming the measurements are independent, the above equation becomes,

$$\arg\max_{\mathbf{x}} \quad lnP\{Y|\mathbf{x}\} = \prod_{i=1}^{n}\prod_{j=1}^{m} ln(\sum_{k=1}^{2} P\{y_{ij}|x_i, I_{ij} = k\}P\{I_{ij} = k\}) \tag{2.16}$$

where $p_{ijk} = P\{y_{ij}|x_i, I_{ij} = k\}$ is defined as follows,

$$p_{ijk}\{y_{ij}|x_i, \sigma_{ik}\} = \frac{1}{\sqrt{2\pi}\sigma_{ik}} exp(-\frac{(y_{ij} - x_i)^2}{2\sigma_{ik}^2}) \quad k = 1, 2 \tag{2.17}$$

and $P\{I_{ij} = k\} = \delta_{ik}$ denotes the probability of occurrence of sensor mode k. In particular, $P\{I_{ij} = 1\} = \delta_{i1}$ is the probability of measurements in the normal mode where only random error is present, and $P\{I_{ij} = 2\} = \delta_{i2}$ denotes the probability of measurements in the abnormal gross error mode for variable i. Prior works that utilize the Gaussian mixture model to characterize the noise distribution use the prespecified values of noise distribution parameters $\theta_{ik} = \{\delta_{ik}, \sigma_{ik}^2\}$ for both the modes[1,20,23]. However, presetting the noise distribution parameters influences the performance of gross error detection and hence, affects the accuracy of reconciled estimates. Therefore, in this work, we aim at determining the noise distribution parameters while simultaneously obtaining the reconciled estimates. Now, the simultaneous data reconciliation and gross error detection problem for linear steady state processes can be posed as follows:

$$\max_{\mathbf{x},\theta_{ik}} \quad lnP\{Y|\mathbf{x}\} = \prod_{i=1}^{n}\prod_{j=1}^{m} ln(\sum_{k=1}^{2} p_{ijk}\{y_{ij}|x_i,\theta_{ik},I_{ij}=k\}P\{I_{ij}=k|\theta_{ik}\}) \tag{2.18}$$

$$s.t. \quad A\mathbf{x} = \mathbf{0} \tag{2.19}$$

From Equation (2.18), it can be observed that the model identity probability cannot be obtained explicitly, and hence, the maximum likelihood estimation problem cannot be solved directly. To resolve this issue, the Expectation Maximization (EM) algorithm is applied to obtain the distribution parameter values for each variable. In other words, all of the model distribution parameters along with the reconciled values are estimated simultaneously using the EM algorithm.

2.4 Solution Methodology

In this section, an Expectation Maximization approach will be proposed to solve the simultaneous gross error detection and data reconciliation problem in the maximum likelihood framework. In general, the EM algorithm is used to solve a maximum likelihood problem with hidden variables[26]. In the following subsections, the EM algorithm is presented for a general problem of maximum likelihood estimation with hidden variables. Next, we derive the relevant update expressions for the data reconciliation problem presented in section 2.3.

2.4.1 General Formulation of the EM Algorithm

Suppose that the complete data is $\{C_{obs}, C_{mis}\}$, where C_{obs} refers to observed data and C_{mis} denotes hidden variables. The problem of maximum likelihood estimation of parameter Θ, when there are hidden variables, can be mathematically stated as:

$$\hat{\Theta} = arg\ max_{\Theta}\ P(C_{obs}, C_{mis}|\Theta) \tag{2.20}$$

For such problems, the EM algorithm is often used. It is an iterative two step algorithm and the general procedure of the algorithm is as follows:

1. In the first step, known as the Expectation or E-step, the conditional expectation of the hidden variables (Q function) is obtained by replacing $\{C_{mis}\}$ with its conditional expectation given $\{C_{obs}\}$ using the current fit for Θ, as,

$$\begin{aligned} Q(\Theta, \Theta^{(l)}) &= E_{C_{mis}|C_{obs},\Theta^{(l)}}[ln\ P(C_{mis}, C_{obs}|\Theta)] \\ &= \int P(C_{mis}|C_{obs}, \Theta^{(l)})[ln\ P(C_{mis}, C_{obs}|\Theta)]dC_{mis} \end{aligned} \tag{2.21}$$

2. In the second step, known as the Maximization or M-step, the $Q(\Theta, \Theta^{(l)})$ function is maximized over the parameter space with respect to Θ,

$$\Theta^{(l+1)} = arg\ max_{\Theta}\ Q(\Theta, \Theta^{(l)}) \tag{2.22}$$

The Q function is re-evaluated with the updated parameter values $\Theta^{(l+1)}$, and the procedure is repeated until convergence.

2.4.2 EM Algorithm for Simultaneous Gross Error Detection and Data Reconciliation

For the simultaneous gross error detection and data reconciliation problem presented in section 2.3, the observed data is $C_{obs} = Y$ and the hidden variable (C_{mis}) is the model identity $I_{ij} = \{1, 2\}$. The reconciled variables and the noise distribution parameters are collected in $\Theta = \{x_i, \theta_{ik}\}$.

2.4.2.1 E-step

As mentioned previously, in the E-step, the Q function is derived:

$$\begin{aligned} Q(\Theta,\Theta^{(l)}) &= E_{I|Y,\Theta^{(l)}}[lnP(Y,I|\Theta)] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} E_{I_{ij}|y_{ij},\theta_i^{(l)}}[ln(P(y_{ij},I_{ij}|\theta_i))] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2} P(I_{ij}=k|y_{ij},\theta_{ik}^{(l)})[ln(P(y_{ij},I_{ij}=k|\theta_{ik}))] \end{aligned} \tag{2.23}$$

In Equation (2.23), $P(I_{ij}=k|y_{ij},\theta_{ik}^{(l)})$ is the posterior distribution of the model identity and it can be defined using Bayes theorem as,

$$\begin{aligned} P(I_{ij}=k|y_{ij},\theta_{ik}^{(l)}) &= \frac{P(y_{ij}|I_{ij}=k,\theta_{ik}^{(l)})P(I_{ij}=k|\theta_{ik}^{(l)})}{P(y_{ij}|\theta_{ik}^{(l)})} \\ &= \frac{P(y_{ij}|I_{ij}=k,\sigma_{ik}^{2(l)},x_i^{(l)})P(I_{ij}=k|\theta_{ik}^{(l)})}{\sum_{t=1}^{2} P(y_{ij}|I_{ij}=t,\theta_{ik}^{(l)})P(I_{ij}=t|\theta_{it}^{(l)})} \\ &= \frac{\frac{1}{\sqrt{2\pi\sigma_{ik}^{2(l)}}}exp(\frac{-(y_{ij}-x_i^{(l)})^2}{2\sigma_{ij}^{2(l)}})P(I_{ij}=k|\theta_{ik}^{(l)})}{\sum_{t=1}^{2}\frac{1}{\sqrt{2\pi\sigma_{it}^{2(l)}}}exp(\frac{-(y_{ij}-x_i^{(l)})^2}{2\sigma_{ij}^{2(l)}})P(I_{ij}=t|\theta_{it}^{(l)})} \end{aligned} \tag{2.24}$$

In the above equation, $P(I_{ij}=k|\theta_{ik}^{(l)})$ is the prior distribution of the model identity which is equal to the weight of k^{th} component, i.e., $P(I_{ij}=k|\theta_{ik}^{(l)})=\delta_{ik}^{(l)}$. So, Equation (2.24) can be further simplified as,

$$P(I_{ij}=k|y_{ij},\theta_{ik}^{(l)}) = \frac{\frac{\delta_{ik}^{(l)}}{\sqrt{2\pi\sigma_{ik}^{2(l)}}}exp(\frac{-(y_{ij}-x_i^{(l)})^2}{2\sigma_{ik}^{2(l)}})}{\sum_{t=1}^{2}\frac{\delta_{it}^{(l)}}{\sqrt{2\pi\sigma_{it}^{2(l)}}}exp(\frac{-(y_{ij}-x_i^{(l)})^2}{2\sigma_{it}^{2(l)}})} \tag{2.25}$$

Since all parameters in Equation (2.25) are obtained from the previous iteration, it can be simplified as $P(I_{ij}=k|y_{ij},\theta_{ik}^{(l)})=\gamma_{ijk}^{(l)}$. Further, the log-likelihood term in the Q function is written as,

$$
\begin{aligned}
lnP(y_{ij}, I_{ij} = k|\theta_{ik}) &= ln(\frac{1}{\sqrt{2\pi\sigma_{ik}^2}} exp(\frac{-(y_{ij} - x_i)^2}{2\sigma_{ik}^2}) P(I_{ij} = k|\theta_{ik})) \\
&= ln(\delta_{ik}) - ln(\sqrt{2\pi\sigma_{ik}^2}) - \frac{(y_{ij} - x_i)^2}{2\sigma_{ik}^2}
\end{aligned} \tag{2.26}
$$

Substituting Equations (2.25) and (2.26) into Equation (2.23), the following expression for the Q function can be obtained:

$$
\begin{aligned}
Q(\Theta, \Theta^{(l)}) &= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} P(I_{ij} = k|y_{ij}, \theta_{ik}^{(l)})[lnP(y_{ij}, I_{ij} = k|\theta_{ik})] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{ijk}^{(l)} [ln(\delta_{ik}) - ln(\sqrt{2\pi\sigma_{ik}^2}) - \frac{(y_{ij} - x_i)^2}{2\sigma_{ik}^2}]
\end{aligned} \tag{2.27}
$$

2.4.2.2 M-step

The M-step involves maximizing the $Q(\Theta, \Theta^{(l)})$ function with respect to Θ and this yields the update equations for the parameters.

Update expression for σ_{rt}: This can be obtained by setting the partial derivative of Q with respect to σ_{rt} to zero. The partial derivative is obtained as follows:

$$
\frac{\partial Q(\Theta, \Theta^{(l)})}{\partial \sigma_{rt}} = \frac{\partial \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{ijk}^{(l)} [ln(\delta_{ik}) - ln(\sqrt{2\pi\sigma_{ik}^2}) - \frac{(y_{ij} - x_i)^2}{2\sigma_{ik}^2}]}{\partial \sigma_{rt}} \tag{2.28}
$$

for $r = 1, ..., n$ and $t = 1, 2$. The derivative can be further simplified as,

$$
\begin{aligned}
\frac{\partial Q(\Theta, \Theta^{(l)})}{\partial \sigma_{rt}} &= \frac{\partial \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ijt}^{(l)} [ln(\delta_{it}) - ln(\sqrt{2\pi\sigma_{it}^2}) - \frac{(y_{ij} - x_i)^2}{2\sigma_{it}^2}]}{\partial \sigma_{rt}} \\
&= \frac{1}{\sigma_{rt}} (\frac{\sum_{j=1}^{m} \gamma_{rjt}^{(l)} (y_{rj} - x_r)^2}{\sigma_{rt}^2} - \sum_{j=1}^{m} \gamma_{rjt}^{(l)}) = 0
\end{aligned} \tag{2.29}
$$

Now the update equation for σ_{rt} is given by,

$$\sigma_{rt}^{(l+1)} = \sqrt{\frac{\sum_{j=1}^{m} \gamma_{rjt}^{(l)} (y_{rj} - x_r^{(l)})^2}{\sum_{j=1}^{m} \gamma_{rjt}^{(l)}}} \tag{2.30}$$

Update expression for δ_{rt}: This can be derived by formulating the following constrained optimization problem:

$$\begin{cases} \delta_{ik}^{(l+1)} = \arg\max_{\delta_{rt}} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{ijk}^{(l)} [ln(\delta_{ik}) - ln(\sqrt{2\pi\sigma_{ik}^{(l+1)^2}}) - \frac{(y_{ij}-x_i^{(l)})^2}{2\sigma_{ik}^{(l+1)^2}}] \\ s.t. \quad \sum_{k=1}^{2} \delta_{ik}^{(l+1)} = 1 \end{cases} \tag{2.31}$$

The above optimization problem can be solved using the method of *Lagrange multipliers* and the corresponding Lagrangian function is defined as,

$$L(\delta_i, \lambda_{\delta i}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{ijk}^{(l)} [ln(\delta_{ik}) - ln(\sqrt{2\pi\sigma_{ik}^{(l+1)^2}}) - \frac{(y_{ij} - x_i^{(l)})^2}{2\sigma_{ik}^{(l+1)^2}}] - \lambda_{\delta i}(\sum_{k=1}^{2} \delta_{ik} - 1) \tag{2.32}$$

In order to solve the above equation, the partial derivative of L with respect to δ_{rt} and $\lambda_{\delta r}$ is determined and then set to zero.

$$\begin{aligned} \frac{\partial L(\delta_i, \lambda_{\delta i})}{\partial \delta_{rt}} &= \frac{\partial \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{ijk}^{(l)} [ln(\delta_{ik}) - ln(\sqrt{2\pi\sigma_{ik}^{(l+1)^2}}) - \frac{(y_{ij}-x_i^{(l)})^2}{2\sigma_{ik}^{(l+1)^2}}]}{\partial \delta_{rt}} - \lambda_{\delta r} \\ \delta_{rt} &= \frac{\sum_{j=1}^{m} \gamma_{rjt}^{(l)}}{\lambda_{\delta r}} \end{aligned} \tag{2.33}$$

It is known that $\sum_{k=1}^{2} \delta_{rk} = 1$; using the property of γ_{ijk} function where $\sum_{k=1}^{2} \gamma_{ijk}^{(l)} = 1$, we get $\lambda_{\delta r} = m$. As a result, $\delta_{rt}^{(l+1)}$ can be expressed as,

$$\delta_{rt}^{(l+1)} = \frac{\sum_{j=1}^{m} \gamma_{ijt}^{(l)}}{\lambda_{\delta r}} = \frac{\sum_{j=1}^{m} \gamma_{ijt}^{(l)}}{m} \tag{2.34}$$

Update expression for x_i: Similarly, the update expression for the reconciled estimate $x_i^{(l+1)}$ can be obtained using the method of *Lagrange multipliers*,

$$\begin{cases} x_i^{(l+1)} = \arg\max_{x_r} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{ijk}^{(l)} [ln(\delta_{ik}^{(l+1)}) - ln(\sqrt{2\pi\sigma_{ik}^{(l+1)^2}}) - \frac{(y_{ij}-x_i)^2}{2\sigma_{ik}^{(l+1)^2}}] \\ s.t. \qquad A\mathbf{x} = 0 \end{cases} \tag{2.35}$$

The Lagrangian function for the above constrained optimization problem is given by,

$$L(\mathbf{x}, \lambda_x) = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{ijk}^{(l)} [ln(\delta_{ik}^{(l+1)}) - ln(\sqrt{2\pi\sigma_{ik}^{(l+1)^2}}) - \frac{(y_{ij}-x_i)^2}{2\sigma_{ik}^{(l+1)^2}}] + \sum_{i=1}^{n} \sum_{z=1}^{q} \lambda_{xz} a_{zi} x_i \tag{2.36}$$

To obtain the update expression for the reconciled estimates, the partial derivatives of the above equation with respect to x_r and λ_{xw} have to be evaluated.

$$\begin{aligned} \frac{\partial L(\mathbf{x}, \lambda_x)}{\partial x_r} &= \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{rjk}^{(l)} [\frac{y_{rj} - x_r}{\sigma_{rk}^{(l+1)^2}}] + \sum_{z=1}^{q} \lambda_{xz} a_{zr} \\ &= \sum_{j=1}^{m} ([\frac{\gamma_{rj1}^{(l)}}{\sigma_{r1}^{(l+1)^2}} + \frac{\gamma_{rj2}^{(l)}}{\sigma_{r2}^{(l+1)^2}}](y_{rj} - x_r)) + \sum_{z=1}^{q} \lambda_{xz} a_{zr} = 0 \end{aligned} \tag{2.37}$$

$$\frac{\partial L(\mathbf{x}, \lambda_x)}{\partial \lambda_{xw}} = \sum_{i=1}^{n} a_{wi} x_i = 0 \tag{2.38}$$

Define $w_{rj}^{(l+1)}$ as,

$$w_{rj}^{(l+1)} = \frac{\gamma_{rj1}^{(l)}}{\sigma_{r1}^{(l+1)^2}} + \frac{\gamma_{rj2}^{(l)}}{\sigma_{r2}^{(l+1)^2}} \tag{2.39}$$

Here, $w_{rj}^{(l+1)}$ corresponds to the weights for each of the measured variables. The above equation is then written in matrix form such that the resulting expression for reconciled estimates can be compared to ones from literature. For this purpose, we define a diagonal matrix $W_j^{(l+1)} = diag\{w_{1j}^{(l+1)}, \cdots, w_{nj}^{(l+1)}\}$, where $diag_{(j=1,\ldots,m)}(a_j)$ denotes the operator that creates the diagonal matrix having the dimension of measured variables (i.e., n) with diagonal elements $w_{rj}^{(l+1)}$. Equation (2.37) can be then simplified as,

$$\sum_{j=1}^{m} W_j^{(l+1)}(\mathbf{y}_j^s - \mathbf{x}) + A^T\lambda_x = 0 \tag{2.40}$$

multiplying by $A\Big(\sum_{j=1}^{m} W_j^{(l+1)}\Big)^{-1}$ yields,

$$\lambda_x^{(l+1)} = -\Big(A\Big(\sum_{j=1}^{m} W_j^{(l+1)}\Big)^{-1}A^T\Big)^{-1}A\Big(\sum_{j=1}^{m} W_j^{(l+1)}\Big)^{-1}\Big(\sum_{j=1}^{m} W_j^{(l+1)}\mathbf{y}_\mathbf{j}^\mathbf{s}\Big) \tag{2.41}$$

define $R^{(l+1)} = (\sum_{j=1}^{m} W_j^{(l+1)})^{-1}$ and $\mathbf{y_w} = \Big(\sum_{j=1}^{m} W_j^{(l+1)}\Big)^{-1}\Big(\sum_{j=1}^{m} W_j^{(l+1)}\mathbf{y}_\mathbf{j}^\mathbf{s}\Big)$, then λ_x is given by,

$$\lambda_x^{(l+1)} = -(AR^{(l+1)}A^T)^{-1}A\mathbf{y_w} \tag{2.42}$$

and hence, the update equation for reconciled values is given by,

$$\mathbf{x}^{(l+1)} = (I - R^{(l+1)}A^T(AR^{(l+1)}A^T)^{-1}A)\mathbf{y_w} \tag{2.43}$$

The resulting expression in Equation (2.43) for reconciled estimates resembles the well-known expression for data reconciliation as given in Equation (2.7). However, the weighting matrix V in the conventional data reconciliation problem is a known constant denoting the diagonal covariance matrix of Normal distribution of errors, whereas the weighting matrix in our formulation uses the standardized residual of m measurement points and the current estimate of noise distribution parameters. In addition, it can be observed that the weighted values of m measurements, denoted by $\mathbf{y_w}$, is used in the proposed formulation. This shows that the measurements are appropriately weighed to obtain more accurate values of the reconciled estimates.

Table 2.1: EM Algorithm for Simultaneous Data Reconciliation and Gross Error Detection

1	Input the raw measurements Y
2	Initialize the parameter Θ^l
3	**E-step** Evaluate the Q function, Equation (2.27), using current updated values of parameters
4	**M-step** Update the parameters σ_{rt}, δ_{rt} and $\mathbf{x}$ using Equations (2.30), (2.34), and (2.43), respectively.
5	Terminate on convergence. Otherwise, proceed to Step 3.

2.4.3 Data Reconciliation for a Single Measurement

In the previous subsection, we presented the solution methodology that determines the error distribution model parameters which can be used in obtaining the robust reconciled values. The solution algorithm in Table 2.1 uses m measurement samples to estimate the parameters of the noise distribution as well as the reconciled estimates. This shows that the data reconciliation can be performed in the moving horizon fashion with window size of m. On the other hand, the presented algorithm can be used to obtain historical noise parameters of each of the variables $\{\delta_{ik}, \sigma_{ik}\}$ for $i = 1, \cdots, n$, and these parameter values can be used to perform data reconciliation for any set of current measurement values, i.e., single measurement set. Therefore, now we present the application of the EM algorithm to a single set of measurements. Using the noise distribution parameters obtained from the EM algorithm, the simplified iterative algorithm for the case of single measurements involves three steps: First, for the current measurement, the probability density values of all measured variables and probabilities of the individual sensor modes, p_{ik}, are evaluated; next, the weights are computed using the values of p_{ik} and the parameter values of δ_{ik} estimated from earlier iteration of the EM algorithm. Finally, the reconciled values are obtained using the newly calculated weights. The corresponding update expressions used at each step is given by,

$$p_{ik}^{(l)} = \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} exp\left(\frac{-(y_i - x_i^{(l)})^2}{2\sigma_{ik}^2}\right) \tag{2.44}$$

$$(R^{(l)})^{-1} = diag\left(\frac{\frac{\delta_{i1}}{\sigma_{i1}^2}p_{i1}^{(l)} + \frac{\delta_{i2}}{\sigma_{i2}^2}p_{i2}^{(l)}}{\delta_{i1}p_{i1}^{(l)} + \delta_{i2}p_{i2}^{(l)}}\right) \tag{2.45}$$

$$\mathbf{x}^{(l+1)} = (I - R^{(l)}A^T(AR^{(l)}A^T)^{-1}A)\mathbf{y} \tag{2.46}$$

This procedure is repeated until convergence. The algorithm presented above for the single measurement set closely resembles the one presented by Alhaj-Dibo et al.[1]. However, in their work, δ_{ik} were assumed to be the same for all the measurements and the variance of the second mode was b times the variance of random errors; further details can be found in the original article.

2.4.4 Further Extensions

Partially measured case: In this subsection, we present suitable modifications to the proposed EM approach such that unmeasured variables can be estimated from the linear process models in the case of partial measurements. For this purpose, the process model constraint is divided into two parts - measured and unmeasured. Next, reconciled values of measured data are estimated, then unmeasured values are computed using the reconciled values.

$$A_y\hat{y} + A_z\hat{z} = 0 \tag{2.47}$$

A matrix P is determined such that $PA_z = 0$, and hence, the second term in Equation (2.47) can be eliminated. Using the following Theorem 1, the process model constraint modifies to,

$$PA_y\hat{y} = 0 \tag{2.48}$$

To simplify the mathematical representation, PA_y is substituted by A which is the process model coefficients matrix,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{q1} & a_{q2} & \cdots & a_{qn} \end{bmatrix} \tag{2.49}$$

Theorem 1[27] If $A_z(m \times n)$, where $m \geq n$, has rank of n, then there is a $Q(m \times m)$ matrix with orthonormal column vector such that $A_z = QR$, where $Q^TQ = I$, and $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$. $R_1(n \times n)$ is an upper triangular and non-singular matrix. 0 is a zero matrix with the dimension $((m-n) \times n)$.

$$A_z = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \times \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \tag{2.50}$$

Q_1 and Q_2 are $(m \times n)$ and $(m \times (m-n))$, respectively.

For proof, the reader is referred to the original article.

After performing data reconciliation and obtaining reconciled values, the unmeasured variables can be estimated using,

$$A_z\hat{z} = -A_y\hat{y} \tag{2.51}$$

$$\hat{z} = -(A_z^TA_z)^{-1}A_z^T(A_y\hat{y}) \tag{2.52}$$

Nonlinear Case: The proposed algorithm can be applied to the case with nonlinear model constraints without linearization since it decreases the accuracy of the results[21]. In order to solve the problem without linearization, the explicit update equation for x in the M-step can be replaced with an implicit nonlinear optimization problem which can be solved efficiently using commercially available optimization solvers. In this work, we use an inbuilt MAT-

LAB function “fmincon” with interior point algorithm to solve the nonlinear programming problem; further details of this algorithm can be found in Dantzig et al. book[28].

2.4.5 Performance Measures

In this work, we use the following performance measures to assess the efficiency of the proposed methodology: (a) Overall power (OP)[10,17], (b) Average number of type I error (AVTI)[10,17], and (c) Correct rate (CR)[19]. They are defined as follows:

$$OP = \frac{Number\ of\ biased\ variables\ correctly\ identified}{Number\ of\ biased\ variables\ simulated} \tag{2.53}$$

$$AVTI = \frac{Number\ of\ unbiased\ variables\ wrongly\ identified}{Number\ of\ simulation\ trials} \tag{2.54}$$

$$CR = \frac{Number\ of\ runs\ where\ all\ the\ gross\ errors\ are\ correctly\ identified}{Total\ number\ of\ runs} \tag{2.55}$$

2.5 Illustrations

In this section, we present several examples to study the effectiveness and performance of the proposed method for simultaneous data reconciliation and gross error detection. In this work, both the linear and nonlinear cases are investigated; in the linear case, the occurrence of single and multiple gross errors as well as partially measured variables and partially occurrence of gross errors are examined, and for the nonlinear part, the presence of multiple gross errors is studied. Finally, a realistic network with both linear and nonlinear constraints while there are some unmeasured variables is presented.

2.5.1 Illustration 1- Linear Example

Let us consider a water flow network of Yuan et al.[19], which has four process units and seven streams as shown in Figure 2.1.

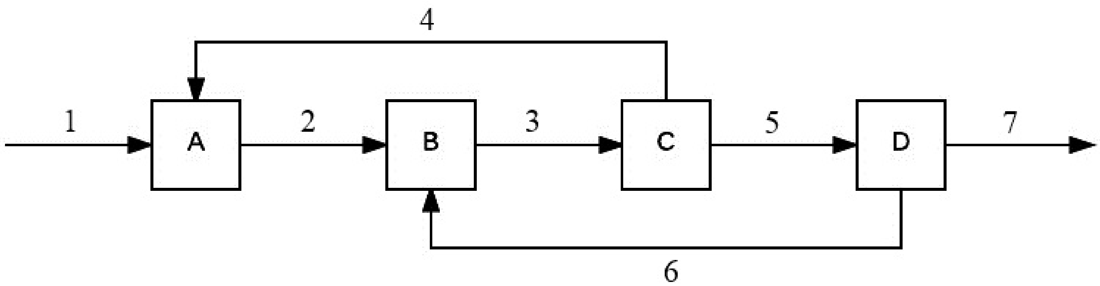

Figure 2.1: Schematic of water flow network

Let x_i denote the flow rate of stream i. For the process network presented in Figure 2.1, the process model constraints (in this case, mass balance equations) are given by,

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

Now expressing the above set of equations in the form of $A\mathbf{x} = 0$, the process model coefficient matrix, A, is given by,

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

All the variables are assumed to be measured unless otherwise stated. The vector of true values of the flow rates is set as $\mathbf{x} = \begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 1 & 1 \end{bmatrix}^T$.

Case 1 − *Single Gross Error* : The measurement data of all the process variables are generated by adding a random noise with zero mean and $0.1I$ variance to the true values. In other words, the standard deviations are around $11 - 32\%$ of the measurements. In this case, we assume x_1 contains gross error with a magnitude of +2. A set of simulated data are plotted in Figure 2.2.

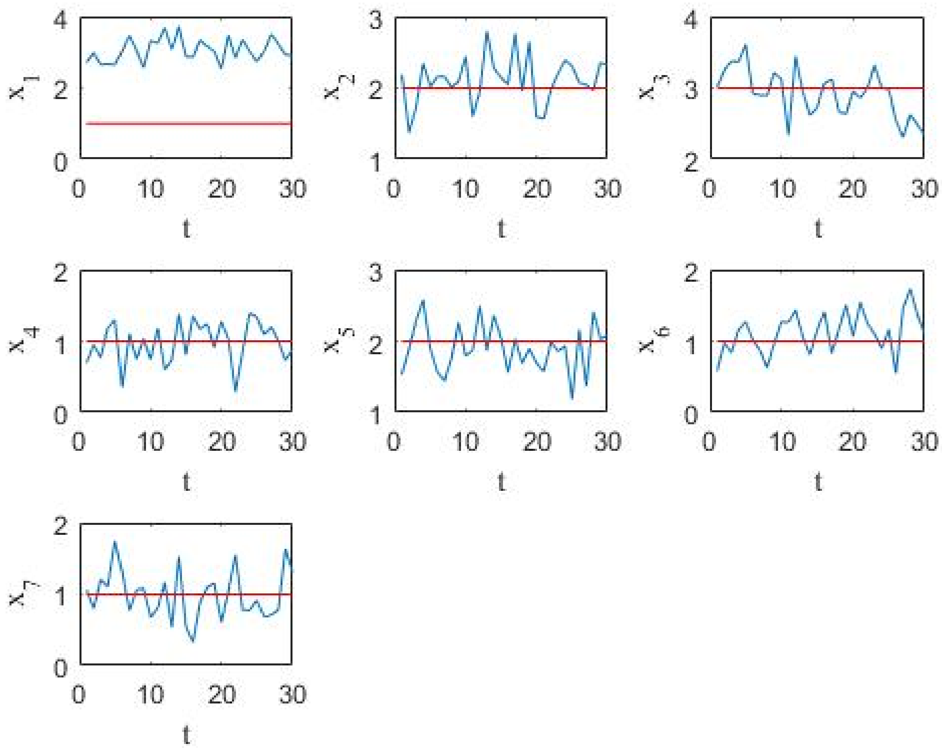

Figure 2.2: Data plot showing gross error in x_1, $\sigma_i = 0.1$

Table 2.2 compares the performance measures obtained using our proposed approach with the results of Algorithm 1 and 2 of Yuan et al.[19]. Only the CR information has been reported in their work when they added a gross error to one of the variables (x_1 to x_7 by sequence) for each 50 runs. The average CR obtained using the proposed EM approach is 1 which means the variable which has gross error has been identified correctly in all the simulation runs.

Table 2.2: Performance measures for single gross error

Method	Simulation runs	Biased	OP	AVTI	Average CR
Proposed Method	50	1	1	0	1
Algorithm 1[19]	50	1	**	**	0.8
Algorithm 2[19]	50	1	**	**	0.98

** denotes not reported

Table 2.3 represents the true values, reconciled values, and the estimates of probability of occurrence of gross errors, δ_{i1}, as well as the estimated variance for the modes when x_1 has gross error. It can be observed from Figure 2.3 and Table 2.3, that the proposed method eliminated bias appropriately (1.9826 which is so close to 2), and resulted in reconciled values that are close to true values. Figure 2.3 shows the histogram of reconciled estimates for 50 runs.

Table 2.3: Reconciliation solution using EM approach for single gross error case

Variables	x	$\hat{x}$	$\hat{\delta}_{i1}$	$\hat{\sigma}_{i1}$	$\hat{\sigma}_{i2}$
x_1	1	1.0174	0.1000	1.9697	2.0237
x_2	2	2.0203	0.9000	0.3069	0.0786
x_3	3	3.0001	0.8667	0.2971	0.0389
x_4	1	1.0029	0.8333	0.3006	0.0590
x_5	2	1.9972	0.9667	0.3545	0.0142
x_6	1	0.9798	0.9333	0.3423	0.0132
x_7	1	1.0174	0.9667	0.2828	0.0855

As mentioned before, σ_{i2} can handle the effect of gross errors. In Case 1, the magnitude of gross error is 2 for the first variable, x_1, and the proposed algorithm returned 2.0237 as the standard deviation of the abnormal mode. In this method, instead of adding another parameter for the mean value and estimating that as a parameter, the gross error effect is captured by a large variance value. Therefore, by applying this method, not only is the effect of gross errors eliminated properly, but the mathematical complexity is also reduced owing to the use of fewer parameters compared to the case where the mean value is considered. However, in this special case, the simulated measurement data does not follow a Gaussian mixture distribution as random noise with zero mean and $0.1I$ variance is being added to all of the variables, as considered in Yuan et al.[19]. Nonetheless, the proposed method can also handle this case properly as shown in Table 2.3. It is important to note that a low probability of occurrence of the first mode for x_1, and low probability of occurrence of the second mode for the other have been determined using the proposed methodology. This signifies that the resulting distribution is not a contaminated Gaussian distribution, rather a single Gaussian distribution. In order to show the performance of the proposed method for Gaussian mixture case, *Case* 4 is presented by simulating the measurement data following a Gaussian mixture distribution.

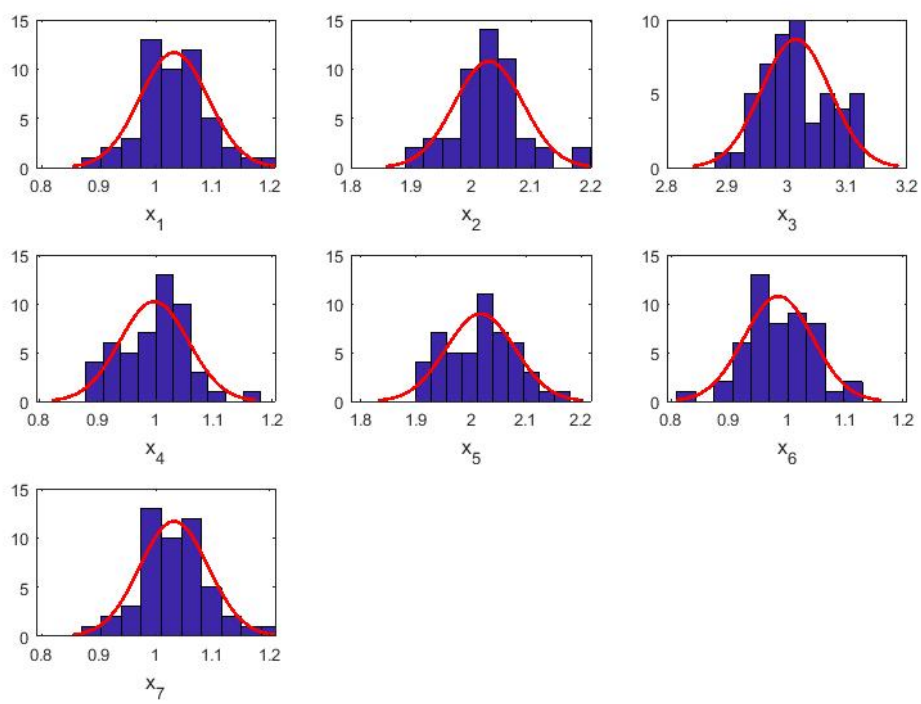

Figure 2.3: Histogram of reconciled values for 50 runs, when x_1 has a bias magnitude of 2

Case 2− *Multiple Gross Errors* : In this case, multiple gross errors in measurements are examined. The simulated data in this case follow the same noise characteristics as in *Case* 1, but now we consider gross errors in x_2 and x_7 with magnitudes of 3 and 1, respectively.

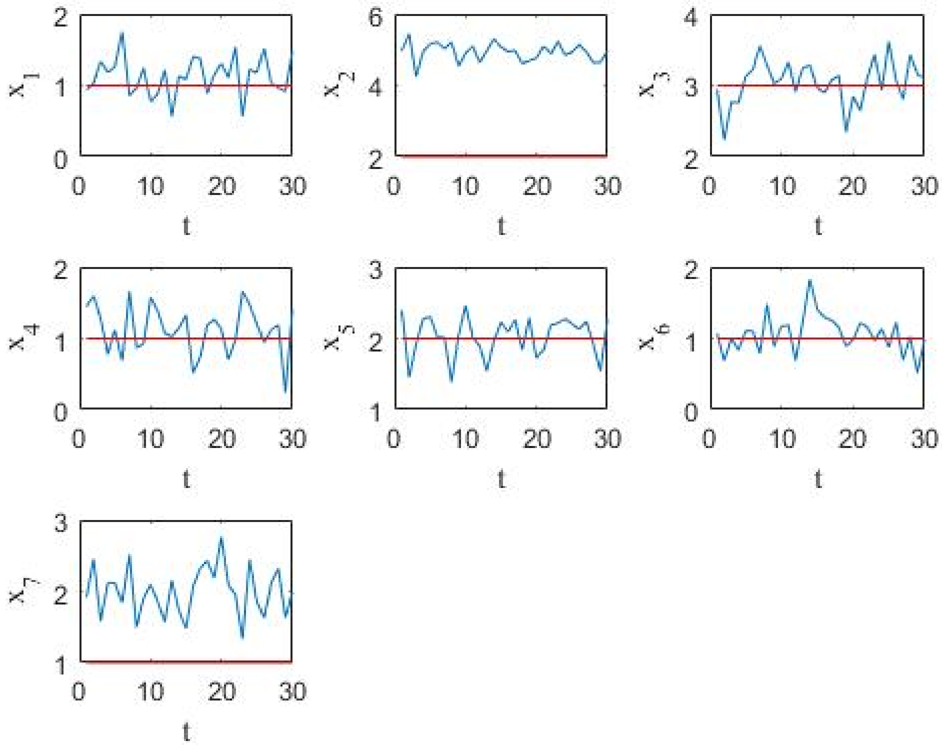

Figure 2.4: Data showing with gross errors in x_2 and x_7

In Table 2.4, the performance measures OP, AVTI, and CR obtained for two simultaneous gross errors are presented.

Table 2.4: Performance measures using EM algorithm for multiple gross errors

Simulation runs	Biased	OP	AVTI	CR
50	x_2, x_7	0.98	0.04	0.98

Table 2.5 provides true and reconciled values, estimated values for the probability of occurrence of the random noise, and estimated variance of each mode of all variables. It can be noted that the variables with gross errors contain high variance. This is because of our characterization of measurement noise distribution. However, all other variables have a low variance, signifying accurate values for those measurements. From Figure 2.5 and Table 2.5, it can be inferred that the proposed method eliminated the bias terms efficiently, and the difference between the reconciled and the true values is close to zero. Figure 2.5 shows the

histogram of reconciled values obtained using our proposed EM approach for 50 simulations.

Table 2.5: Reconciliation solution using EM approach for multiple gross errors case

Variables	x	$\hat{x}$	$\hat{\delta}_{i1}$	$\hat{\sigma}_{i1}$	$\hat{\sigma}_{i2}$
x_1	1	1.0521	0.9740	0.3100	0.3968
x_2	2	2.0806	0.0333	2.5160	3.1681
x_3	3	3.0351	0.8667	0.3205	0.5269
x_4	1	1.0285	0.9333	0.3095	0.5736
x_5	2	2.0066	0.9000	0.3180	0.2952
x_6	1	0.9545	0.9667	0.2717	0.5048
x_7	1	1.0521	0.0333	0.9855	1.0063

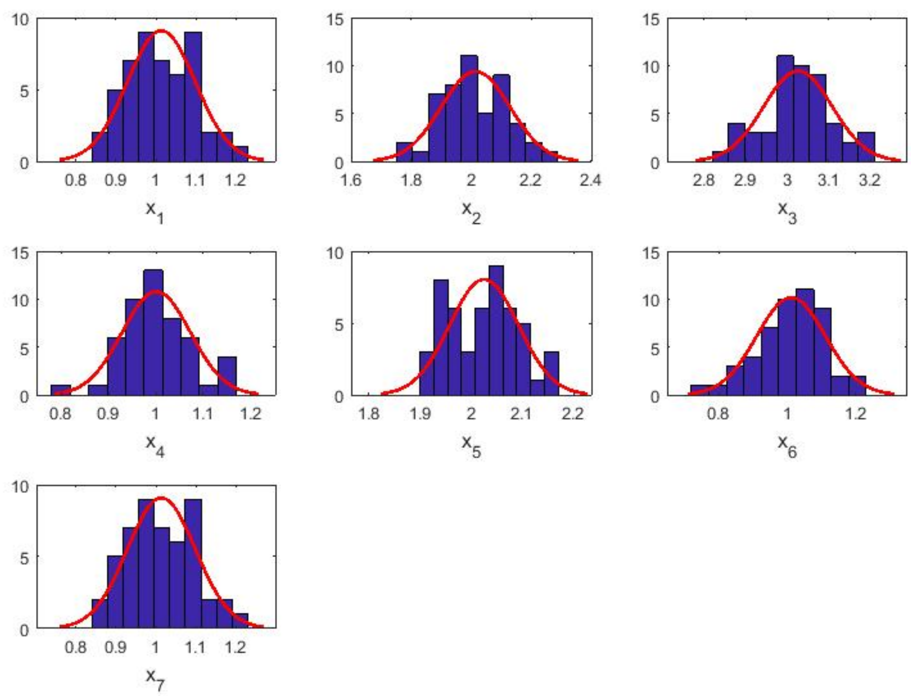

Figure 2.5: Histogram of reconciled values for 50 runs, when x_2 and x_7 contain gross errors

Case 3$-$ *Partially Measured with a Gross Error* : The objective of this case is to show the performance of the proposed method when there are both measured and unmeasured variables. In this case, x_6 is unmeasured and all other process variables are measured. Also, x_2 contains gross errors with a bias magnitude of 2 at all sample points. The measurement data is generated similar to the previous two cases for all other variables. The process model constraint is given by $A_y\hat{y} + A_z\hat{z} = 0$, in which A_y and A_z are as follows,

$$A_y = \begin{bmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, A_z = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}$$

The P matrix in $PA_y\hat{y} = 0$ is,

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{2}/2 & 0 & \sqrt{2}/2 \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

A set of simulated data is presented in Figure 2.6.

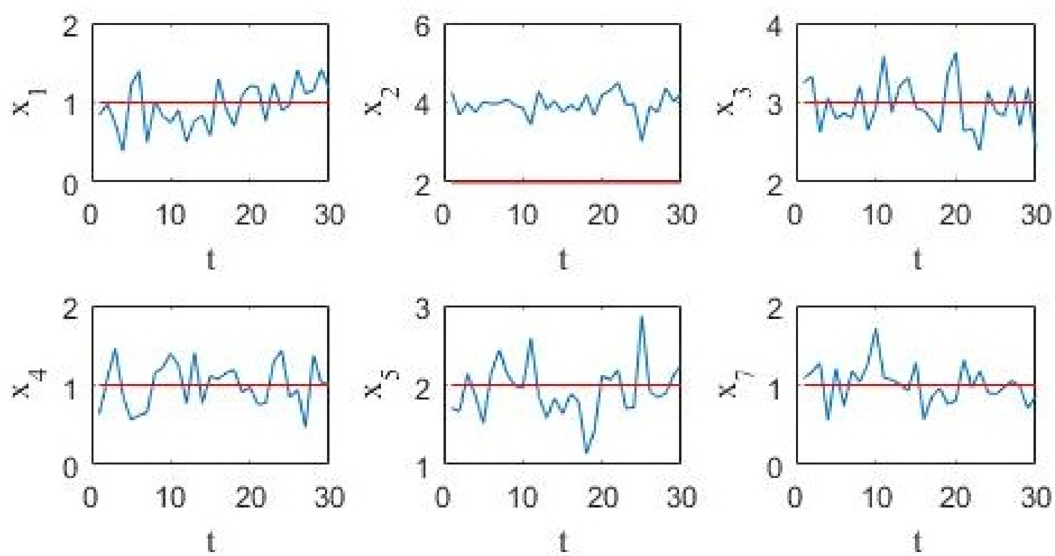

Figure 2.6: Data plot showing gross error in x_2

In Table 2.9 the results of OP, AVTI, and CR tests for *Case* 3 are presented. The results show that the gross error has been eliminated in all the simulation runs.

Table 2.6: Performance measures for partially measured case with single gross error

Simulation runs	Biased	OP	AVTI	CR
50	x_2	1	0	1

Table 2.7 presents true and reconciled values, estimated values for the probability of occurrence of random noise and estimated variance of each mode of all variables. From Figure 2.7 and Table 2.7, it can be seen that the proposed method eliminated the bias terms in an efficient manner, and the difference between the reconciled and the true values is close to zero. Figure 2.5 shows the histogram of rectified values for 50 runs, when there is gross error in x_2.

Table 2.7: Reconciliation solution using EM approach for partially measured case with gross error in x_2

Variables	x	$\hat{x}$	$\hat{\delta}_{i1}$	$\hat{\sigma}_{i1}$	$\hat{\sigma}_{i2}$
x_1	1	1.0336	0.9858	0.3058	0.6204
x_2	2	1.9975	0.1168	1.6135	2.0453
x_3	3	2.9822	0.8132	0.2672	0.4918
x_4	1	0.9639	0.8974	0.2915	0.5288
x_5	2	2.0183	0.9792	0.2952	0.7453
x_7	1	1.0336	0.9620	0.3091	0.4944

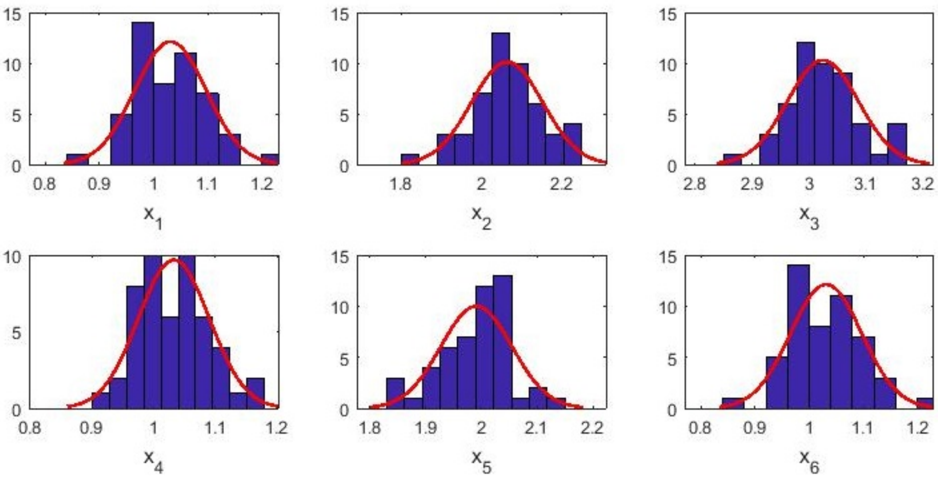

Figure 2.7: Histogram of reconciled values for 50 runs, when x_2 includes gross error and x_6 is unmeasured

The estimate of unmeasured variable x_6 can be obtained using Equation (2.52) and the value is determined to be 0.9847 which is close enough to its true value. As a result, the proposed method can successfully determine the values of unmeasured variables in partially measured case.

Case 4 − Mixture Gaussian Distribution with Multiple Gross Errors : The three previous cases assumed that if a particular variable contains gross error, then the gross error is persistent through all sample points. However, it is possible that gross error exists only for a short period of time. This can be characterized by the use of the Gaussian mixture distribution. The objective of this case is to determine the performance of the proposed method for a Gaussian mixture distribution. In order to do so, the measurement data are generated such that the error characteristics follow a Gaussian mixture distribution noise with zero mean and 10% of measurement variance. For variables x_2 and x_7, the first mode is possible to occur 65%, and the bias magnitude for both of these two variables is 3. Table 2.8 presents true and reconciled values of all variables. Figure 2.8 shows the histogram of the rectified values for 50 runs, when there are gross errors in x_2 and x_7.

Table 2.8: Reconciliation results for mixture distribution with multiple gross errors

Variables	x	$\hat{x}$	$\hat{\delta}_{i1}$	$\hat{\sigma}_{i1}$	$\hat{\sigma}_{i2}$
x_1	1	1.0180	0.9000	0.2972	0.3055
x_2	2	2.0251	0.6566	0.4014	2.8074
x_3	3	3.0378	0.9232	0.6092	0.2631
x_4	1	1.0071	0.9700	0.3791	0.0904
x_5	2	2.0307	0.9970	0.5020	0.1722
x_6	1	1.0127	0.9700	0.2929	0.3954
x_7	1	1.0180	0.6404	0.3081	3.0165

Table 2.9: Performance measures for mixture distribution with multiple gross errors

Simulation runs	Biased	OP	AVTI	CR
50	x_2	1	0	1

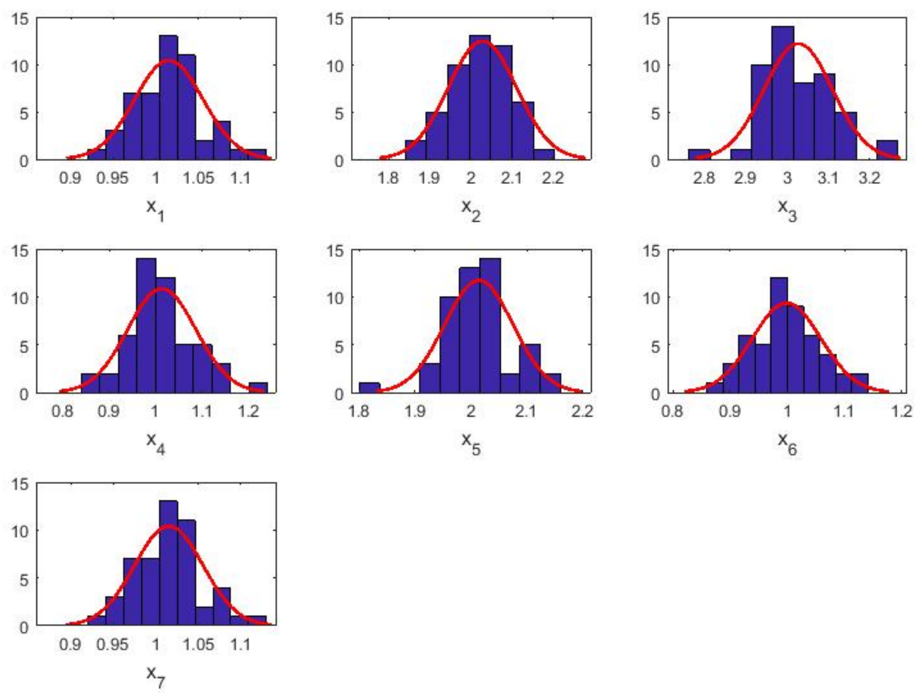

Figure 2.8: Histogram of reconciled values for 50 runs of Case 4

Based on the results and figures presented, it can be concluded that the developed method and framework can handle partially gross errors as well.

2.5.2 Illustration 2 - Nonlinear Example

The purpose of this case study is to demonstrate the efficacy of the proposed approach for nonlinear systems. To this end, we consider the nonlinear example of Yuan et al.[19] and Tjoa and Biegler[20]. This problem contains eight variables, five of which are measured and the rest, unmeasured. There are six nonlinear constraints given by,

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

The exact values for all of the variables are given by,

$$
\begin{aligned}
x_{exact} &= [4.5124,\ 5.5819,\ 1.9260,\ 1.4560,\ 4.8545]^T \\
u_{exact} &= [11.070,\ 0.61467,\ 2.0504]^T
\end{aligned}
$$

In this case, three different situations are investigated to present the effectiveness of the proposed approach for data reconciliation using the EM algorithm, and then the results are compared with the performance measures reported in literature. In Scenario 1, the set of measurements is divided into 5 blocks with 20 gross errors in each block in sequence from x_1 to x_5. Scenario 2 is the case that all the measured variables are added with gross errors in every fifth run and no gross errors in other runs. Finally, in Scenario 3, one gross error is added for each run in sequence from x_1 to x_5 in rotation.

The data is simulated with the mean values equal to the true values, and has the variance 0.1. The magnitudes of the gross errors considered in this study are +0.4 and +1.

Table 2.10: Comparison of the results of the proposed method with previous methods for bias = + 0.4

	x_1		x_2		x_3		x_4		x_5	
Scenario	R	W	R	W	R	W	R	W	R	W
1 Proposed Method	17	-	20	-	20	-	20	-	20	-
1 Yuan et al.[19]	5	2	20	6	20	-	20	2	20	5
1 Tjoa and Biegler[20]	8	3	19	-	20	-	9	-	17	1
2 Proposed Method	15	-	20	-	20	-	19	-	20	-
2 Yuan et al.[19]	8	-	20	1	20	-	15	-	20	-
2 Tjoa and Biegler[20]	-	-	20	-	20	2	-	-	20	1
3 Proposed Method	10	-	20	-	20	-	20	-	20	-
3 Yuan et al.[19]	3	2	20	8	20	-	20	2	20	-
3 Tjoa and Biegler[20]	4	-	18	-	20	1	9	-	14	2

For the case where bias = +0.4, the reconciled values of measured variables and the estimated values of unmeasured values are given by,

$$\hat{x} = [4.5402,\ 5.5666,\ 1.9221,\ 1.4735,\ 4.8441]^T$$
$$\hat{u} = [11.1174,\ 0.6151,\ 2.0487]^T$$

Table 2.11: Comparison of the results of the proposed method with previous methods for bias = + 1

	x_1		x_2		x_3		x_4		x_5	
Scenario	R	W	R	W	R	W	R	W	R	W
1 Proposed Method	20	-	20	-	20	-	20	-	20	-
1 Yuan et al.[19]	20	2	20	-	20	1	20	1	20	2
1 Tjoa and Biegler[20]	19	-	20	1	20	-	19	1	20	2
2 Proposed Method	19	-	20	-	20	-	20	-	20	-
2 Yuan et al.[19]	10	-	20	-	20	-	19	-	20	-
2 Tjoa and Biegler[20]	12	-	20	-	20	1	20	1	17	1
3 Proposed Method	17	-	20	-	20	-	20	-	20	-
3 Yuan et al.[19]	20	1	20	2	20	2	20	2	20	1
3 Tjoa and Biegler[20]	20	2	20	-	20	-	19	-	20	1

For the case in which bias = +1, the reconciled and estimated values are given by,

$$\hat{x} = [4.5704,\ 5.5638,\ 1.9178,\ 1.5110,\ 4.8163]^T$$
$$\hat{u} = [11.1620,\ 0.6154,\ 2.0406]^T$$

In Table 2.10 and Table 2.11, R and W refer to Right and Wrong, respectively. Right denotes there is gross error in a measurement and it has been detected correctly, and Wrong means there is no gross error in a measurement, but it has been incorrectly detected as gross error measurement. It is important to note that the gross errors in variables x_2, x_3 and x_5 have been identified correctly at all times whereas the gross error in variable x_4 has been identified in 19 of 20 simulation runs. The gross error in variables x_1 has been identified correctly most of the time and the results reveal that the proposed method shows improved performance compared to that of Yuan et al.[19] and Tjoa and Biegler[20]. Moreover, our method does not wrongly identify gross errors in any of the variables.

2.5.3 Illustration 3 - Realistic Mineral Processing Example

In this example, the proposed method is applied to a realistic mineral processing plant[1]. This process contains 16 streams, and each of them is characterized by a flow rate and two concentration variables. The process schematic is shown in Figure 2.9. All of the concentration variables are measured, and the flow rates except streams 1, 4 and 11 are measured. Therefore, there are 45 measured variables and 3 unmeasured variables in total. In order to demonstrate the approach, seven measured variables are assumed to be corrupted by gross errors with the magnitude of variables reported in Table 2.13.

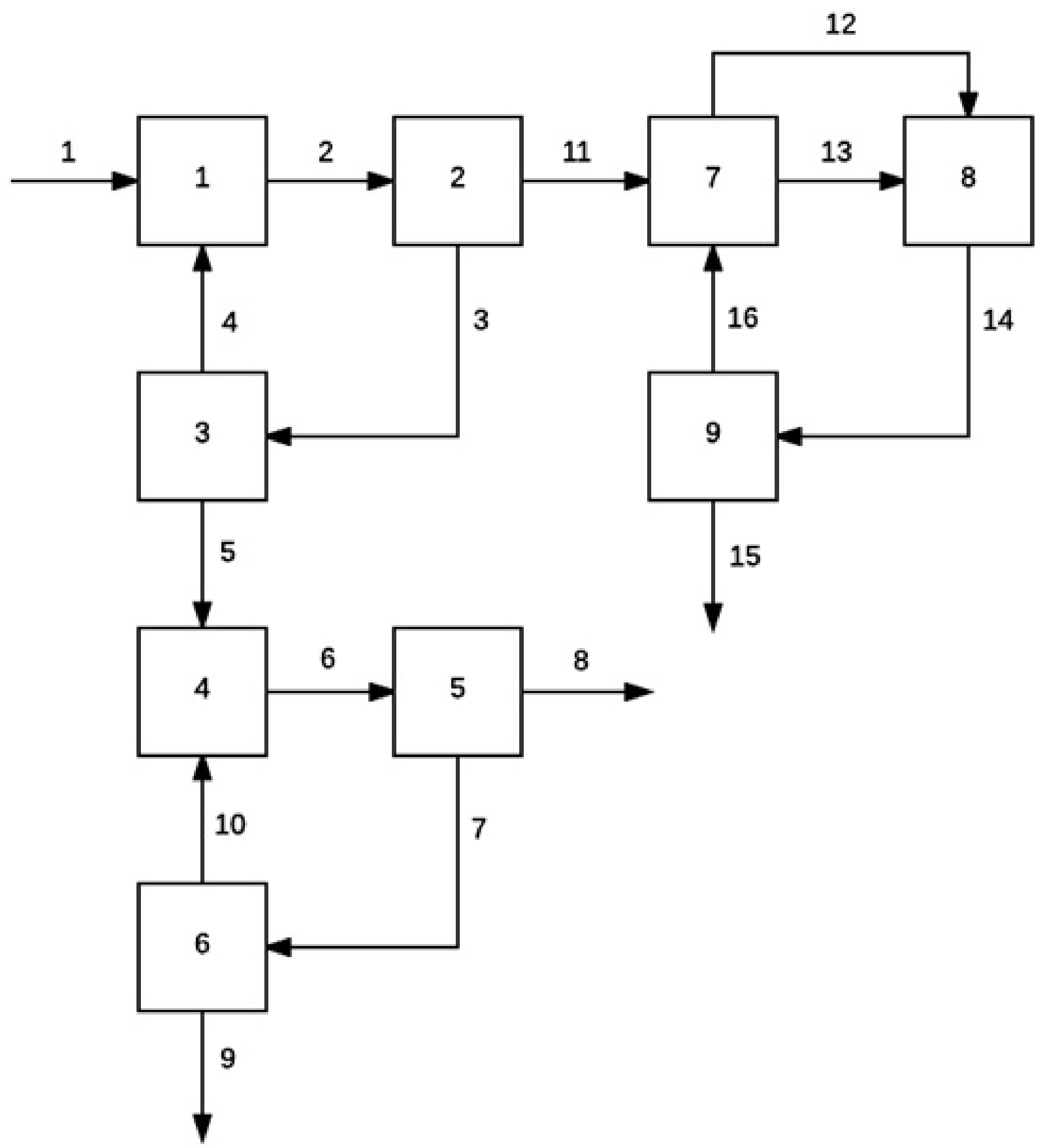

Figure 2.9: Process flow diagram

Table 2.12 represents measurements, standard deviations, and reconciled estimates. In order to analyze the performance of the proposed method, we define the relative correction in terms of a percentage of the measurement for each variable as $c(r) = 100(\hat{y}_i - y_i)/y_i$.

Table 2.12: Measurements, standard deviations, and reconciled estimates

	Measurement			Standard deviation			Rectified estimate		
	x	y_1	y_2	Std(x)	Std(y_1)	Std(y_2)	$\hat{x}$	$\hat{y_1}$	$\hat{y_2}$
1		3.890	3.440		0.389	0.344	22.565	2.408	3.452
2	26.500	2.700	3.530	1.325	0.270	0.353	25.874	2.710	3.651
3	29.200	2.520	3.550	1.460	0.252	0.355	20.895	2.517	3.533
4		4.750	5.000		0.475	0.500	3.309	4.770	5.007
5	18.320	2.090	3.290	0.916	0.209	0.329	17.586	2.093	3.255
6	22.020	2.460	3.510	1.101	0.246	0.351	21.722	2.488	3.502
7	20.800	2.900	3.740	1.040	0.290	0.374	12.203	2.863	3.722
8	9.430	2.010	5.200	0.472	0.201	0.520	9.519	2.006	3.220
9	8.010	3.710	3.290	0.401	0.371	0.329	8.067	2.196	3.297
10	4.140	4.150	4.550	0.207	0.415	0.455	4.136	4.164	4.550
11		3.490	4.190		0.349	0.419	4.980	3.518	4.146
12	6.560	3.630	4.340	0.328	0.363	0.434	6.539	3.537	4.243
13	1.040	8.710	6.320	0.052	0.871	0.632	1.040	8.312	6.254
14	7.380	4.240	6.650	0.369	0.424	0.665	7.579	4.192	4.519
15	4.990	3.490	4.100	0.250	0.349	0.410	4.980	3.518	4.146
16	7.690	5.150	5.180	0.385	0.515	0.518	2.599	5.484	5.233

Table 2.13: True and estimated values of gross errors

	x_3	x_7	x_{16}	$y_{1,1}$	$y_{1,9}$	$y_{2,8}$	$y_{2,14}$
True Bias	8.0	8.0	5.0	1.5	1.5	2.0	2.0
Proposed method	8.31	8.59	5.09	1.48	1.51	1.98	2.13
Alhaj-Dibo et al.[1]	8.52	8.51	4.87	1.46	1.45	1.95	2.10

Figure 2.10 shows the relative corrections $c(r)$ obtained for each measurements using our proposed method while Figure 2.11 represents the relative correction obtained using conventional data reconciliation. It should be noted that the relative corrections for the measurements with gross errors are high, and for other measurements it is minimal. On the other hand, the relative correction obtained using the conventional data reconciliation approach shows that the relative correction occurs in all the measurements thus showing smearing effect.

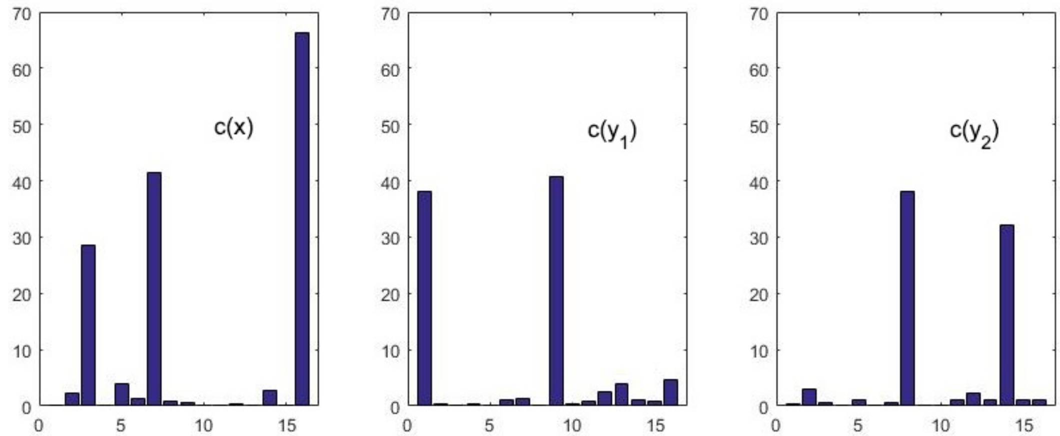

Figure 2.10: Relative corrections by proposed method

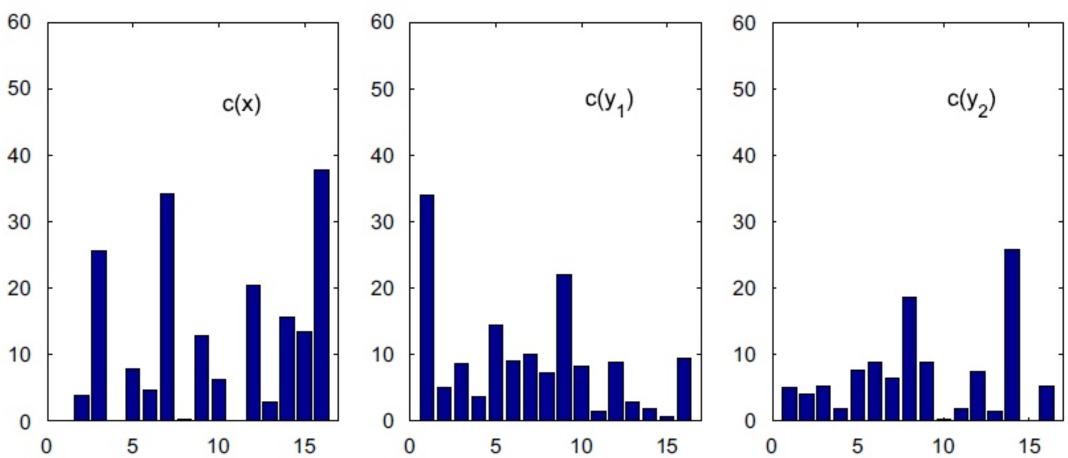

Figure 2.11: Relative corrections by conventional data reconciliation[1]

2.6 Conclusions

In this chapter, we addressed the problem of data reconciliation in the maximum likelihood framework by simultaneously eliminating the gross errors present in the measurement data. The Gaussian mixture distribution model characterizes the error distribution of different noise modes and hence, the resulting formulation involves solving for hidden variables. Therefore, the expectation maximization approach was introduced. It is important to note that the parameters of the error distribution model are not preset and are determined as part of the solution. Several case studies were presented to demonstrate the various features of the algorithm. For the linear example, with single gross error cases (*Case* 1 and *Case* 3), the performance measures showed that the gross error was identified correctly in all the simulation runs. For the case of multiple gross errors in which the data are generated with the Gaussian mixture distribution (*Case* 4), the performance measures of the proposed method achieved theoretical optimal values. From the nonlinear example, it was seen that the proposed approach showed superior performance over methods from literature. Using the realistic mineral processing example, we demonstrated that because of the assumed characterization of noise distribution, our proposed method did significant adjustments to the biased measurements appropriately.

Chapter 3

Simultaneous Data Reconciliation and Gross Error Detection Based on Maximum A Posteriori (MAP) Estimation via EM Algorithm

The raw measurements collected from industrial plants operations contain significant information for process identification, online optimization, and process control. However, they are prone to random and gross errors that may affect the subsequent applications, i.e., biased results for estimation. In this work, we formulate the data rectification problem to eliminate both random and gross errors by assuming a mixture Gaussian distribution for the measurement noise model. Some of the previously proposed models for data reconciliation and gross error estimation cannot determine the true set of reconciled estimates for arbitrary sets of gross error candidates. That said, some of the candidates' combinations result in the same value for the objective function of data reconciliation. Previous works, which did not utilize prior information, mainly fail in distinguishing the true set among the equivalent sets of gross errors. Therefore, in this work, the proposed framework is able to rectify measurements in the presence of equivalent sets of gross errors by incorporating prior knowledge into the objective function. Based on this assumption, a Maximum a Posteriori (MAP) framework is proposed for simultaneous data reconciliation and gross error detection. The problem is solved using the Expectation Maximization (EM) algorithm due to the presence of the hidden variable in the noise model, i.e., noise mode. The proposed method provides the magnitude of gross errors in addition to the ability of detecting partial gross errors. Thus, it is not required to assume that gross errors exist in the entire data set. In order to illustrate the efficiency and the performance of the proposed method, several case studies are presented.

3.1 Introduction

Raw measurements play a significant role in chemical plants for process monitoring, identification, control, and so on. However, there are errors in these measured variables that are incurred by improper instrumentation, poor calibration, and other unmeasured errors. Consequently, the collected data generally are not expected to follow process balances, i.e., mass and energy balances[29–32]. Errors which are caused by random and nonrandom events are addressed by data reconciliation and gross error detection, respectively. Using error contaminated data leads to suboptimal and even unsafe process operation[33]. Therefore, obtaining rectified data is the basis for efficient process operation and control[19,33,34]. The theory of data reconciliation has been developed to resolve the contradictions between the raw measurements and their constraints in order to obtain consistent information[35]. Gross error detection has been well investigated in the literature using different theoretical and numerical methods, which is an essential step for industrial applications[32]. There are three main approaches to remove the effect of gross errors in the measurements: (1) sequential gross error detection and data reconciliation, (2) simultaneous gross error detection and data reconciliation, and (3) robust data reconciliation that obtains reconciled estimates by ignoring gross errors without actually identifying them. As a result, data reconciliation and gross error detection, or data rectification, have to be performed before conducting any further analysis.

Data rectification and data reconciliation are two distinct terms which are often used indiscriminately. Data rectification is used to obtain an estimate of true values by removing both random and gross errors, and literally means "to make data right", or data correction, while the objective of data reconciliation is to adjust data such that they follow process balance models[21,36]. Data reconciliation is a widely used approach for improving the accuracy of measurements with the help of the process model[37]. Initially, it was assumed there are only random errors in the measurements, i.e., absence of gross errors. However, the existence of gross errors in data limits the usage of conventional data reconciliation solutions due to the smearing effect[34]. If measurements are adjusted in the presence of gross error(s) in order to follow the process model, all the adjustments are affected by the bias measurements (gross errors) and would not be a reliable indicator of the true values[31]. Therefore, gross errors should be removed prior to or during data reconciliation. In order to remove gross errors prior to data reconciliation, a statistical test must be performed to first identify gross error candidates and then remove them in the subsequent step. However, to remove gross errors during data reconciliation, a robust objective function must be defined such that gross errors are eliminated while trying to reconcile the data. Several statistical tests have been developed to detect gross errors, like the global test[4,5], nodal test[5–7], measurement test[5,8,9], generalized likelihood ratios[10], the Bonferroni tests[11], and principal component tests[12]. Therefore, data reconciliation and gross error detection are techniques that help with producing accurate estimates while identifying instrument malfunction[37]. The statistical tests mentioned above have their unique approaches/methodologies to detect the presence of gross errors, namely, the global and nodal tests are based on model constraint residuals, and the measurement test is based on the residuals between measurements and estimations[29]. However, it is desirable to not only detect the gross errors, but to identify their locations and magnitudes as well[13].

In this work, we have formulated the data rectification problem to eliminate both random and gross errors. There are plenty of books and articles that discuss different aspects of data reconciliation and gross error detection, and extensive effort has been devoted to this problem[30].

The data rectification solution is vulnerable to equivalent sets of gross errors which are defined as follows: "Two sets of gross errors are equivalent if they have the same effect in the data reconciliation, i.e., when eliminating either one, leads to the same value of the objective function." In other words, utilizing each of the rectified values leads to the same objective function value of data rectification[13,34]. Therefore, the equivalent sets of gross errors are theoretically indiscernible, and in the presence of an equivalent set of gross errors, there is an equal probability that the true locations of gross errors will be in one of its equivalent sets[37]. Thus, the existence of equivalent sets of gross errors leads to several undesirable situations. For instance, it is possible that the wrong set of errors are identified using gross error detection approaches which could result in the removal of good measurements and therefore, incorrectly leave biased measurements that will smear the reconciliation[37]. As a result, if there are equivalent sets of gross errors in the measurements, it would not be solved by Maximum Likelihood Estimation (MLE). In order to resolve this issue, the Maximum a Posteriori (MAP) estimation is applied, or in other words, the prior information must be considered in the objective function, which is the main motivation of this work.

Linear and/or nonlinear constraints are an inevitable part of the data rectification problem. Almost all the research on gross error detection consider only linear or linearized models under steady state conditions. That said, nonlinear process models are linearized first, and then data reconciliation is performed on the measurements[19]. However, this linearization may introduce other errors to the data as well[21]. Rollins and Roelfs[38] solved the nonlinear data reconciliation problem by extending the unbiased estimation method[11] to bilinear constraints. Herein, we consider both linear and nonlinear systems operating under steady state conditions. Further, the proposed algorithm performs gross error elimination and data reconciliation simultaneously and also estimates the magnitudes of the gross errors. Moreover, the distribution parameters of errors and the occurrence probability of gross errors have to be estimated. The rest of this chapter is organized as follows: In section 3.2, data reconciliation and rectification are defined, and then in section 3.3, the problem statement is discussed and all assumptions are provided. The proposed approach is elaborated in section 3.4 by elaborating the general framework and then applying it to our case. In section 3.5, the simulation results for different examples are discussed in order to evaluate the performance of the proposed approach. Finally, in section 3.6, conclusions are provided.

3.2 Data Reconciliation and Rectification

As mentioned previously, raw measurements are corrupted by errors, and data reconciliation is a procedure of optimally adjusting contaminated measurements such that they satisfy process models. Due to the existence of errors, measurements can be modelled as given in Equation (3.1),

$$y_i = x_i + \epsilon_i \tag{3.1}$$

where y_i and x_i denote the measurement and true value of the i^{th} variable, respectively. Given the process model of the form, $\mathbf{f}(\mathbf{x}) = 0$, and assuming the measurement error ϵ_i to be independent of x_i, the general formulation of a steady state data rectification problem can be expressed as[21]:

$$\arg\max_{\mathbf{x}} \quad P\{\mathbf{x}|\mathbf{y}\} \tag{3.2}$$

$$\text{s.t.} \quad \mathbf{f}(\mathbf{x}) = 0 \tag{3.3}$$

$$\mathbf{g}(\mathbf{x}) \leq 0 \tag{3.4}$$

where the objective function signifies the probability of state given measurements, and $\mathbf{g}(\mathbf{x})$ denotes the inequality constraints signifying the safety limits or variable bounds. Due to the presence of errors, the raw measurements do not satisfy the process model and hence, solving the above optimization problem obtains the rectified values by maximizing the posterior probability of states. By applying the Bayes rule, the posterior probability of states can be converted in terms of the probability of the measurements given the states, the prior probability of the states, and the probability of the raw measurements.

$$\arg\max_{\mathbf{x}} \quad P\{\mathbf{x}|\mathbf{y}\} = \arg\max_{\mathbf{x}} \quad P\{\mathbf{y}|\mathbf{x}\}P\{\mathbf{x}\}/P\{\mathbf{y}\} \tag{3.5}$$

In the above equation, $P\{\mathbf{y}\}$ is independent of $\mathbf{x}$, and therefore, it can be excluded from Equation (3.5):

$$\arg\max_{\mathbf{x}} \quad P\{\mathbf{x}|\mathbf{y}\} = \arg\max_{\mathbf{x}} \quad P\{\mathbf{y}|\mathbf{x}\}P\{\mathbf{x}\} \tag{3.6}$$

If the prior probability of the states are assumed to have uniform distribution, the rectification problem is changed to a reconciliation problem. For linear processes with no inequality constraints and also assuming zero-mean Gaussian distribution of measurement errors, the reconciliation problem can be stated as a weighted least squares problem

$$\min_{\mathbf{x}} \quad (\mathbf{y}-\mathbf{x})^T V^{-1}(\mathbf{y}-\mathbf{x}) \tag{3.7}$$

$$\text{s.t.} \quad A\mathbf{x} = 0 \tag{3.8}$$

where A is the model coefficient matrix, and V is the covariance matrix of measurements which is often assumed to be a diagonal matrix. The reconciled estimates is given in vector form by the following closed-form solution[2]:

$$\hat{\mathbf{x}} = [I - VA^T(AVA^T)^{-1}A]\mathbf{y} \tag{3.9}$$

In the above data reconciliation problem, it is assumed that there are no gross errors, and therefore, the presence of gross errors invalidates the statistical basis of the reconciliation problem and results in the smearing effect, since the above solution only adjusts raw measurements such that they obey the process model constraints[2]. In order to elucidate on the concept of the smearing effect, consider the network shown in Figure 3.1 which has three units with five flow rates.

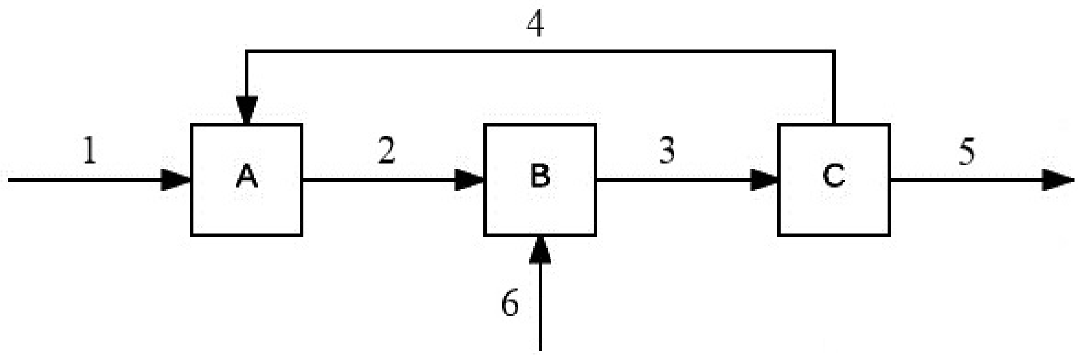

Figure 3.1: Diagram of process network

Suppose the true flow rate values are $\mathbf{x} = [1 \ \ 2 \ \ 3 \ \ 1 \ \ 2 \ \ 1]^T$, but x_2 includes gross error with magnitude 3, i.e., x_2 fluctuates around 5. The reconciled estimate vector for this case is $\hat{\mathbf{y}} = [1.83 \ \ 3.59 \ \ 3.79 \ \ 1.76 \ \ 2.03 \ \ 0.20]^T$ which is a poor estimate of the true values. This phenomenon is called smearing effect. A wide variety of methods and algorithms have been developed to address the effect of gross errors prior or simultaneously by performing data reconciliation. It is often assumed that ϵ_i follows a non-zero mean Gaussian distribution (i.e., $\epsilon_i \sim N(b_i, \sigma_i^2)$) for the measurement model given by Equation (3.1). Equivalently, ϵ_i can be characterized using a zero-mean Gaussian distribution with measurement model given by $y_i = x_i + b_i + \epsilon_i$. Using these noise characteristics and form of the measurement model, the simultaneous gross error detection and data reconciliation can be formulated as a mixed-integer optimization problem[17,18]. Recently, Yuan et al.[19] proposed a hierarchical Bayesian framework that iteratively solves for reconciled values, gross error detection, and hyperparameters of the noise model. Apart from these, robust objective functions can be used in order to address the effect of gross errors while obtaining the reconciled estimates[1,24].

The generalized form of the robust objective function, θ, is given by[1,24,25],

$$\min_{\mathbf{x}} \quad \sum_{i=1}^{n} \rho(\epsilon_i) \tag{3.10}$$

$$\text{s.t.} \quad A\mathbf{x} = 0 \tag{3.11}$$

where $\epsilon_i = (y_i - x_i)/\sigma_i$ represents the standardized error, and ρ is any reasonable function of standardized error which reduces the effect of gross errors on estimation of true values. The robust objective function ρ can be derived from the maximum likelihood functions for contaminated Gaussian, Cauchy and Logistic distributions. Further, the Fair function, Lorentzian function and redescending M-estimator can also be used with the proper tuning parameters to eliminate the influence of gross errors and perform robust data reconciliation.[23]

There are various methods and algorithms that can be used to reconcile the measurements and also eliminate the effect of gross errors. Nonetheless, these methods cannot handle equivalent sets of gross errors. To this end, we need to use prior information in the objective function which transforms the objective function of the reconciliation problem to the rectification problem. Thus, the main objective of this chapter is to handle equivalent sets of gross errors by introducing a new method for data rectification. In order to clarify the concept of equivalent sets, consider a tank with only one input and output (Figure 3.2). In this example, the process constraint is the mass balance equation, i.e., $x_1 = x_2$. The measurements for the input and output flow rates are given in Table 3.1; it is assumed that there are no random errors. From Table 3.1, it can be seen that the reconciled values can be both sets, and the estimated gross error is ± 10.

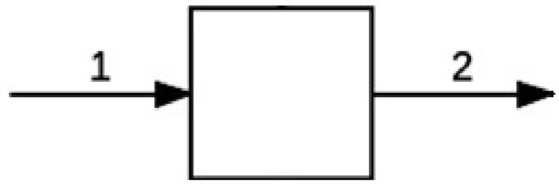

Figure 3.2: Isolated tank

Therefore, prior information must be incorporated to determine which sensor reading is biased. Depending on the importance of the measurement, prior information in industrial plants is typically obtained using two or more sensors to measure the same variable in the process. When there are three or more sensors, one can select the "majority view" by voting. In this method, m measurements are selected out of the total n number of signals such that $m > n/2$. Practically, n is 3 so that m is 2.[39] In the case that there are two sensors measuring the same variable, we should use history of actual flows to know in which regions the actual flows would tend to lie[21]. For instance, in this example, if the historical data shows the stream flow rates are in the range of 40, i.e., the stream follows a Gaussian distribution with the mean value of 40 and a particular standard deviation, it means that currently, the first sensor is not working properly.

Table 3.1: Isolated tank with an input and an output

		x_1	x_2
	Measurement	50	40
Case 1	Reconciled data	40	40
	Estimated gross error	-10	
Case 2	Reconciled data	50	50
	Estimated gross error		+10

Therefore, if prior information is considered in the objective function, a set of estimates close to the true values can be obtained. In the case where there is not any prior knowledge of $P(x)$, a uniform distribution is used to describe non-informative priors, and the Maximum A Posteriori (MAP) estimation problem is changed to a Maximum Likelihood Estimation

(MLE) problem[21]. If the network has more than a unit, it would not be so easy to determine equivalent sets of gross errors. However, determining equivalent sets of gross errors has already been investigated thoroughly in the literature. To this end, we need to define gross error cardinality and basic subset concepts:

Gross error cardinality : A set of variables has gross error cardinality $\Gamma = t$ if t is the minimum number of gross errors that are required to represent all possible sets of gross errors in the variables[34].

Basic subset : A set of variables constitutes a basic subset of a system when every set of gross errors is equivalent to a set of gross errors in the basic set[34].

Theorem 1. [34] *Let m columns $[a_1\ a_2\ ...\ a_m]$ of the coefficient matrix A correspond to a set of variables. The set has gross error cardinality $\Gamma = t$ if rank $[a_1\ a_2\ ...\ a_m]\ = t$.*

Suppose a set of gross errors with t elements has been identified. If t is equal to the gross error cardinality of the network, then the identified set is a basic subset of the network. A more detailed discussion of these concepts is provided in the first example of section 3.5.

3.3 Problem Statement

Let Y denote the data matrix of raw measurements as follows:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix} = \begin{bmatrix} \mathbf{y_1^v} \\ \mathbf{y_2^v} \\ \cdots \\ \mathbf{y_n^v} \end{bmatrix} = \begin{bmatrix} \mathbf{y_1^s} & \mathbf{y_2^s} & \cdots & \mathbf{y_m^s} \end{bmatrix} \tag{3.12}$$

where superscripts s and v refer to sample points and variables, respectively. The i^{th} row of the Y matrix represents the variable i measured at different sampling instants, and the j^{th} column represents the set of n variables measured at j^{th} sampling instant. Hence, the measurement model for all variables is given by,

$$\mathbf{y_j^s} = \mathbf{x} + \epsilon \tag{3.13}$$

where $\mathbf{x}$ is defined as,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \tag{3.14}$$

It should be noted that the measurement model given by Equation (3.13) is the vector form of the measurement model presented in Equation (3.1) at the j^{th} time instant.

Following our previous work, a Gaussian mixture distribution with two noise modes is assumed for the measurement error vector, ϵ, to account for the presence of gross errors. The sensor mode with small variance corresponds to random noise, and the other abnormal one with large variance corresponds to gross errors. Therefore, the measurement error distribution can be written for the i^{th} variable as,

$$\epsilon_i \sim \delta_{i1} N(0, \sigma_{i1}^2) + \delta_{i2} N(0, \sigma_{i2}^2) \tag{3.15}$$

such that $\delta_{i1} + \delta_{i2} = 1$. Representing gross errors with a probability density function with zero mean might appear as a contradiction to the definition of gross error. However, the main idea is that by assigning a large variance for the second mode of the error distribution, $\pm 3\sigma_{i2}$ could cover the magnitude of the gross error. Therefore, if there is an error between $-3\sigma_{i1}$ and $-3\sigma_{i2}$ or $+3\sigma_{i1}$ and $+3\sigma_{i2}$, it would be considered as gross error.

As mentioned previously, $P\{\mathbf{x}|\mathbf{y}\}$ is directly proportional to the probability of raw measurements and the prior probability of the states based on the Bayes rule, and also since both the noise modes are mutually exclusive, the sensor model for a single measurement set can be expressed as:

$$P\{\mathbf{x}|\mathbf{y}\} = P\{\mathbf{x}|\mathbf{y_j^s}\} \propto P\{\mathbf{y_j^s}|\mathbf{x}\}P\{\mathbf{x}\} = \sum_{k=1}^{2} P\{\mathbf{y_j^s}|\mathbf{x}, I_{ij} = k\}P\{I_{ij} = k\}P\{\mathbf{x}\} \tag{3.16}$$

where I_{ij} denotes the hidden sensor mode of variable i at the j^{th} sample point. This can be either in the normal mode or in abnormal mode. Using the above noise model, the log of the posterior probability of the set of m measurements can be defined as:

$$\arg\max_{\mathbf{x}} \quad lnP\{\mathbf{x}|Y\} = ln(\sum_{k=1}^{2} P\{Y|\mathbf{x}, I_{ij} = k\}P\{I_{ij} = k\}P\{\mathbf{x}\}) \tag{3.17}$$

Assuming the measurements are independent, the above equation becomes,

$$\arg\max_{\mathbf{x}} \quad lnP\{\mathbf{x}|Y\} = \prod_{i=1}^{n}\prod_{j=1}^{m} ln(\sum_{k=1}^{2} P\{y_{ij}|x_i, I_{ij} = k\}P\{I_{ij} = k\}P\{x_i\}) \tag{3.18}$$

where $p_{ijk} = P\{y_{ij}|x_i, I_{ij} = k\}$ is defined as follows,

$$p_{ijk}\{y_{ij}|x_i, \sigma_{ik}\} = \frac{1}{\sqrt{2\pi}\sigma_{ik}} exp(-\frac{(y_{ij} - x_i)^2}{2\sigma_{ik}^2}) \quad k = 1, 2 \tag{3.19}$$

and $P\{I_{ij} = k\} = \delta_{ik}$ denotes the probability of occurrence of noise mode k. In particular, $P\{I_{ij} = 1\} = \delta_{i1}$ is the probability of measurements in the normal mode where only random error is present, and $P\{I_{ij} = 2\} = \delta_{i2}$ denotes the probability of measurements in abnormal gross error mode for variable i. Prior works that utilize the Gaussian mixture model to characterize the noise distribution do not utilize prior information in their objective functions to handle equivalent sets of gross errors. Moreover, some of them use the prespecified values of noise distribution parameters $\theta_{ik} = \{\delta_{ik}, \sigma_{ik}^2\}$ for both the modes[1,20,23]. However, presetting the noise distribution parameters influences the performance of gross error detection and hence, affects the accuracy of the reconciled estimates. Therefore, in this work, we aim to determine the noise distribution parameters while obtaining the rectified estimates. Now, the simultaneous data reconciliation and gross error detection problem for linear steady state processes can be posed as follows:

$$\max_{\mathbf{x},\theta_{ik}} \quad lnP\{\mathbf{x}|Y\} = \prod_{i=1}^{n}\prod_{j=1}^{m} ln(\sum_{k=1}^{2} p_{ijk}\{y_{ij}|x_i, \theta_{ik}, I_{ij} = k\}P\{I_{ij} = k|\theta_{ik}\}P\{x_i\}) \tag{3.20}$$

$$s.t. \quad A\mathbf{x} = \mathbf{0} \tag{3.21}$$

From Equation (3.20), it can be observed that the model identity probability cannot be obtained explicitly, and hence, the maximum a posteriori estimation problem cannot be solved directly. To resolve this issue, the Expectation Maximization (EM) algorithm is applied to obtain the distribution parameter values for each variable. In other words, all of the model distribution parameters along with the rectified values are estimated simultaneously using the EM algorithm.

The assumptions considered in this study are as follows

1. Steady state process model is available, which can be expressed as,

$$f(\mathbf{x}, \mathbf{z}) = 0 \tag{3.22}$$

   where, $\mathbf{x}$ is a vector of measured variables, and $\mathbf{z}$ is a vector of unmeasured variables.

2. In Equation (3.13), the true value, $\mathbf{x}$, is unknown and is assumed to follow the Gaussian distribution,

$$x_i \sim N(\mu_i, \sigma_i^2) \tag{3.23}$$

3. The prior distribution variances for x_i and ϵ_i follow the Inverse Gamma distribution. It should be noted that the Inverse Gamma distribution is the conjugate prior for the

Gaussian and the Gaussian mixture likelihood function[40], and thus we have:

$$\sigma_i^2 \sim IG(\alpha_i, \beta_i) \tag{3.24}$$

$$\sigma_{ik}^2 \sim IG(\alpha_{ik}, \beta_{ik}) \tag{3.25}$$

Therefore, the probability density functions are given by,

$$P(\sigma_i^2|\alpha_i, \beta_i) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}(\sigma_i^2)^{-\alpha_i-1}exp(-\frac{\beta_i}{\sigma_i^2}) \tag{3.26}$$

$$P(\sigma_{ik}^2|\alpha_{ik}, \beta_{ik}) = \frac{\beta_{ik}^{\alpha_{ik}}}{\Gamma(\alpha_{ik})}(\sigma_{ik}^2)^{-\alpha_{ik}-1}exp(-\frac{\beta_{ik}}{\sigma_{ik}^2}) \tag{3.27}$$

Further, each component of the mixture Gaussian follows a Dirichlet(also known as multivariate Beta) distribution[40]; if there are only two modes for mixture Gaussian, they follow the Beta distribution, so we used the Beta distribution as a prior for the probability of occurrence of each mode, δ_{ik}:

$$P(\delta_{i1}, \delta_{i2}|\phi_i) = \frac{\Gamma(\phi_i)}{\Gamma(\phi_i/2)^2}\prod_{j=1}^{2}\delta_{ik}^{\phi_i/2-1} \tag{3.28}$$

Therefore,

$$P(\delta_{ik}|\phi_i) \propto \delta_{ik}^{\phi_i/2-1} \tag{3.29}$$

If there is no a priori information for the hyperparameters, a uniform distribution can be applied to express noninformative priors[19].

3.4 Proposed Approach

In this section, the general formulation of the Expectation Maximization method in a maximum a posteriori framework will be presented. Subsequently, the method will be applied on the data rectification problem in order to solve simultaneous data reconciliation and gross error detection, and parameter estimation.

3.4.1 General Formulation for the Maximum A Posteriori Estimation Using EM Algorithm

In order to solve the MAP estimation problem with hidden variable, the EM algorithm is applied. The EM algorithm can be modified to obtain the maximum a posteriori (MAP) estimate[26]. Suppose the complete data is $\{C_{obs}, C_{mis}\}$, where C_{obs} refers to observed data and C_{mis} denotes hidden variables. The objective is to find the MAP estimate of the parameter Θ, which can be mathematically expressed as:

$$\begin{aligned} \hat{\Theta} &= arg\ max_{\Theta}\ P(\Theta|C_{obs}, C_{mis}) \\ &= arg\ max_{\Theta}\ P(C_{obs}, C_{mis}|\Theta)\ P(\Theta) \end{aligned} \tag{3.30}$$

Since the formulation involves hidden variables, the EM algorithm is often used. It is an iterative two step algorithm where in the first step, also known as the Expectation or E-step, the conditional expectation of hidden variables (R function) is obtained, and in the second step, also known as the Maximization or M-step, the parameters are optimized.

3.4.1.1 E-step

In this step, $\{C_{mis}\}$ is replaced by its conditional expectation given $\{C_{obs}\}$, using the current fit for Θ. Therefore, the R function is defined as

$$
\begin{aligned}
R(\Theta, \Theta^{(l)}) &= E_{C_{mis}|C_{obs},\Theta^{(l)}}[ln\ P(C_{mis}, C_{obs}|\Theta)] + ln\ P(\Theta) \\
&= \int P(C_{mis}|C_{obs}, \Theta^{(l)})[ln\ P(C_{mis}, C_{obs}|\Theta)]\ dC_{mis} + ln\ P(\Theta)
\end{aligned} \tag{3.31}
$$

It can be seen that in the E-step the first term is the Q function in MLE framework, so

$$
R(\Theta, \Theta^{(l)}) = Q(\Theta, \Theta^{(l)}) + ln\ P(\Theta) \tag{3.32}
$$

3.4.1.2 M-step

After obtaining the R function in the E-step, the M-step requires the maximization of $R(\Theta, \Theta^{(l)})$ with respect to Θ over the parameter space.

$$
\Theta^{(l+1)} = arg\ max_{\Theta}\ R(\Theta, \Theta^{(l)}) \tag{3.33}
$$

The R function in the E-step is evaluated with the updated parameter values $\Theta^{(l+1)}$ and the procedure is repeated until convergence.

3.4.2 Data Rectification based on MAP Estimation Using EM Algorithm

This thesis attempts to develop a method for data rectification in the presence of random errors and gross errors. Therefore, the error distribution is assumed to follow the mixture Gaussian distribution, that is $\epsilon_i^v \sim \delta_{i1}N(0, \sigma_{i1}^2) + \delta_{i2}N(0, \sigma_{i2}^2)$. In this section, the detailed solution algorithm is presented.

3.4.2.1 E-step

In the mixture normal distribution, the observed data is Y, and the hidden variable is the model identity, $I_{ij} = \{1, 2\}$. Parameters which have to be estimated are denoted as Θ.

Hence, the E-step in the EM algorithm can be illustrated as,

$$
\begin{aligned}
R(\Theta, \Theta^{(l)}) &= Q(\Theta, \Theta^{(l)}) + lnP(\Theta) \\
&= E_{I|Y,\Theta^{(l)}}[lnP(Y, I|\Theta)] + lnP(\Theta) \\
&= E_{I|Y,\Theta^{(l)}}[ln \prod_{i=1}^{n} P(Y_i^v, I_i^v|\theta_i)] + ln \prod_{i=1}^{n} P(\theta_i) \\
&= E_{I|Y,\Theta^{(l)}}[\sum_{i=1}^{n} lnP(Y_i^v, I_i^v|\theta_i)] + \sum_{i=1}^{n} lnP(\theta_i) \\
&= \sum_{i=1}^{n} E_{I_i^v|Y_i^v,\theta_i^{(l)}}[lnP(Y_i^v, I_i^v|\theta_i)] + \sum_{i=1}^{n} lnP(\theta_i) \\
&= \sum_{i=1}^{n} E_{I_i^v|Y_i^v,\theta_i^{(l)}}[ln \prod_{j=1}^{m} P(y_{ij}, I_{ij}|\theta_i)] + \sum_{i=1}^{n} lnP(\theta_i) \\
&= \sum_{i=1}^{n} E_{I_i^v|Y_i^v,\theta_i^{(l)}}[\sum_{j=1}^{m} lnP(y_{ij}, I_{ij}|\theta_i)] + \sum_{i=1}^{n} lnP(\theta_i) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} E_{I_i^v|Y_i^v,\theta_i^{(l)}}[lnP(y_{ij}, I_{ij}|\theta_i)] + \sum_{i=1}^{n} lnP(\theta_i) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} P(I_{ij} = k|y_{ij}, \theta_{ij}^{(l)})[lnP(y_{ij}, I_{ij} = k|\theta_{ik})] + \sum_{i=1}^{n} \sum_{k=1}^{2} lnP(\theta_i)
\end{aligned}
\tag{3.34}
$$

$P(I_{ij} = k|y_{ij}, \theta_{ij}^{(l)})$ is the posterior distribution of the model identity in Equation (3.34); based on Bayes theorem, it is defined as,

$$
\begin{aligned}
P(I_{ij} = k|y_{ij}, \theta_{ij}^{(l)}) &= \frac{P(y_{ij}|I_{ij} = k, \theta_{ik}^{(l)} P(I_{ij} = k|\theta_{ik}^{(l)})}{P(y_{ij}|\theta_{ik}^{(l)})} \\
&= \frac{P(y_{ij}|I_{ij} = k, \sigma_{ik}^{2\ (l)}, x_i^{(l)})P(I_{ij} = k|\theta_{ik}^{(l)})}{\sum_{t=1}^{2} P(y_{ij}|I_{ij} = t, \theta_{it}^{(l)})P(I_{ij} = t|\theta_{it}^{(l)})} \\
&= \frac{\frac{1}{\sqrt{2\pi\sigma_{ik}^{2\ (l)}}} exp(-\frac{(y_{ij}-x_i^{(l)})^2}{2\sigma_{ik}^{2\ (l)}})P(I_{ij} = k|\theta_{ik}^{(l)})}{\sum_{t=1}^{2} \frac{1}{\sqrt{2\pi\sigma_{it}^{2\ (l)}}} exp(-\frac{(y_{ij}-x_i^{(l)})^2}{2\sigma_{it}^{2\ (l)}})P(I_{ij} = t|\theta_{it}^{(l)})}
\end{aligned}
\tag{3.35}
$$

Since $P(I_{ij} = k|\theta_{ik}^{(l)})$ is the prior distribution of the model identity, it is equal to the weight of the k^{th} component, i.e., $P(I_{ij} = k|\theta_{ik}^{(l)}) = \delta_{ik}^{(l)}$; so Equation (3.35) can be written as:

$$P(I_{ij} = k|y_{ij}, \theta_{ij}^{(l)}) = \frac{\frac{\delta_{ik}^{(l)}}{\sqrt{2\pi\sigma_{ik}^{2\,(l)}}} exp(-\frac{(y_{ij}-x_i^{(l)})^2}{2\sigma_{ik}^{2\,(l)}})}{\sum_{t=1}^{2} \frac{\delta_{it}^{(l)}}{\sqrt{2\pi\sigma_{it}^{2\,(l)}}} exp(-\frac{(y_{ij}-x_i^{(l)})^2}{2\sigma_{it}^{2\,(l)}})} \tag{3.36}$$

In the above equation, all parameters are obtained from the prior iteration, so parameters are assumed to be known, and can be simplified as $P(I_{ij} = k|y_{ij}, \theta_{ij}^{(l)}) = \gamma_{ijk}^{(l)}$

$$\begin{aligned} lnP(y_{ij}, I_{ij} = k|\theta_{ik}) &= ln(\frac{1}{\sqrt{2\pi\sigma_{ik}^2}} exp(\frac{-(y_{ij} - x_i)^2}{2\sigma_{ik}^2}) P(I_{ij} = k|\theta_{ik})) \\ &= ln(\frac{\delta_{ik}}{\sqrt{2\pi\sigma_{ik}^2}} exp(\frac{-(y_{ij} - x_i)^2}{2\sigma_{ik}^2})) \\ &= ln(\delta_{ik}) - ln(\sqrt{2\pi\sigma_{ik}^2}) - \frac{(y_{ij} - x_i)^2}{2\sigma_{ik}^2} \end{aligned} \tag{3.37}$$

Substituting Equations (3.36) and (3.37) into the first term of Equation (3.34), i.e., the Q function, the following equation is obtained:

$$Q(\Theta, \Theta^{(l)}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{ijk}^{(l)} [ln(\delta_{ik}) - ln(\sqrt{2\pi\sigma_{ik}^2}) - \frac{(y_{ij} - x_i)^2}{2\sigma_{ik}^2}] \tag{3.38}$$

The joint prior probability of the parameters is given by,

$$\begin{aligned} \sum_{i=1}^{n} \sum_{k=1}^{2} lnP(\theta_{ik}) &= \sum_{i=1}^{n} \sum_{k=1}^{2} (lnP(x_i) + lnP(\sigma_i^2) + lnP(\delta_{ik}) + lnP(\sigma_{ik}^2)) \\ &= -\sum_{i=1}^{n} \sum_{k=1}^{2} (ln\sigma_i + \frac{(x_i - \mu_i)^2}{2\sigma_i^2} + 2(\alpha_i + 1)ln\sigma_i + \frac{\beta_i}{\sigma_i^2} - (\phi_i/2 - 1)ln(\delta_{ik}) \\ &+ 2(\alpha_{ik} + 1)ln\sigma_{ik} + \frac{\beta_{ik}}{\sigma_{ik}^2}) \end{aligned} \tag{3.39}$$

Substituting Equations (3.38) and (3.39) in the R function, the following expression is obtained:

$$\begin{aligned}
R(\Theta,\Theta^{(l)}) &= Q(\Theta,\Theta^{(l)}) + \sum_{i=1}^{n}\sum_{k=1}^{2} lnP(\theta_{ik}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}\gamma_{ijk}^{(l)}[ln\delta_{ik} - ln\sqrt{2\pi\sigma_{ik}^2} - \frac{(y_{ij}-x_i)^2}{2\sigma_{ik}^2}] \\
&- \sum_{i=1}^{n}\sum_{k=1}^{2}[ln\sigma_i + \frac{(x_i-\mu_i)^2}{2\sigma_i^2} + 2(\alpha_i+1)ln\sigma_i + \frac{\beta_i}{\sigma_i^2} - (\phi_i/2-1)ln(\delta_{ik}) \\
&+ 2(\alpha_{ik}+1)ln\sigma_{ik} + \frac{\beta_{ik}}{\sigma_{ik}^2}] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}[\gamma_{ijk}^{(l)}(ln\delta_{ik} - ln\sqrt{2\pi\sigma_{ik}^2} - \frac{(y_{ij}-x_i)^2}{2\sigma_{ik}^2}) - \frac{(x_i-\mu_i)^2}{2\sigma_i^2} \\
&- (2\alpha_i+3)ln\sigma_i + (\phi_i/2-1)ln(\delta_{ik}) - \frac{\beta_i}{\sigma_i^2} - 2(\alpha_{ik}+1)ln\sigma_{ik} - \frac{\beta_{ik}}{\sigma_{ik}^2}]
\end{aligned} \tag{3.40}$$

3.4.2.2 M-step

In the M-step, the $R(\Theta,\Theta^{(l)})$ function is maximized with respect to Θ, and the update equations for the parameters are calculated. Now, the update expression for σ_{rt} can be derived as follows:

$$
\begin{aligned}
\frac{\partial R(\Theta,\Theta^{(l)})}{\partial \sigma_{rt}} &= \frac{\partial}{\partial \sigma_{rt}}(\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}\gamma_{ijk}^{(l)}(ln\delta_{ik} - ln\sqrt{2\pi\sigma_{ik}^2} - \frac{(y_{ij}-x_i)^2}{2\sigma_{ik}^2}) - \frac{(x_i-\mu_i)^2}{2\sigma_i^2} \\
&- (2\alpha_i+3)ln\sigma_i - \frac{\beta_i}{\sigma_i^2} - 2(\alpha_{ik}+1)ln\sigma_{ik} - \frac{\beta_{ik}}{\sigma_{ik}^2}) \\
&= \frac{\partial}{\partial \sigma_{rt}}(\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}\gamma_{ijk}^{(l)}(-ln\sqrt{2\pi\sigma_{ik}^2} - \frac{(y_{ij}-x_i)^2}{2\sigma_{ik}^2}) - 2(\alpha_{ik}+1)ln\sigma_{ik} - \frac{\beta_{ik}}{\sigma_{ik}^2}) \\
&= \frac{\partial}{\partial \sigma_{rt}}(\sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{ijt}^{(l)}(-ln\sqrt{2\pi\sigma_{it}^2} - \frac{(y_{ij}-x_i)^2}{2\sigma_{it}^2}) - 2(\alpha_{it}+1)ln\sigma_{it} - \frac{\beta_{it}}{\sigma_{it}^2}) \\
&= \frac{\partial}{\partial \sigma_{rt}}(\sum_{j=1}^{m}\gamma_{rjt}^{(l)}(-ln\sqrt{2\pi\sigma_{rt}^2} - \frac{(y_{rj}-x_r)^2}{2\sigma_{rt}^2}) - 2(\alpha_{rt}+1)ln\sigma_{rt} - \frac{\beta_{rt}}{\sigma_{rt}^2}) \\
&= \sum_{j=1}^{m}\gamma_{rjt}^{(l)}(-\frac{1}{\sigma_{rt}} + \frac{(y_{rj}-x_r)^2}{\sigma_{rt}^3}) - \frac{2(\alpha_{rt}+1)}{\sigma_{rt}} + \frac{2\beta_{rt}}{\sigma_{rt}^3} = 0
\end{aligned}
\tag{3.41}
$$

$$
(\sum_{j}^{m}\gamma_{rjt}^{(l)} + 2(\alpha_{rt}+1))\sigma_{rt}^2 - (\sum_{j}^{m}\gamma_{rjt}^{(l)}(y_{rj}-x_r)^2 + 2\beta_{rt}) = 0 \tag{3.42}
$$

The updating equation for σ_{rt} is given by,

$$
\sigma_{rt}^{(l+1)} = \sqrt{\frac{\sum_j^m \gamma_{rjt}^{(l)}(y_{rj}-x_r^{(l)})^2 + 2\beta_{rt}}{\sum_j^m \gamma_{rjt}^{(l)} + 2(\alpha_{rt}+1)}} \tag{3.43}
$$

In order to obtain the updating equation for δ_{ik}, the constraint optimization problem should be solved as follows:

$$
\begin{cases}
\delta_{ik}^{(l+1)} = \arg\max_{\delta_{rt}} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}[\gamma_{ijk}^{(l)}ln(\delta_{ik}) + (\phi_i/2-1)ln(\delta_{ik})] \\
s.t. \qquad \sum_{k=1}^{2}\delta_{ik}^{(l+1)} = 1
\end{cases}
\tag{3.44}
$$

In order to solve this constrained optimization problem, the *Lagrange multipliers* method is applied.

$$L(\delta_i, \lambda_{\delta i}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}[\gamma_{ijk}^{(l)} ln(\delta_{ik}) + (\phi_i/2-1)ln(\delta_{ik})] - \lambda_{\delta i}(\sum_{k=1}^{2}\delta_{ik} - 1) \tag{3.45}$$

In order to solve the above equation, the partial derivative of L with respect to δ_{rt} and $\lambda_{\delta r}$ is determined and then set to zero.

$$\begin{aligned}
\frac{\partial L(\delta_i, \lambda_{\delta i})}{\partial \delta_{rt}} &= \frac{\partial \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}[\gamma_{ijk}^{(l)} ln(\delta_{ik}) + (\phi_i/2-1)ln(\delta_{ik})]}{\partial \delta_{rt}} - \lambda_{\delta r} \\
&= \frac{\partial \sum_{j=1}^{m}[\gamma_{rjt}^{(l)} ln(\delta_{rt})] + (\phi_r/2-1)ln(\delta_{rt})}{\partial \delta_{rt}} - \lambda_{\delta r} \\
&= \frac{\sum_{j=1}^{m}\gamma_{rjt}^{(l)} + (\phi_r/2-1)}{\delta_{rt}} - \lambda_{\delta r} = 0
\end{aligned} \tag{3.46}$$

$$\begin{aligned}
\delta_{rt} &= \frac{\sum_{j=1}^{m}\gamma_{rjt}^{(l)} + \phi_r/2 - 1}{\lambda_{\delta r}} \\
\sum_{k=1}^{2}\delta_{rk} &= \frac{\sum_{k=1}^{2}\sum_{j=1}^{m}\gamma_{rjt}^{(l)} + \phi_r/2 - 1}{\lambda_{\delta r}}
\end{aligned} \tag{3.47}$$

$$\frac{\partial L(\delta_i, \lambda_{\delta i})}{\partial \lambda_{\delta r}} = \sum_{k=1}^{2}\delta_{rk} - 1 = 0 \tag{3.48}$$

Due to the property of the γ_{ijk} function, the following equation is obtained:

$$\sum_{k=1}^{2}\gamma_{ijk}^{(l)} = 1 \rightarrow \sum_{j=1}^{m}\sum_{k=1}^{2}\gamma_{ijk}^{(l)} = m \tag{3.49}$$

By substituting equations (3.48) and (3.49) into equation (3.47), it can be concluded that $1 = \frac{m+\phi_r/2-1}{\lambda_{\delta r}} \rightarrow \lambda_{\delta r} = m + \phi_r/2 - 1$, and as a result $\delta_{rt}^{(l+1)}$ would be

$$\delta_{rt}^{(l+1)} = \frac{\sum_{j=1}^{m} \gamma_{rjt}^{(l)} + \phi_r/2 - 1}{\lambda_{\delta r}} = \frac{\sum_{j=1}^{m} \gamma_{rjt}^{(l)} + \phi_r/2 - 1}{m + \phi_r/2 - 1} \tag{3.50}$$

Similarly for $\sigma_r^{(l+1)}$:

$$\begin{aligned}
\frac{\partial R(\Theta, \Theta^{(l)})}{\partial \sigma_r} &= \frac{\sum_{i=1}^{n} \sum_{k=1}^{2} (-\frac{(x_i^{(l)}-\mu_i)^2}{2\sigma_i^2} - (2\alpha_i + 3) ln\sigma_i - \frac{\beta_i}{\sigma_i^2} - 2(\alpha_{ik} + 1) ln\sigma_{ik} - \frac{\beta_{ik}}{\sigma_{ik}^2})}{\partial \sigma_r} \\
&= \frac{\sum_{i=1}^{n} (-\frac{(x_i^{(l)}-\mu_i)^2}{2\sigma_i^2} - (2\alpha_i + 3) ln\sigma_i - \frac{\beta_i}{\sigma_i^2})}{\partial \sigma_r} \\
&= \frac{\partial(-\frac{(x_r^{(l)}-\mu_r)^2}{2\sigma_r^2} - (2\alpha_r + 3) ln\sigma_r - \frac{\beta_r}{\sigma_r^2})}{\partial \sigma_r} \\
&= \frac{(x_r^{(l)} - \mu_r)^2}{\sigma_r^3} - \frac{(2\alpha_r + 3)}{\sigma_r} + \frac{2\beta_r}{\sigma_r^3} = 0
\end{aligned} \tag{3.51}$$

$$(2\alpha_r + 3)\sigma_r^2 - ((x_r^{(l)} - \mu_r)^2 + 2\beta_r) = 0 \tag{3.52}$$

The solution for $\sigma_r^{(l+1)}$ is:

$$\sigma_r^{(l+1)} = \sqrt{\frac{(x_r(l) - \mu_r)^2 + 2\beta_r}{2\alpha_r + 3}} \tag{3.53}$$

Since there are process model constraints, solve the constrained optimization problem for $\mathbf{x}^{(l+1)}$ must be solved.

$$\begin{cases}
x_i^{(l+1)} = argmax_{x_r} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{2} \gamma_{ijk}^{(l)} (ln\delta_{ik}^{(l+1)} - ln\sqrt{2\pi {\sigma_{ik}^{(l+1)}}^2} - \frac{(y_{ij} - x_i)^2}{2{\sigma_{ik}^{(l+1)}}^2}) \\
\qquad - \frac{(x_i - \mu_i^{(l+1)})^2}{2{\sigma_i^{(l+1)}}^2} \\
s.t. \qquad A\mathbf{x} = 0
\end{cases} \tag{3.54}$$

The Lagrangian function for the above constrained optimization problem is given by,

$$
\begin{aligned}
L(\mathbf{x},\lambda_x) &= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}[\gamma_{ijk}^{(l)}(ln\delta_{ik}^{(l+1)} - ln\sqrt{2\pi\sigma_{ik}^{(l+1)^2}} - \frac{(y_{ij}-x_i)^2}{2\sigma_{ik}^{(l+1)^2}}) - \frac{(x_i-\mu_i^{(l+1)})^2}{2\sigma_i^{(l+1)^2}}] \\
&+ \lambda_x^T A\mathbf{x} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}[\gamma_{ijk}^{(l)}(ln\delta_{ik}^{(l+1)} - ln\sqrt{2\pi\sigma_{ik}^{(l+1)^2}} - \frac{(y_{ij}-x_i)^2}{2\sigma_{ik}^{(l+1)^2}}) - \frac{(x_i-\mu_i^{(l+1)})^2}{2\sigma_i^{(l+1)^2}}] \\
&+ \begin{bmatrix}\lambda_{x1} & \lambda_{x2} & ... & \lambda_{xq}\end{bmatrix}\begin{bmatrix}a_{11} & a_{12} & ... & a_{1n}\\ a_{21} & a_{22} & ... & a_{2n}\\ ... & ... & ... & ...\\ a_{q1} & a_{q2} & ... & a_{qn}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ ...\\ x_n\end{bmatrix} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{2}[\gamma_{ijk}^{(l)}(ln\delta_{ik}^{(l+1)} - ln\sqrt{2\pi\sigma_{ik}^{(l+1)^2}} - \frac{(y_{ij}-x_i)^2}{2\sigma_{ik}^{(l+1)^2}}) - \frac{(x_i-\mu_i^{(l+1)})^2}{2\sigma_i^{(l+1)^2}}] \\
&+ \sum_{i=1}^{n}\sum_{z=1}^{q}\lambda_{xz}a_{zi}x_i
\end{aligned}
\tag{3.55}
$$

In order to obtain the update expression for the rectified estimates, the partial derivatives of the above equation with respect to x_r and λ_{xw} must be evaluated.

$$
\begin{aligned}
\frac{\partial L(\mathbf{x},\lambda_x)}{\partial x_r} &= \sum_{j=1}^{m}\sum_{k=1}^{2}[\gamma_{rjk}^{(l)}(\frac{y_{rj}-x_r}{\sigma_{rk}^{(l+1)^2}}) - \frac{x_r-\mu_r^{(l+1)}}{\sigma_r^{(l+1)^2}}] + \sum_{z=1}^{q}\lambda_{xz}a_{zr} \\
&= \sum_{j=1}^{m}[\gamma_{rj1}^{(l)}(\frac{y_{rj}-x_r}{\sigma_{r1}^{(l+1)^2}}) + \gamma_{rj2}^{(l)}(\frac{y_{rj}-x_r}{\sigma_{r2}^{(l+1)^2}})] - \frac{x_r-\mu_r^{(l+1)}}{\sigma_r^{(l+1)^2}} + \sum_{z=1}^{q}\lambda_{xz}a_{zr} \\
&= \sum_{j=1}^{m}[\frac{\gamma_{rj1}^{(l)}}{\sigma_{r1}^{(l+1)^2}} + \frac{\gamma_{rj2}^{(l)}}{\sigma_{r2}^{(l+1)^2}}](y_{rj}-x_r) - \frac{x_r-\mu_r^{(l+1)}}{\sigma_r^{(l+1)^2}} + \sum_{z=1}^{q}\lambda_{xz}a_{zr} = 0
\end{aligned}
\tag{3.56}
$$

$$
\frac{\partial L(\mathbf{x},\lambda_x)}{\partial \lambda_{xw}} = \sum_{i=1}^{n}a_{wi}x_i = 0 \tag{3.57}
$$

Consider $W_j^{(l+1)}$ and $Z^{(l+1)}$ matrices,

$$W_j^{(l+1)} = diag\Big(\frac{\gamma_{rj1}^{(l)}}{\sigma_{r1}^{(l+1)^2}} + \frac{\gamma_{rj2}^{(l)}}{\sigma_{r2}^{(l+1)^2}}\Big) \tag{3.58}$$

$$Z^{(l+1)} = diag\Big(\frac{-1}{\sigma_r^{(l+1)^2}}\Big) \tag{3.59}$$

where $diag_{(j=1,\ldots,m)}(a_j)$ represents the operator that converts an m-dimensional vector a into a diagonal matrix; Equation (3.56) can be simplified as,

$$\begin{aligned}
&\textstyle\sum_{j=1}^m W_j^{(l+1)}(\mathbf{y}_j^s - \mathbf{x}) + Z^{(l+1)}(\mathbf{x} - \mu^{(l+1)}) + A^T\lambda_x = 0 \\
&\textstyle\sum_{j=1}^m \Big(W_j^{(l+1)}\mathbf{y}_j^s - W_j^{(l+1)}\mathbf{x}\Big) + Z^{(l+1)}\mathbf{x} - Z^{(l+1)}\mu^{(l+1)} + A^T\lambda_x = 0 \\
&\textstyle\sum_{j=1}^m \Big(W_j^{(l+1)}\mathbf{y}_j^s\Big) - \Big(\sum_{j=1}^m W_j^{(l+1)}\Big)\mathbf{x} + Z^{(l+1)}\mathbf{x} - Z^{(l+1)}\mu^{(l+1)} + A^T\lambda_x = 0 \\
&\textstyle\sum_{j=1}^m \Big(W_j^{(l+1)}\mathbf{y}_j^s\Big) + \Big(Z^{(l+1)} - \sum_{j=1}^m W_j^{(l+1)}\Big)\mathbf{x} - Z^{(l+1)}\mu^{(l+1)} + A^T\lambda_x = 0 \\
&\textstyle\Big(Z^{(l+1)} - \sum_{j=1}^m W_j^{(l+1)}\Big)^{-1}\sum_{j=1}^m \Big(W_j^{(l+1)}\mathbf{y}_j^s\Big) + \mathbf{x} - \Big(Z^{(l+1)} - \sum_{j=1}^m W_j^{(l+1)}\Big)^{-1} Z^{(l+1)} \\
&\textstyle\mu^{(l+1)} + \Big(Z^{(l+1)} - \sum_{j=1}^m W_j^{(l+1)}\Big)^{-1} A^T\lambda_x = 0 \\
&\textstyle A\Big(Z^{(l+1)} - \sum_{j=1}^m W_j^{(l+1)}\Big)^{-1}\sum_{j=1}^m \Big(W_j^{(l+1)}\mathbf{y}_j^s\Big) - A\Big(Z^{(l+1)} - \sum_{j=1}^m W_j^{(l+1)}\Big)^{-1} Z^{(l+1)} \\
&\textstyle\mu^{(l+1)} + A\Big(Z^{(l+1)} - \sum_{j=1}^m W_j^{(l+1)}\Big)^{-1} A^T\lambda_x = 0
\end{aligned}$$

B and C are defined as,

$$\begin{aligned}
&\textstyle B^{(l+1)} = (Z^{(l+1)} - \sum_{j=1}^m W_j^{(l+1)})^{-1} \\
&\textstyle C^{(l+1)} = \sum_{j=1}^m (W_j^{(l+1)}\mathbf{y}_j^s)
\end{aligned}$$

Therefore,

$$AB^{(l+1)}C^{(l+1)} - AB^{(l+1)}Z^{(l+1)}\mu^{(l+1)} + AB^{(l+1)}A^T\lambda_x = 0 \tag{3.60}$$

$$\lambda_x = (AB^{(l+1)}A^T)^{-1}(AB^{(l+1)}Z^{(l+1)}\mu^{(l+1)} - AB^{(l+1)}C^{(l+1)}) \tag{3.61}$$

$$C^{(l+1)} + {B^{(l+1)}}^{-1}\mathbf{x} - A^T(AB^{(l+1)}A^T)^{-1}(AB^{(l+1)}Z^{(l+1)}\mu^{(l+1)} - AB^{(l+1)}C^{(l+1)}) = 0$$

$$\mathbf{x}^{(l+1)} = B^{(l+1)}(I - A^T(AB^{(l+1)}A^T)^{-1}AB^{(l+1)})(Z^{(l+1)}\mu^{(l+1)} - C^{(l+1)}) \qquad (3.62)$$

Table 3.2: Simultaneous Data Reconciliation and Gross Error Detection Using EM Algorithm

1	Input the raw measurements Y
2	Initialize the parameter $\Theta^{(l)}$
3	**E-step** Evaluate the R function (3.40) using current updated values of parameters
4	**M-step** Update the parameters using Equations (3.43), (3.50), (3.53), and(3.62).
5	Terminate on convergence. Otherwise, proceed to Step 3.

3.4.3 Performance Assessment

In order to illustrate the performance of the proposed method and to draw a comparison with the other approaches, several performance assessment tests have been used.

The AVTI (average number of type I) and OP (overall power) tests, which are utilized by Narasimhan and Mah[10] as well as Soderstrom, Himmelblau, and Edgar[17], are performed in order to judge the performance of the proposed method. The AVTI and OP tests are defined as follows:

$$AVTI = \frac{Number\ of\ unbiased\ variables\ wrongly\ identified}{Number\ of\ simulation\ trials} \qquad (3.63)$$

$$OP = \frac{Number\ of\ biased\ variables\ correctly\ identified}{Number\ of\ biased\ variables\ simulated} \qquad (3.64)$$

The AVTI test is computed for each simulation run separately; for this test, it does not matter whether there are gross errors or not in the data set. It gives the average number of gross errors mispredicted per each run. By contrast, the OP test is computed only for those simulations which include gross errors in the data set[10].

3.5 Simulation Examples

In order to illustrate the effectiveness and the performance of the proposed method, in this section, different simulation case studies are presented.

3.5.1 Example 1

In this example, we will illustrate equivalent sets of a network and compare the performance of MLE and MAP estimation. First, we determine all of the equivalent sets of the network, then show the difference between MLE and MAP estimation for one of the equivalent cases of the network for simultaneous data reconciliation and gross error detection using EM algorithm. Simulated data sets of a water flow network shown in Figure 3.3 are examined to validate the proposed method. The network includes four units as well as seven flow rates.

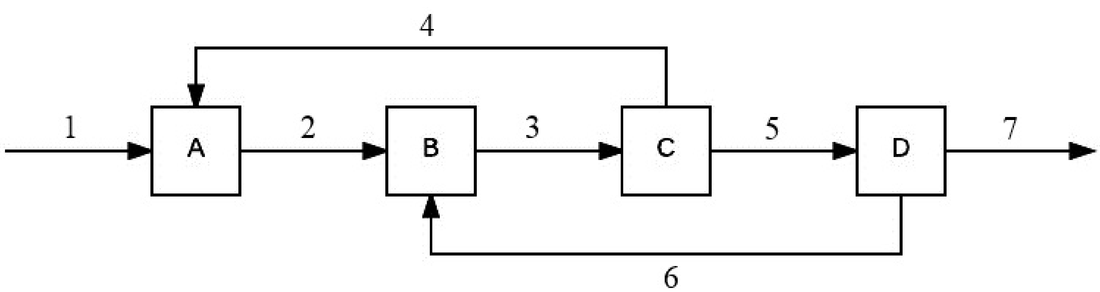

Figure 3.3: Diagram of process network

Based on the above flow diagram, the process model equations, which are the constraints of the problem, are given by:

$$x_1 - x_2 + x_4 = 0$$

$$x_2 - x_3 + x_6 = 0$$

$$x_3 - x_4 - x_5 = 0$$

$$x_5 + x_6 - x_7 = 0$$

According to the above equations, the process model coefficient matrix in $A\mathbf{x} = 0$ is given by,

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

In order to illustrate the performance of the proposed method in the presence of an equivalent set of gross errors, equivalent sets of gross errors of the system must be determined first. According to Theorem 1, the rank of matrix A is 4. Consider the presence of gross error of size (+2) in x_1, a gross error of size (-1) in x_3, a gross error of size (+3) in x_5, and a gross error of size (+1) in x_6. As illustrated in Case 1 of Table 3.3, these four gross errors can be modelled by gross errors in four other streams (Case 2). Case 3 shows that the four gross errors in Case 1 can be reduced to three gross errors since the gross error cardinality of x_3, x_5, and x_6 streams is 2, i.e., $\Gamma = 2$. Therefore, we can model any gross error in these three streams by two other gross errors. Case 4, also, shows this fact which is used to express errors in Case 2.

Table 3.3: Illustration of a set of gross error cardinality $\Gamma = 4$

		x_1	x_2	x_3	x_4	x_5	x_6	x_7
	Measurement	12	20	29	10	23	11	10
Case 1	Reconciled estimate	10	20	30	10	20	10	10
Candidates: x_1, x_3, x_5, x_6	Estimated bias	+2		-1		+3	+1	
Case 2	Reconciled estimate	12	19	30	7	23	11	12
Candidates: x_2, x_3, x_4, x_7	Estimated bias		+1	-1	+3			-2
Case 3	Reconciled estimate	10	20	31	10	21	11	10
Candidates: x_1, x_3, x_5	Estimated bias	+2		-2		+2		
Case 4	Reconciled estimate	12	20	31	8	23	11	12
Candidates: x_3, x_4, x_7	Estimated bias			-2	+2			-2

Assume x_2 and x_3 are biased, and the magnitude of gross errors in these two variables are +6 and +7, respectively (Case 1, Table 3.4). The gross error cardinality of these two variables is 2; adding another variable to this set, i.e., $\{x_2, x_3\}$, will not increase the gross error cardinality. The two gross errors in x_2 and x_3 can be modeled by two other variables by constructing a subset. x_4 is the only variable in this network by adding which the gross error cardinality of the mentioned set will not be changed. As expected, if the Maximum Likelihood Estimation (MLE)is used, incorrect, or very poor estimates, of these three variables will be obtained. By contrast, because prior information is used in the Maximum A Posteriori (MAP) framework, better results are obtained for the estimation. Let us first show what would be the expectation of MLE and MAP estimation, then compare the results of these two methods. As can be seen in Table 3.4, there are three different equivalent sets. The MLE approach should return the results of Case 3 as reconciled estimates. Since in the MLE estimation we use the mean values of the raw measurements as initial values, it converges to the first equivalent set of reconciled estimate in the iteration procedure, i.e., it converges to those results which need least adjustments. Even if we use the available prior information as an initial guess for MLE approach, we cannot obtain a true estimate of the state, and we will demonstrate this shortly.

Table 3.4: Illustration of a set of gross errors cardinality $\Gamma = 2$

		x_1	x_2	x_3	x_4	x_5	x_6	x_7
	Measurement	10	26	37	10	20	10	10
Case 1	Reconciled estimate	10	20	30	10	20	10	10
Candidates: x_2, x_3	Estimated bias		+6	+7				
Case 2	Reconciled estimate	10	27	37	17	20	10	10
Candidates: x_2, x_4	Estimated bias		-1		-7			
Case 3	Reconciled estimate	10	26	36	16	20	10	10
Candidates: x_3, x_4	Estimated bias			-1	+6			

Let us illustrate the performance of MLE and MAP estimation by comparing the performance of these two methods for this example. Suppose all the variables are measured and the vector of true values of the flow rates is set as $\mathbf{x} = \begin{bmatrix} 10 & 20 & 30 & 10 & 20 & 10 & 10 \end{bmatrix}^T$. The measurements are generated by adding a random noise with zero mean and 10% of true value variance. It is assumed that x_2 and x_3 contain gross errors with magnitudes of +6 and +7, respectively. Table 3.5 shows the performance of the proposed method by the results of the OP and the AVTI tests.

Table 3.5: Solution for MLE and MAP DR using EM algorithm for multiple gross errors in the presence of equivalent sets of gross errors

	Simulation runs	Biased	OP	AVTI
MLE	50	2	0	1
MAP estimation	50	2	1	0

From Table 3.5, it can be seen that the MLE approach cannot converge to the true values, and as expected, it converged to the results of Case 3 of Table 3.4. Because it wrongly estimated x_4 as a biased variable, and also could not correctly identify gross errors in x_2 and x_3, the results of the AVTI and the OP tests are 1 and 0, respectively. Figure 3.4 represents the histogram of the rectified values for 50 runs for MLE and MAP estimation.

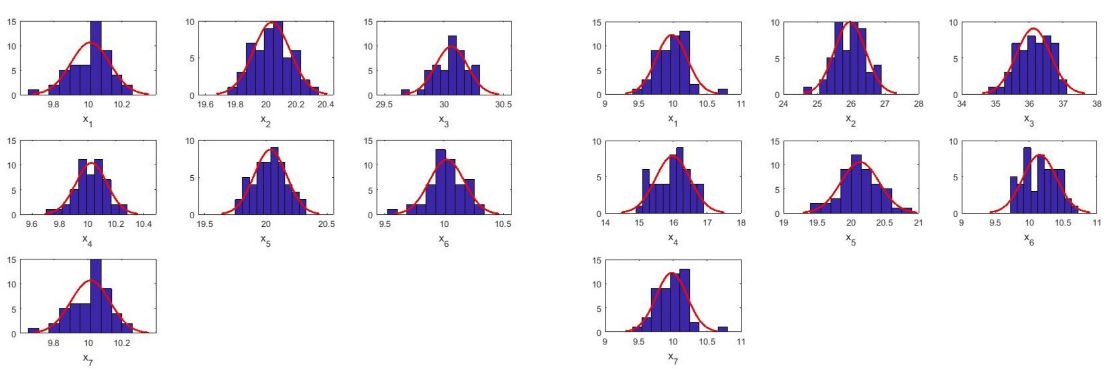

Figure 3.4: Histogram of developed code for 50 runs for MLE (right hand side) and MAP estimation (left hand side), when x_2 and x_3 are corrupted by gross errors with magnitude of 6 and 7, respectively.

Now, let us assume that we use the prior knowledge to initialize the state. In the M-LE framework, even if we set the initial values of the variable based on the prior information, it cannot handle the equivalent sets of gross errors. For instance, in the current example, where x_2 and x_3 are corrupted by gross errors. Even if we consider $\mathbf{x}^{(\mathbf{0})} = \begin{bmatrix} 10 & 20 & 30 & 10 & 20 & 10 & 10 \end{bmatrix}^T$ as a set of initial values, the MLE method would not converge to these initial values. Although in the first iteration it forces measurements to get closer to the true values, it is very likely that the MLE approach converges to other equivalent sets. From Table 3.4, it can be seen that there are two equivalent sets between the raw measurements and the true values. Therefore, in the iteration procedure, the estimated values can converge to one of these sets. To make it clear, we plotted the results of MLE method when we use true values as our initial guesses for the states in Figure 3.5.

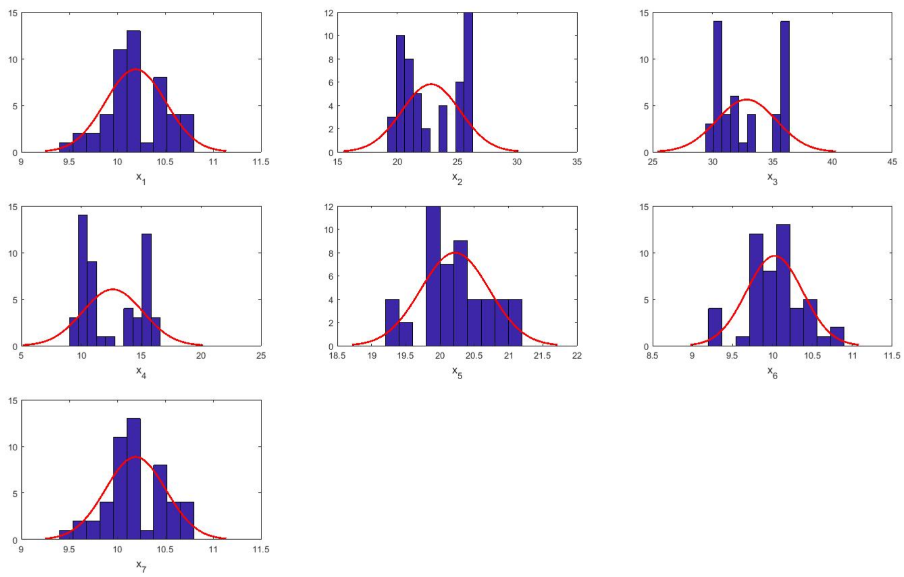

Figure 3.5: Histogram of developed code for 50 runs for MLE while the initial values for the state are true value, when x_2 and x_3 are corrupted by gross errors with magnitude of 6 and 7, respectively.

As can be seen in the Figure 3.5, the MLE method has problem when there are equivalent sets of gross errors. The mean value of the estimated reconciled values for 50 runs with MLE approach is $\mathbf{x} = \begin{bmatrix} 10.1197 & 21.9289 & 32.0910 & 11.8093 & 20.2817 & 10.162110.1197 \end{bmatrix}^T$. Table 3.6 shows the true and rectified values, as well as the estimated parameters of the MAP estimation.

Table 3.6: MAP results for multiple gross errors while there are equivalent sets.

Variables	x	$\hat{x}$	$\hat{\delta}_{i1}$	$\hat{\sigma}_{i1}$	$\hat{\sigma}_{i2}$
x_1	10	9.9390	1.00	0.9097	0.7071
x_2	20	19.9624	0.00	0.7071	5.7537
x_3	30	30.0316	0.00	0.7071	6.5693
x_4	10	10.0234	1.00	0.9568	0.7071
x_5	20	20.0082	1.00	1.5565	0.7071
x_6	10	10.0692	1.00	1.0534	0.7071
x_7	10	9.9390	1.00	0.9960	0.7071

In Table 3.6, $\hat{\sigma}_{i2} = 0.7071$ for those variables with $\hat{\delta}_{i1} = 1$, or in other words, if $\hat{\delta}_{i2} = 0$ then $\hat{\sigma}_{i2} = 0.7071$. Whenever there is an unbiased set of measurements, this set follows a single distribution, not a mixture one. Mathematically speaking,

$$\epsilon_i \sim \delta_{i1} N(0, \sigma_{i1}^2) + \delta_{i2} N(0, \sigma_{i2}^2) \ if \ \delta_{i2} = 0 \rightarrow \epsilon_i \sim N(0, \sigma_{i1}^2)$$

The reason that the same value is obtained for their $\hat{\sigma}_{i2}$ is that in the γ_{ij2} equation when $\hat{\delta}_{i2} = 0$ we have,

$$\gamma_{ij2} = \frac{\frac{\delta_{i2}^{(l)}}{\sqrt{2\pi\sigma_{i2}^{2\,(l)}}} exp(-\frac{(y_{ij}-x_i^{(l)})^2}{2\sigma_{i2}^{2\,(l)}})}{\sum_{t=1}^{2} \frac{\delta_{it}^{(l)}}{\sqrt{2\pi\sigma_{it}^{2\,(l)}}} exp(-\frac{(y_{ij}-x_i^{(l)})^2}{2\sigma_{it}^{2\,(l)}})} = \frac{0}{\frac{1}{\sqrt{2\pi\sigma_{i1}^{2\,(l)}}} exp(-\frac{(y_{ij}-x_i^{(l)})^2}{2\sigma_{i1}^{2\,(l)}})} = 0$$

so, in this case, the updating equation for σ_{i2} is given by,

$$\sigma_{i2}^{(l+1)} = \sqrt{\frac{\sum_j^m \gamma_{ij2}^{(l)} (y_{ij} - x_i^{(l)})^2 + 2\beta_{i2}}{\sum_j^m \gamma_{ij2}^{(l)} + 2(\alpha_{i2} + 1)}} = \sqrt{\frac{\beta_{i2}}{\alpha_{i2} + 0.5}}$$

Therefore, it is a function of some constant values, β_{i2} and α_{i2}, which result in the same value (0.7071) for the σ_{i2} subsequently. There is also the same reason for $\hat{\sigma}_{i1}$ when $\hat{\delta}_{i1} = 0$, i.e., if $\hat{\delta}_{i1} = 0$ then $\hat{\sigma}_{i1} = 0.7071$.

3.5.2 Example 2 - Comparing with Literature

In this example, we will illustrate the performance of the proposed method for different cases with and without equivalent sets of gross errors by selecting any possible combinations of two biased variables and we compare the results with previously proposed methods in the literature. Consider the well-known example[14,30] including a recycle system with four units and seven streams as shown in Figure(3.6). In this example, the vector of true values of the flow rates is set as $\mathbf{x} = \begin{bmatrix} 5 & 15 & 15 & 5 & 10 & 5 & 5 \end{bmatrix}^T$ and the standard deviations of the flow rates are taken as 2.5% of the true flow rate values. As in previous publications, all possible combinations of two measurement biases are simulated. The magnitude of the gross errors is fixed at 7 and 4 standard deviations for the corresponding flow rates.

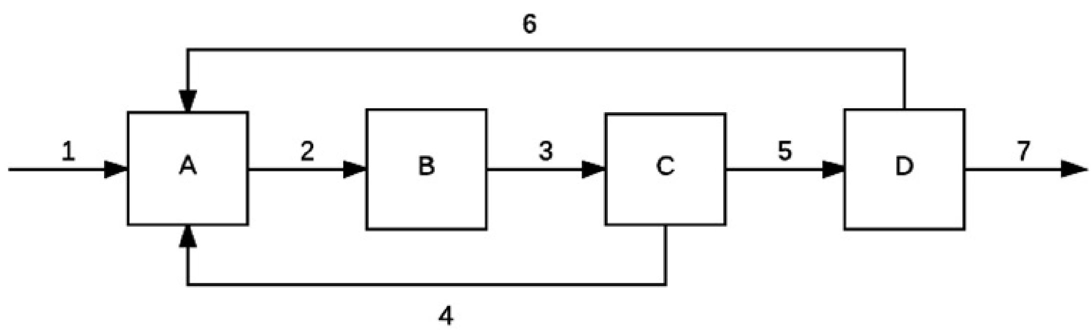

Figure 3.6: Diagram of steam metering process network

Table 3.7 compares the results of OP and AVTI tests of the proposed method to the Modified Iterative Measurement Test (MIMT)[7], the Generalized Likelihood Ratio Test Method (GLR)[10], and the Simultaneous Estimation of Gross Error Method (SEGE)[30].

Table 3.7: Performance results for MIMT, GLR, SEGE, and Proposed Method

	MIMT		GLR		SEGE		Proposed Method	
Streams	OP	AVTI	OP	AVTI	OP	AVTI	OP	AVTI
1-2	0.969	0.152	0.971	0.167	0.996	0.008	1.000	0.000
1-3	0.969	0.153	0.972	0.168	1.000	0.000	1.000	0.000
1-4	0.974	0.070	0.973	0.063	0.974	0.049	1.000	0.000
1-5	0.035	2.088	0.145	1.950	0.704	0.615	1.000	0.000
1-6	0.500	1.048	0.993	0.143	0.997	1.000	1.000	0.000
1-7	0.821	0.403	0.504	1.074	0.997	1.000	1.000	0.000
2-3	0.500	1.045	0.997	1.041	0.999	1.000	1.000	0.000
2-4	0.501	0.963	0.500	0.961	0.958	0.948	1.000	0.000
2-5	1.000	0.043	0.999	0.089	0.999	0.000	1.000	0.000
2-6	0.989	0.050	0.966	0.130	0.987	0.027	1.000	0.000
2-7	0.999	0.046	0.999	0.057	0.999	0.002	1.000	0.000
3-4	0.501	0.965	0.500	0.964	0.960	0.951	1.000	0.000
3-5	0.999	0.044	1.000	0.086	1.000	0.000	1.000	0.000
3-6	0.988	0.051	0.967	0.125	0.987	0.027	1.000	0.000
3-7	0.999	0.049	0.999	0.054	0.999	0.001	1.000	0.000
4-5	0.500	1.037	0.999	0.077	0.999	1.000	1.000	0.000
4-6	0.923	0.189	0.577	1.063	0.998	0.999	1.000	0.000
4-7	0.999	0.035	0.998	0.067	0.998	0.004	1.000	0.000
5-6	0.978	0.062	0.500	1.026	0.977	0.996	1.000	0.000
5-7	0.996	0.051	0.996	0.144	0.997	0.006	1.000	0.000
6-7	1.000	0.045	0.886	0.820	1.000	1.000	1.000	0.000

3.6 Conclusion

In this chapter, data rectification problem in the maximum a posteriori framework was proposed. The contaminated Gaussian distribution model characterized the error distribution of different noise mode and therefore, the resulting formulation involved hidden variable. As a result, the EM method was used. It is important to notice that the proposed approach could handle the cases where there are sets of gross errors which lead to multiple sets of true value estimates, which is known as an equivalent set of gross errors, because of using prior information. Moreover, the parameters of the error distribution model were not preset and could be determined as part of the solution. Moreover, the proposed method was capable of reporting the magnitude of gross errors. Furthermore, it detected partial gross errors, so it was not required to assume that gross errors exist in the entire data set. Several case studies were presented to demonstrate the different features of the algorithm.

Chapter 4

Maximum A Posteriori Framework for Data Rectification Using Contaminated Gaussian Mixture Distribution with Multiple Operating Regions

Process measurements collected from chemical or other industrial plant operations contain significant information which is used in process optimization, control, and identification. These raw measurements, however, are mostly corrupted by errors including random errors and gross errors, which could lead to biases in subsequent calculations. Therefore, the p-resence of gross errors causes unreliable solutions for control and optimization problems. In reality, a system works under different circumstances, i.e., there are multiple operating regions. The main focus of this chapter is rectifying a data set which includes different operating regions. Distinguishing between the occurrence of gross errors and changing the operating mode is the challenging part of this work. In other words, when either a measurement is corrupted by gross errors or the operating mode is changed, the mean value of the measurement is changed subsequently. Therefore, first, the operating mode of each sample point must be determined, and then, the noise mode of each measurement in the sample point, as a gross error identifier, has to be specified accordingly. To this end, the data set is divided into several clusters based on the number of operating modes. Then, the same operation, i.e., data rectification, is applied to each operating mode. In order to remove both kinds of measurement errors in different operating regions, two hidden variables are introduced - one for identifying the operating mode and one for identifying the noise mode. Based on these assumptions, the Maximum a Posteriori (MAP) framework is applied, and the subsequent optimization problem is solved using Expectation Maximization (EM) algorithm due to the presence of hidden variables. It should be noted that the proposed approach distinguishes between the operating modes in different clusters. As a result, the future model would be capable of identifying operating modes and switching between them. This approach does not require presetting the parameters of the error distribution model, rather they are determined as a part of the solution. Several case studies are presented to demonstrate the effectiveness of the proposed approach.

4.1 Introduction

The intense competitive nature of the world market, the growing importance of producing quality products, and the increasingly relevant necessity to consider safety and environmental issues and their regulations have magnified the need for fast and low-cost changes in chemical processes to enhance their performance[30]. Any possible changes and modifications in a system to control, optimize, and evaluate the behavior of the process, or to improve the performance of the system requires clear understanding and knowledge of its actual state. This information is obtained by processing a data set - collecting it, ameliorating its accuracy, and storing/using it for further analysis[30].

It is necessary to improve the accuracy of the collected raw measurements from different process plant operations such that the most reliable and the highest performance is attained. It is important to note that raw measurements inherently contain inaccurate information due to imperfect instruments[30,33]. Two types of errors can be identified in plant data; random errors, which are assumed to be independent of each other and normally distributed with zero mean and small variance due to the normal fluctuation of the process or the variation in instrument operation, and gross errors, or systematic errors, which occur occasionally because of incorrect calibration or malfunctioning of instruments, process leaks, and so on[30,41].

A failure in one of the instruments can lead to a deviation in process variables beyond acceptable limits unless the failure detection and correction is performed in an appropriate time. It is the purpose of data rectification to estimate the true states and detect any instrument faults which could seriously degrade the performance of a system[42]. Several rectification methods have been proposed in the literature based on the assumptions made for gross error detection and data reconciliation. Steady state data reconciliation and gross error detection methods can be categorized into three parts. These approaches have been discussed in detail in Chapter 2. The current chapter belongs to the third category.

So far, several approaches have been proposed for data rectification when there is no change in the process operating mode. However, in reality, the process operating condition changes over time, and the problem of data rectification when there are multiple operating regions has not been thoroughly investigated. This chapter extends the proposed method in Chapter 3 by introducing another hidden variable in order to determine the operating mode. Therefore, the main contribution of this chapter is proposing a method to solve the simultaneous gross error detection and data reconciliation problem, and estimating distribution parameters based on the historical data when there are more than one operating regions. It should be noted that the proposed approach discerns between the operating regions and therefore, the future model is capable of identifying and switching between the operating modes.

Here, we tackle the issue of distinguishing the change in operating modes from the occurrence of gross errors. In other words, there are two scenarios when the mean values of the measurements change: (1) the operating mode is switched to another region, and the mean values of the measurements are changed accordingly, and (2) some of the variables are partially corrupted by gross errors which lead to the changes in the mean values. Therefore, a vital step in this work is to identify the reason for changes in the mean values of the measurements. To this end, the operating mode of each sample point has to be identified, followed by determining whether each measurement in the sample point contains a gross error or not. As a result, the data set has to be divided into several clusters based on the number of operating regions. Then, the same operation of data rectification is applied to each operating mode. Hence, in order to remove both types of measurement errors in different operating regions, two hidden variables are introduced - one for identifying operating mode and one for identifying noise mode. In section 4.3, a detailed explanation for distinguishing between operating mode and noise mode is provided.

The rest of the Chapter is organized as follows: Section 4.2 provides the background and a general introduction to data rectification problem with multiple operating regions. In section 4.3, the problem of the data reconciliation using Gaussian mixture distribution as a measurement noise model is presented, and then in section 4.4, an Expectation Maximization algorithm is presented such that the parameters are estimated along with the rectified estimates. In section 4.5, several examples are given to evaluate the performance of the proposed approach. Finally, conclusions to this chapter are provided in section 4.6.

4.2 Data Rectification with Multiple Operating Regions

Data rectification is a procedure of processing raw measurements to remove errors, both random and gross errors, from error-corrupted measurements, and estimating true states based on the available process models. Due to the presence of errors, raw measurements can be modelled shown in Equation (4.1),

$$y_{zi} = x_{zi} + \epsilon_{zi} \tag{4.1}$$

where y_{zi} is the measurement, and x_{zi} is the true value of the i^{th} variable in the z^{th} operating mode. Here, i and z are variable and operating mode indices, respectively. Given the process model of the form, $\mathbf{f}(\mathbf{x}) = 0$, and assuming the measurement error ϵ_{zi} is independent of the true states, x_{zi}, the general formulation of the data rectification problem is given by:

$$\arg\max_{\mathbf{x}} \quad P\{\mathbf{x}|\mathbf{y}\} \tag{4.2}$$

$$\text{s.t.} \quad \mathbf{f}(\mathbf{x}) = 0 \tag{4.3}$$

$$\mathbf{g}(\mathbf{x}) \leq 0 \tag{4.4}$$

where the objective function expresses the probability of the true state given the raw measurements, and the inequality constraint is given by $\mathbf{g}(\mathbf{x})$. Using Bayes rule and writing the above posterior probability in terms of the probability of the raw measurements given the state, $P\{\mathbf{y}|\mathbf{x}\}$, the prior probability of the states, $P\{\mathbf{x}\}$, and the probability of the raw measurements, $P\{\mathbf{y}\}$, the last term can be removed from the objective function since it is independent of the states. Therefore, the objective function is simplified to the following expression,

$$\arg\max_{\mathbf{x}} \quad P\{\mathbf{x}|\mathbf{y}\} = \arg\max_{\mathbf{x}} \quad P\{\mathbf{y}|\mathbf{x}\}P\{\mathbf{x}\} \tag{4.5}$$

Further, assuming a uniform distribution for the prior probability, i.e., $P\{\mathbf{x}\} = 1$, the rectification problem is converted to the reconciliation problem. Further details of data rectification and reconciliation, equivalent sets of gross errors, and examples can be found in Chapter 3 or in the work which has been done by Bagajewicz et al.[34]. In this chapter, we solve the data rectification problem for a system which works in different operating modes. To this end, the objective function in Equation (4.5) has to be expanded such that it represents different operating regions, or it should be a function of operating regions as well. Equation (4.6) describes the objective function for this chapter,

$$\begin{aligned} \arg\max_{\mathbf{x}} \quad P\{\mathbf{x}|\mathbf{y}\} &= \arg\max_{\mathbf{x}} \quad P\{\mathbf{y}|\mathbf{x}\}P\{\mathbf{x}\} \\ &= \arg\max_{\mathbf{x}} \quad \sum_{z=1}^{q} P\{\mathbf{y_z}|\mathbf{x_z}, I = z\}P\{\mathbf{x_z}|I = z\}P\{I = z\} \end{aligned} \tag{4.6}$$

where, I denotes the hidden operating mode variable. In Section 4.4, the proposed method is presented in detail.

4.3 Problem Statement

Let Y_z denote the data matrix of raw measurements for each operating mode as follows:

$$Y_z = \begin{bmatrix} y_{z11} & y_{z12} & \cdots & y_{z1m} \\ y_{z21} & y_{z22} & \cdots & y_{z2m} \\ \cdots & \cdots & \cdots & \cdots \\ y_{zn1} & y_{zn2} & \cdots & y_{znm} \end{bmatrix} = \begin{bmatrix} \mathbf{y^v_{z1}} \\ \mathbf{y^v_{z2}} \\ \cdots \\ \mathbf{y^v_{zn}} \end{bmatrix} = \begin{bmatrix} \mathbf{y^s_{z1}} & \mathbf{y^s_{z2}} & \cdots & \mathbf{y^s_{zm}} \end{bmatrix} \tag{4.7}$$

where superscripts s and v refer to sample points and variables, respectively. The i^{th} row of the Y_z matrix represents the variable i measured at different sampling instants which belongs to z^{th} operating mode, $z = \{1, 2, ..., q\}$, and the j^{th} column represents the set of n variables measured at j^{th} sampling instant. Since each sample point belongs to an operating point, the raw measurements can be modelled according to sampling instants. Hence, the measurement model for all variables is given by,

$$\mathbf{y^s_{zj}} = \mathbf{x_z} + \epsilon_z \tag{4.8}$$

where $\mathbf{x_z}$ for each operating mode is defined as,

$$\mathbf{x_z} = \begin{bmatrix} x_{z1} \\ x_{z2} \\ \cdots \\ x_{zn} \end{bmatrix} \tag{4.9}$$

It should be noted that the measurement model given by Equation (4.8) is the vector form of the measurement model presented in Equation (4.1) at the j^{th} time instant of the z^{th} operating mode. Since true values are unknown and there is uncertain information of x_{zi}, it is assumed in Equation (4.1) this uncertain information follows a Gaussian distribution, i.e., $x_{zi} \sim N(\mu_{zi}, \sigma^2_{zi})$.

The second term in the right-hand-side of the Equation (4.1), ϵ_{zi}, follows a mixture Gaussian distribution with two modes; the first mode with small variance accounts for the random errors and the second one with large variance accounts for the presence of gross errors. Therefore, the raw measurement error distribution for the i^{th} variable of the z^{th} operating mode can be expressed in the following form:

$$\epsilon_{zi} \sim \delta_{zi1} N(0, \sigma^2_{zi1}) + \delta_{zi2} N(0, \sigma^2_{zi2}) \tag{4.10}$$

such that $\delta_{zi1} + \delta_{zi2} = 1$. Since both the sensor modes are mutually exclusive, the sensor model for a single measurement set can be expressed as,

$$P\{\mathbf{x_z}|\mathbf{y^s_{zj}}\} \propto P\{\mathbf{y^s_{zj}}|\mathbf{x_z}\}P\{\mathbf{x_z}\} = \sum_{k=1}^{2} P\{\mathbf{y^s_{zj}}|\mathbf{x_z}, I_{zij} = k\}P\{I_{zij} = k\}P\{\mathbf{x_z}\} \tag{4.11}$$

where I_{zij} denotes the hidden noise mode of variable i at the j^{th} sample point of the z^{th} operating mode which can be in either normal or abnormal mode. Using the above noise model, the log of the posterior probability of m measurements set can be expressed as:

$$\begin{aligned} \arg\max_{\mathbf{x}} \quad lnP\{\mathbf{x}|Y\} &= ln(\sum_{z=1}^{q} P\{Y_z|\mathbf{x_z}, I_j = z\}P\{I_j = z\}P\{\mathbf{x_z}\}) \\ &= ln(\sum_{z=1}^{q}\sum_{k=1}^{2} P\{Y_z|\mathbf{x_z}, I'_{zij} = k, I_j = z\}P\{I'_{zij} = k|I_j = z\} \\ &P\{I_j = z\}P\{\mathbf{x_z}\}) \end{aligned} \tag{4.12}$$

Assuming the measurements are independent, the above equation can be expanded as follows:

$$\begin{aligned} \arg\max_{\mathbf{x}} \quad lnP\{\mathbf{x}|Y\} = \prod_{i=1}^{n}\prod_{j=1}^{m} ln(\sum_{z=1}^{q}\sum_{k=1}^{2} P\{y_{zij}|x_{zi}, I'_{zij} = k, I_j = z\} \\ P\{I'_{zij} = k|I_j = z\}P\{I_j = z\}P\{x_{zi}\}) \end{aligned} \tag{4.13}$$

where $p_{zijk} = P\{y_{zij}|x_{zi}, I_{zij} = k, I_j = z\}$ is defined as follows,

$$p_{zijk}\{y_{zij}|x_{zi}, \sigma_{zik}\} = \frac{1}{\sqrt{2\pi}\sigma_{zik}} exp(-\frac{(y_{zij} - x_{zi})^2}{2\sigma^2_{zik}}) \quad k = 1, 2 \quad z = 1, ..., q \tag{4.14}$$

and $P\{I'_{zij} = k|I_j = z\} = \delta_{zik}$ and $P\{I_j = z\} = \eta_{zj}$ are the probability of occurrence of noise mode k and the probability of occurrence of the operating mode z, respectively. Prior works which used the Gaussian mixture model to characterize the noise distribution neither considered multiple operating modes nor used prior information in their objective functions to handle a multi-operating mode process and equivalent sets of gross errors, respectively. Moreover, some of previous studies used the pre-specified values of noise distribution parameters $\theta_{zik} = \{\delta_{zik}, \sigma^2_{zik}\}$ for both of the noise modes[1,20,23]. However, presetting the noise distribution parameters might lead to inaccurate results for the rectified estimates. Therefore, the objective of this work is to determine the noise distribution parameters and obtain the rectified estimates while the system works under different operating modes. Now, the simultaneous data reconciliation and gross error detection problem for linear steady state processes can be stated as follows:

$$\begin{aligned} \max_{\mathbf{x_z}, \theta_{zik}} \quad lnP\{\mathbf{x}|Y\} = \prod_{i=1}^{n}\prod_{j=1}^{m} ln(\sum_{z=1}^{q}\sum_{k=1}^{2} p_{zijk}\{y_{zij}|x_{zi}, \theta_{zik}, I'_{zij} = k, I_j = z\} \\ P\{I'_{zij} = k|\theta_{zik}, I_j = z\}P\{I_j = z|\theta_{zik}\}P\{x_{zi}\}) \\ s.t. \quad A\mathbf{x_z} = \mathbf{0} \end{aligned} \tag{4.15}$$

From Equation (4.15), it can be seen that the noise model identity probability as well as operating mode probability cannot be obtained explicitly, and as a result the maximum a posteriori estimation problem cannot be solved directly. Therefore, the Expectation Maximization (EM) algorithm is applied to solve this issue in order to obtain the distribution parameter values for each variable for each operating mode. In other words, all of the model distribution parameters along with the rectified values are estimated simultaneously using the EM algorithm.

In this work, we assumed that the prior distribution variances for x_{zi} and ϵ_{zi} follow the Inverse Gamma distribution which is the conjugate prior for the Gaussian and the Gaussian mixture likelihood function[40], and therefore:

$$\sigma^2_{zi} \sim IG(\alpha_{zi}, \beta_{zi}) \tag{4.16}$$

$$\sigma^2_{zik} \sim IG(\alpha_{zik}, \beta_{zik}) \tag{4.17}$$

Hence, their probability density functions are given by,

$$P(\sigma^2_{zi}|\alpha_{zi}, \beta_{zi}) = \frac{\beta_{zi}^{\alpha_{zi}}}{\Gamma(\alpha_{zi})}(\sigma^2_{zi})^{-\alpha_{zi}-1} exp(-\frac{\beta_{zi}}{\sigma^2_{zi}}) \tag{4.18}$$

$$P(\sigma^2_{zik}|\alpha_{zik}, \beta_{zik}) = \frac{\beta_{zik}^{\alpha_{zik}}}{\Gamma(\alpha_{zik})}(\sigma^2_{zik})^{-\alpha_{zik}-1} exp(-\frac{\beta_{zik}}{\sigma^2_{zik}}) \tag{4.19}$$

Further, each component of the noise model follows a Dirichlet(also known as multivariate Beta) distribution[40], and for the situation where there are only two modes for mixture Gaussian they follow Beta distribution, so Beta distribution is a prior for the probability of occurrence of each noise mode, δ_{ik}:

$$P(\delta_{i1}, \delta_{i2}|\phi_i) = \frac{\Gamma(\phi_i)}{\Gamma(\phi_i/2)^2} \prod_{j=1}^{2} \delta_{ik}^{\phi_i/2-1} \tag{4.20}$$

Therefore,

$$P(\delta_{ik}|\phi_i) \propto \delta_{ik}^{\phi_i/2-1} \tag{4.21}$$

If there is no a priori information available for the hyperparameters, a uniform distribution can be applied to express noninformative priors[19].

4.4 Proposed Method and Algorithm

In this section, the general formulation of the Expectation Maximization method in a maximum a posteriori framework is presented, and then Single Measurement Rectification is applied in that framework to solve the simultaneous data reconciliation and gross error detection problem along with parameter estimation while the process operating point changes.

4.4.1 General Formulation of EM Algorithm for The Maximum A Posteriori Estimation

In general, the EM algorithm is used to solve a maximum likelihood problem with hidden variables, but it can be modified to produce the MAP estimation[26]. If the complete data set is denoted by $\{C_{obs}, C_{mis}\}$, where C_{obs} refers to observed data and C_{mis} denotes hidden variables, the objective to find the MAP estimate of the parameter Θ is mathematically given by:

$$\hat{\Theta} = arg\ max_{\Theta}\ P(\Theta|C_{obs}, C_{mis}) \tag{4.22}$$

Using Bayes rule,

$$P(\Theta|C_{obs}, C_{mis}) = \frac{P(C_{obs}, C_{mis}|\Theta)\ P(\Theta)}{P(C_{obs}, C_{mis})} \tag{4.23}$$

Since the denominator of the above equation is independent of the parameters,

$$P(\Theta|C_{obs}, C_{mis}) \propto P(C_{obs}, C_{mis}|\Theta)\ P(\Theta) \tag{4.24}$$

Therefore,

$$\hat{\Theta} = arg\ max_{\Theta}\ P(C_{obs}, C_{mis}|\Theta)\ P(\Theta) \tag{4.25}$$

Since the formulation involves hidden variables, the EM algorithm is typically used. It is an iterative two step algorithm where in the first step, also known as the Expectation or E-step, the conditional expectation of hidden variables (R function) is obtained and in the second step, also known as the Maximization or M-step, the parameters are optimized.

4.4.1.1 E-step

In the E-step, the hidden variable is substituted by its conditional expectation given observed data, using the current fit for the parameters, i.e., $E_{C_{mis}|C_{obs},\Theta^{(l)}}$. Therefore, the R function is defined as

$$
\begin{aligned}
R(\Theta, \Theta^{(l)}) &= E_{C_{mis}|C_{obs},\Theta^{(l)}}[ln\ P(C_{mis}, C_{obs}|\Theta)] + ln\ P(\Theta) \\
&= \int P(C_{mis}|C_{obs}, \Theta^{(l)})[ln\ P(C_{mis}, C_{obs}|\Theta)]\ dC_{mis} + ln\ P(\Theta)
\end{aligned}
\tag{4.26}
$$

It can be seen that in the E-step the first term is the Q function in the MLE framework, so

$$
R(\Theta, \Theta^{(l)}) = Q(\Theta, \Theta^{(l)}) + ln\ P(\Theta) \tag{4.27}
$$

4.4.1.2 M-step

After obtaining the R function in the E-step, the M-step requires the maximization of $R(\Theta, \Theta^{(l)})$ with respect to Θ over the parameter space.

$$
\Theta^{(l+1)} = arg\ max_{\Theta}\ R(\Theta, \Theta^{(l)}) \tag{4.28}
$$

The R function in the E-step is evaluated with the updated parameter values $\Theta^{(l+1)}$ and the procedure is repeated until convergence.

4.4.2 Data Rectification based on MAP Estimation Using EM Algorithm while There Are Different Operating Regions

This work attempts to develop a method for data rectification for a single measurement in the presence of random errors and gross errors while there are multiple changes in process operating point. To this end, a model is built for each operating point beforehand, then the best fit model according to the available single measurement is selected, and finally, the data rectification is performed. The assumed distribution model for the noise of historical measurements for each operating mode is the mixture Gaussian, i.e., $\epsilon_{zi}^{v} \sim \delta_{zi1} N(0, \sigma_{zi1}^2) +$

$\delta_{zi2}N(0, \sigma^2_{zi2})$. In this section, the detailed solution is presented.

4.4.2.1 E-step

In the present mixture Gaussian distribution, the observed data is Y and the hidden variables are the operating mode and the noise mode for each operating point, i.e., $\mathbf{I} = \{I, I'\}$, where I and I' refer to the operating mode identity and noise mode identity, respectively. Each sample point corresponds to an operating mode, and each measurement in a sample point is prone to have gross error, so $I_j = \{1, 2, ..., q\}$ and $I'_{zij} = \{1, 2\}$. Parameters which have to be estimated are denoted as Θ, which contains rectified variables besides noise distribution parameters. Hence, the E-step in the EM Algorithm, in which the R function is derived, can be illustrated as,

$$
\begin{aligned}
R(\Theta, \Theta^{(l)}) &= Q(\Theta, \Theta^{(l)}) + lnP(\Theta) \\
&= E_{\mathbf{I}|Y,\Theta^{(l)}}[lnP(Y, \mathbf{I}|\Theta)] + lnP(\Theta) \\
&= \int P(\mathbf{I}|Y, \Theta^{(l)})\ d\mathbf{I}\ [lnP(Y, \mathbf{I}|\Theta)] + lnP(\Theta) \\
&= \int\int P(I', I|Y, \Theta^{(l)})\ dI'\ dI\ [lnP(Y, I, I'|\Theta)] + lnP(\Theta) \\
&= \int\int P(I'|I, Y, \Theta^{(l)})P(I|Y, \Theta^{(l)})\ dI'\ dI\ [lnP(Y, I, I'|\Theta)] + lnP(\Theta) \\
&= \int P(I|Y, \Theta^{(l)})\ dI \int P(I'|I, Y, \Theta^{(l)})\ dI'\ [lnP(Y, I'|I, \Theta) + lnP(I|\Theta)] + lnP(\Theta) \\
&= \int P(I|Y, \Theta^{(l)})\ dI \int P(I'|I, Y, \Theta^{(l)})\ dI'\ [lnP(Y, I'|I, \Theta) + lnP(I|\Theta)] \\
&+ ln\prod_{i=1}^{n} P(\theta_i) \\
&= \sum_{j=1}^{m} P(I_j|\mathbf{y}_j^s, \Theta^{(l)}) \sum_{i=1}^{n} P(I'_{zij}|I_j, y_{ij}, \theta_i^{(l)})\ [lnP(y_{ij}, I'_{zij}|I_j, \theta_i) + lnP(I_j|\Theta)] \\
&+ \sum_{i=1}^{n} lnP(\theta_i) \\
&= \sum_{z=1}^{q}\sum_{j=1}^{m} P(I_j = z|\mathbf{y}_j^s, \Theta_z^{(l)}) \sum_{i=1}^{n}\sum_{k=1}^{2} P(I'_{zij} = k|I_j = z, y_{zij}, \theta_{zik}^{(l)}) \\
&[lnP(y_{zij}, I'_{zij} = k|I_j = z, \theta_{zik}) + lnP(I_j = z|\Theta_z)] + \sum_{z=1}^{q}\sum_{i=1}^{n}\sum_{k=1}^{2} lnP(\theta_{zik}) \\
&= \sum_{z=1}^{q}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{2} P(I_j = z|\mathbf{y}_j^s, \Theta_z^{(l)})P(I'_{zij} = k|I_j = z, y_{zij}, \theta_{zik}^{(l)}) \\
&[lnP(y_{zij}|I'_{zij} = k, I_j = z, \theta_{zik}) + lnP(I'_{zij} = k|I_j = z, \theta_{zik}) + lnP(I_j = z|\Theta_z)] \\
&+ lnP(\theta_{zik})
\end{aligned}
\tag{4.29}
$$

In Equation (4.29), $P(I'_{zij} = k|I_j = z, y_{zij}, \theta_{zik}^{(l)})$ is the posterior probability of the noise mode identity which can be expressed using Bayes rule as,

$$
\begin{aligned}
P(I'_{zij}=k|I_j=z,y_{zij},\theta^{(l)}_{zik}) &= \frac{P(y_{zij}|I'_{zij}=k,I_j=z,\theta^{(l)}_{zik})P(I'_{zij}=k|I_j=z,\theta^{(l)}_{zik})}{P(y_{zij}|I_j=z,\theta^{(l)}_{zik})} \\
&= \frac{P(y_{zij}|I'_{zij}=k,I_j=z,\theta^{(l)}_{zik})P(I'_{zij}=k|I_j=z,\theta^{(l)}_{zik})}{\sum_{t=1}^{2}P(y_{zij}|I'_{zij}=t,I_j=z,\theta^{(l)}_{zit})P(I'_{zij}=t|I_j=z,\theta^{(l)}_{zir})} \\
&= \frac{\frac{1}{\sqrt{2\pi\sigma^{2\ (l)}_{zik}}}exp(\frac{-(y_{zij}-x^{(l)}_{zi})^2}{2\sigma^{2\ (l)}_{zik}})P(I'_{zij}=k|I_j=z,\theta^{(l)}_{zik})}{\sum_{t=1}^{2}\frac{1}{\sqrt{2\pi\sigma^{2\ (l)}_{zit}}}exp(\frac{-(y_{zij}-x^{(l)}_{zi})^2}{2\sigma^{2\ (l)}_{zit}})P(I'_{zij}=t|I_j=z,\theta^{(l)}_{zit})}
\end{aligned}
\tag{4.30}
$$

In the Equation (4.30), $P(I'_{zij}=k|I_j=z,\theta^{(l)}_{zik})$ is the prior probability of the noise mode identity which is equal to the weight of k^{th} component, i.e., $P(I'_{zij}=k|I_j=z,\theta^{(l)}_{zik})=\delta^{(l)}_{zik}$. By substituting it in the Equation (4.30), it can be further simplified as,

$$
P(I'_{zij}=k|I_j=z,y_{zij},\theta^{(l)}_{zik}) = \frac{\frac{\delta^{(l)}_{zik}}{\sqrt{2\pi\sigma^{2\ (l)}_{zik}}}exp(\frac{-(y_{zij}-x^{(l)}_{zi})^2}{2\sigma^{2\ (l)}_{zik}})}{\sum_{t=1}^{2}\frac{\delta^{(l)}_{zit}}{\sqrt{2\pi\sigma^{2\ (l)}_{zit}}}exp(\frac{-(y_{zij}-x^{(l)}_{zi})^2}{2\sigma^{2\ (l)}_{zit}})} = \gamma^{(l)}_{zijk} \tag{4.31}
$$

Since all the parameters in the above equation are obtained from the previous iteration, we have simplified and shown them as $\gamma^{(l)}_{zijk}$. In Equation (4.29), $P(I_j=z|\mathbf{y}^s_j,\Theta^{(l)}_z)$ is the posterior probability of the operating mode identity which can be calculated using Bayes rule as,

$$
\begin{aligned}
P(I_j = z|\mathbf{y}_j^s, \Theta_z^{(l)}) &= \frac{P(\mathbf{y}_j^s|I_j = z, \Theta_z^{(l)})P(I_j = z|\Theta_z^{(l)})}{P(\mathbf{y}_j^s|\Theta_z^{(l)})} \\
&= \frac{P(\mathbf{y}_j^s|I_j = z, \Theta_z^{(l)})P(I_j = z|\Theta_z^{(l)})}{\sum_{s=1}^q P(\mathbf{y}_j^s|I_j = s, \Theta_s^{(l)})P(I_j = s|\Theta_s^{(l)})} \\
&= \frac{\left(\prod_{i=1}^n P(y_{zij}|I_j = z, \theta_{zi}^{(l)})\right)P(I_j = z|\Theta_z^{(l)})}{\sum_{s=1}^q \left(\prod_{i=1}^n P(y_{zij}|I_j = s, \theta_{si}^{(l)})\right)P(I_j = s|\Theta_s^{(l)})} \\
&= \frac{\prod_{i=1}^n \left\{\frac{\delta_{zi1}}{\sqrt{2\pi\sigma_{zi1}^{2}{}^{(l)}}}exp(\frac{-(y_{zij}-x_{zi}^{(l)})^2}{2\sigma_{zi1}^{2}{}^{(l)}}) + \frac{\delta_{zi2}}{\sqrt{2\pi\sigma_{zi2}^{2}{}^{(l)}}}exp(\frac{-(y_{zij}-x_{zi}^{(l)})^2}{2\sigma_{zi2}^{2}{}^{(l)}})\right\}P(I_j = z|\Theta_z^{(l)})}{\sum_{s=1}^q \prod_{i=1}^n \left\{\frac{\delta_{si1}}{\sqrt{2\pi\sigma_{si1}^{2}{}^{(l)}}}exp(\frac{-(y_{sij}-x_{si}^{(l)})^2}{2\sigma_{si1}^{2}{}^{(l)}}) + \frac{\delta_{si2}}{\sqrt{2\pi\sigma_{si2}^{2}{}^{(l)}}}exp(\frac{-(y_{sij}-x_{si}^{(l)})^2}{2\sigma_{si2}^{2}{}^{(l)}})\right\}P(I_j = s|\Theta_s^{(l)})}
\end{aligned}
\tag{4.32}
$$

Equation (4.32) can be further simplified by substituting $P(I_j = z|\Theta_z^{(l)}) = \eta_{zj}^{(l)}$, since $P(I_j = z|\Theta_z^{(l)})$ is the prior probability of the operating mode which is equal to the weight of z^{th} operating mode.

$$
P(I_j = z|\mathbf{y}_j^s, \Theta_z^{(l)}) = \frac{\eta_{zj}^{(l)} \prod_{i=1}^n \left\{\frac{\delta_{zi1}}{\sqrt{2\pi\sigma_{zi1}^{2}{}^{(l)}}}exp(\frac{-(y_{zij}-x_{zi}^{(l)})^2}{2\sigma_{zi1}^{2}{}^{(l)}}) + \frac{\delta_{zi2}}{\sqrt{2\pi\sigma_{zi2}^{2}{}^{(l)}}}exp(\frac{-(y_{zij}-x_{zi}^{(l)})^2}{2\sigma_{zi2}^{2}{}^{(l)}})\right\}}{\sum_{s=1}^q \eta_{sj}^{(l)} \prod_{i=1}^n \left\{\frac{\delta_{si1}}{\sqrt{2\pi\sigma_{si1}^{2}{}^{(l)}}}exp(\frac{-(y_{sij}-x_{si}^{(l)})^2}{2\sigma_{si1}^{2}{}^{(l)}}) + \frac{\delta_{si2}}{\sqrt{2\pi\sigma_{si2}^{2}{}^{(l)}}}exp(\frac{-(y_{sij}-x_{si}^{(l)})^2}{2\sigma_{si2}^{2}{}^{(l)}})\right\}}
\tag{4.33}
$$

Here we can also simply replace the above equation with $\zeta_{zj}^{(l)}$, i.e., $P(I_j = z|\mathbf{y}_j^s, \Theta_z^{(l)}) = \zeta_{zj}^{(l)}$, since all the parameters are obtained in the previous iteration.

$$
P(y_{zij}|I'_{zij} = k, I_j = z, \theta_{zik}) = \frac{1}{\sqrt{2\pi\sigma_{zik}^2}}exp(\frac{-(y_{zij} - x_{zi})^2}{2\sigma_{zik}^2}) \tag{4.34}
$$

$$
P(I'_{zij} = k|I_j = z, \theta_{zik}) = \delta_{zik} \tag{4.35}
$$

$$P(I_j = z|\Theta_z) = \eta_{zj} \tag{4.36}$$

The *log* of joint prior probability of the parameters is given by,

$$\begin{aligned}
\sum_{z=1}^{q}\sum_{i=1}^{n}\sum_{k=1}^{2} lnP(\theta_{zik}) &= \sum_{z=1}^{t}\sum_{i=1}^{n}\sum_{k=1}^{2}(lnP(x_{zi}) + lnP(\sigma_{zi}^2) + lnP(\delta_{zik}) + lnP(\sigma_{zik}^2)) \\
&= -\sum_{z=1}^{q}\sum_{i=1}^{n}\sum_{k=1}^{2}(ln\sigma_{zi} + \frac{(x_{zi}-\mu_{zi})^2}{2\sigma_{zi}^2} + 2(\alpha_i+1)ln\sigma_{zi} + \frac{\beta_i}{\sigma_{zi}^2} \\
&- (\phi_{zi}/2-1)ln(\delta_{zik}) + 2(\alpha_{ik}+1)ln\sigma_{zik} + \frac{\beta_{ik}}{\sigma_{zik}^2})
\end{aligned} \tag{4.37}$$

Substituting Equations (4.31)-(4.37) into Equation (4.29), the following expression for R function is obtained:

$$\begin{aligned}
R(\Theta, \Theta^{(l)}) &= Q(\Theta, \Theta^{(l)}) + lnP(\Theta) \\
&= \sum_{z=1}^{q}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{2} P(I_j = z|\mathbf{y}_j^s, \Theta_z^{(l)})P(I'_{ij} = k|I_j = z, y_{zij}, \theta_{zik}^{(l)}) \\
&[lnP(y_{zij}|I'_{zij} = k, I_j = z, \theta_{zik}) + lnP(I'_{zij} = k|I_j = z, \theta_{zik}) \\
&+ lnP(I_j = z|\Theta_z)] + lnP(\theta_{zik}) \\
&= \sum_{z=1}^{q}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{2} \zeta_{zj}^{(l)}\gamma_{zijk}^{(l)}[lnP(y_{zij}|I'_{zij} = k, I_j = z, \theta_{zik}) \\
&+ lnP(I'_{zij} = k|I_j = z, \theta_{zik}) + lnP(I_j = z|\Theta_z)] + lnP(\theta_{zik})
\end{aligned} \tag{4.38}$$

$$
\begin{aligned}
R(\Theta,\Theta^{(l)}) &= \sum_{z=1}^{q}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{2}\zeta_{zj}^{(l)}\gamma_{zijk}^{(l)}[ln(\frac{1}{\sqrt{2\pi\sigma_{zik}^2}}exp(\frac{-(y_{zij}-x_{zi})^2}{2\sigma_{zik}^2})) + ln(\delta_{zik}) + ln(\eta_{zj})] \\
&- \sum_{z=1}^{q}\sum_{i=1}^{n}\sum_{k=1}^{2}(ln\sigma_{zi} + \frac{(x_{zi}-\mu_{zi})^2}{2\sigma_{zi}^2} + 2(\alpha_i+1)ln\sigma_{zi} + \frac{\beta_i}{\sigma_{zi}^2} - (\phi_{zi}/2-1)ln(\delta_{zik}) \\
&+ 2(\alpha_{ik}+1)ln\sigma_{zik} + \frac{\beta_{ik}}{\sigma_{zik}^2}) \\
&= \sum_{z=1}^{q}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{2}\zeta_{zj}^{(l)}\gamma_{zijk}^{(l)}[-ln(\sqrt{2\pi\sigma_{zik}^2}) - \frac{(y_{zij}-x_{zi})^2}{2\sigma_{zik}^2} + ln(\delta_{zik}) + ln(\eta_{zj})] \\
&- \sum_{z=1}^{q}\sum_{i=1}^{n}\sum_{k=1}^{2}(ln\sigma_{zi} + \frac{(x_{zi}-\mu_{zi})^2}{2\sigma_{zi}^2} + 2(\alpha_i+1)ln\sigma_{zi} + \frac{\beta_i}{\sigma_{zi}^2} - (\phi_{zi}/2-1)ln(\delta_{zik}) \\
&+ 2(\alpha_{ik}+1)ln\sigma_{zik} + \frac{\beta_{ik}}{\sigma_{zik}^2})
\end{aligned} \tag{4.39}
$$

4.4.2.2 M-step

The M-step involves maximizing $R(\Theta, \Theta^{(l)})$ with respect to the parameters, Θ, and this results in the update equations for the parameters.

Update expression for σ_{zik}: This can be obtained by setting the partial derivative of R w.r.t σ_{srt} to zero. The partial derivative is obtained as follows:

$$
\frac{\partial R(\Theta,\Theta^{(l)})}{\partial \sigma_{srt}} = \sum_{j=1}^{m}\zeta_{sj}^{(l)}\gamma_{srjt}^{(l)}[-\frac{1}{\sigma_{srt}} - \frac{(y_{srj}-x_{sr})^2}{\sigma_{srt}^3}] - [\frac{2(\alpha_{rt}+1)}{\sigma_{srt}} - \frac{2\beta_{rt}}{\sigma_{srt}^3}] = 0 \tag{4.40}
$$

for $s = 1, ..., q$, $r = 1, ..., n$, and $t = 1, 2$. The derivative can be simplified as,

$$
(\sum_{j=1}^{m}\zeta_{sj}^{(l)}\gamma_{srjt}^{(l)} + 2(\alpha_{rt}+1))\sigma_{srt}^2 - \sum_{j=1}^{m}\zeta_{sj}^{(l)}\gamma_{srjt}^{(l)}(y_{srj}-x_{sr})^2 + 2\beta_{rt} = 0 \tag{4.41}
$$

Now the update equation for σ_{srt} is given by,

$$\sigma_{srt}^{(l+1)} = \sqrt{\frac{\sum_{j=1}^{m} \zeta_{sj}^{(l)} \gamma_{srjt}^{(l)} (y_{srj} - x_{sr}^{(l)})^2 + 2\beta_{rt}}{\sum_{j=1}^{m} \zeta_{sj}^{(l)} \gamma_{srjt}^{(l)} + 2(\alpha_{rt} + 1)}} \tag{4.42}$$

Update expression for δ_{zik}: The updating equation for δ_{srt} can be derived by formulating the following constrained optimization problem:

$$\begin{cases} \delta_{zik}^{(l+1)} = \arg\max_{\delta_{zik}} R(\Theta, \Theta^{(l)}) \\ s.t. \qquad \sum_{k=1}^{2} \delta_{zik}^{(l+1)} = 1 \end{cases} \tag{4.43}$$

In order to solve the above constrained optimization problem, the *Lagrange multipliers* method can be applied.

$$L(\delta_{zik}, \lambda_{\delta zi}) = \sum_{z=1}^{q} \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{2} \zeta_{zj}^{(l)} \gamma_{zijk}^{(l)} ln(\delta_{zik}) + (\phi_{zi}/2 - 1) ln(\delta_{zik}) - \lambda_{\delta zi} (\sum_{k=1}^{2} \delta_{zik} - 1) \tag{4.44}$$

The above equation is solved by taking partial derivative of $L(\delta_{zik}, \lambda_{\delta zi})$ w.r.t δ_{zik} and $\lambda_{\delta zi}$, and then setting them to zero.

$$\begin{aligned} \frac{\partial L(\delta_{zik}, \lambda_{\delta zi})}{\partial \delta_{srt}} &= \frac{\sum_{j=1}^{m} \zeta_{sj}^{(l)} \gamma_{srjt}^{(l)} + \phi_{zi}/2 - 1}{\delta_{srt}} - \lambda_{\delta sr} = 0 \\ &\rightarrow \delta_{srt} = \frac{\sum_{j=1}^{m} \zeta_{sj}^{(l)} \gamma_{srjt}^{(l)} + \phi_{zi}/2 - 1}{\lambda_{\delta sr}} \\ &\rightarrow \sum_{k=1}^{2} \delta_{srk} = \frac{\sum_{k=1}^{2} \sum_{j=1}^{m} \zeta_{sj}^{(l)} \gamma_{srjk}^{(l)} + \phi_{zi}/2 - 1}{\lambda_{\delta sr}} \end{aligned} \tag{4.45}$$

$$\frac{\partial L(\delta_{zik}, \lambda_{\delta zi})}{\partial \lambda_{\delta sr}} = \sum_{k=1}^{2} \delta_{srk} - 1 = 0 \tag{4.46}$$

Owing to the property of the γ_{srjt} equation, the following expression is obtained:

$$\sum_{k=1}^{2} \gamma_{zijk}^{(l)} = 1 \rightarrow \sum_{j=1}^{m}\sum_{k=1}^{2} \zeta_{zj}\gamma_{zijk}^{(l)} = \sum_{j=1}^{m} \zeta_{zj} \tag{4.47}$$

By substituting Equation (4.47) into Equation (4.45), it is obtained that $1 = \frac{\sum_{j=1}^{m}\zeta_{zj}+\phi_r/2-1}{\lambda_{\delta r}} \rightarrow \lambda_{\delta r} = \sum_{j=1}^{m}\zeta_{zj} + \phi_r/2 - 1$, and as a result, the updating equation for $\delta_{rt}^{(l+1)}$ would be:

$$\delta_{srt} = \frac{\sum_{j=1}^{m}\zeta_{sj}^{(l)}\gamma_{srjt}^{(l)} + \phi_r/2 - 1}{\sum_{j=1}^{m}\zeta_{sj}^{(l)} + \phi_r/2 - 1} \tag{4.48}$$

Update expression for η_{zj}: This can be derived by solving the following constrained optimization problem:

$$\begin{cases} \eta_{zj}^{(l+1)} = \arg\max_{\eta_{zj}} R(\Theta, \Theta^{(l)}) \\ s.t. \qquad \sum_{z=1}^{q}\eta_{zj}^{(l+1)} = 1 \end{cases} \tag{4.49}$$

Again, in order to solve the above optimization problem, the *Lagrange multipliers* method is applied and the corresponding function is given by,

$$L(\eta_{zj}, \lambda_{\eta z}) = \sum_{z=1}^{q}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{2} \zeta_{zj}^{(l)}\gamma_{zijk}^{(l)} ln(\eta_{zj}) - \lambda_{\eta z}(\sum_{z=1}^{q}\eta_{zj} - 1) \tag{4.50}$$

In order to obtain the update expression for η_{zj}, the partial derivatives of the above equation with respect to η_{sp} and $\lambda_{\eta s}$ must be determined.

$$\begin{aligned} \frac{\partial L(\eta_{jz}, \lambda_{\eta z})}{\partial \eta_{ps}} &= \sum_{i=1}^{n}\sum_{k=1}^{2} \zeta_{sj}^{(l)}\gamma_{sipk}^{(l)}\frac{1}{\eta_{sp}} - \lambda_{\eta s} = 0 \\ &\rightarrow \eta_{sp} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{2}\zeta_{sp}^{(l)}\gamma_{sipk}^{(l)}}{\lambda_{\eta s}} \\ &\rightarrow \eta_{sp} = \frac{\sum_{i=1}^{n}\zeta_{sp}^{(l)}}{\lambda_{\eta s}} \end{aligned} \tag{4.51}$$

$$\frac{\partial L(\eta_{zj}, \lambda_{\eta z})}{\partial \lambda_{\eta s}} = \sum_{z=1}^{q}\eta_{zp} - 1 = 0 \tag{4.52}$$

Similar to the approach in obtaining the update equation for δ_{srt}, if we get a summation of η_{sp} over z we have,

$$\begin{aligned}\sum_{z=1}^{q}\eta_{zp} &= \frac{\sum_{z=1}^{q}\sum_{i=1}^{n}\zeta_{sp}^{(l)}}{\lambda_{\eta s}} \\ \rightarrow 1 = \frac{\sum_{i=1}^{n}1}{\lambda_{\eta s}} &= \frac{n}{\lambda_{\eta s}} \rightarrow \lambda_{\eta s} = n\end{aligned} \tag{4.53}$$

Therefore,

$$\eta_{sp} = \frac{\sum_{i=1}^{n}\zeta_{sp}^{(l)}}{n} \tag{4.54}$$

Update expression for σ_{zi}: This can be obtained by taking the partial derivative of the R function w.r.t the σ_{sr} and setting it to zero.

$$\frac{\partial R(\Theta,\Theta^{(l)})}{\partial\sigma_{sr}} = \frac{(x_{sr}-\mu_{sr})^2}{\sigma_{sr}^3} - \frac{2\alpha_r+3}{\sigma_{sr}} + \frac{2\beta_r}{\sigma_{sr}^3} = 0 \tag{4.55}$$

As a result, the update equation for x_{zi} is given by,

$$\sigma_{sr}^{(l+1)} = \sqrt{\frac{(x_{sr}^{(l)}-\mu_{sr}^{(l)})^2+2\beta_r}{2\alpha_r+3}} \tag{4.56}$$

Update expression for x_{zi}: Similarly, the update equation for the rectified estimate $x_{zi}^{(l+1)}$ is obtained using the *Lagrange multipliers* method:

$$\begin{cases} x_{zi}^{(l+1)} = \arg\max_{x_{zi}} R(\Theta,\Theta^{(l)}) \\ s.t. \qquad A\mathbf{x_z} = 0 \end{cases} \tag{4.57}$$

The Lagrangian function for the above constrained optimization problem is given by,

$$
\begin{aligned}
L(\mathbf{x_z}, \lambda_{xz}) &= -\sum_{z=1}^{q}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{2}\zeta_{jz}^{(l)}\gamma_{zijk}^{(l)}\frac{(y_{zij}-x_{zi})^2}{2\sigma_{zik}^2} - \sum_{z=1}^{q}\sum_{i=1}^{n}\sum_{k=1}^{2}\frac{(x_{zi}-\mu_{zi})^2}{2\sigma_{zi}^2} + \lambda_{xz}^T A\mathbf{x_z} \\
&= \sum_{z=1}^{q}\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{2} -\zeta_{jz}^{(l)}\gamma_{zijk}^{(l)}\frac{(y_{zij}-x_{zi})^2}{2\sigma_{zik}^2} - \frac{(x_{zi}-\mu_{zi})^2}{2\sigma_{zi}^2} + \sum_{i=1}^{n}\sum_{f=1}^{e}\lambda_{xzf}a_{if}x_{zi}
\end{aligned}
\tag{4.58}
$$

In order to obtain the update equation for rectified values, the partial derivatives of the above function w.r.t x_{zi} and λ_{xzf} have to be determined.

$$
\frac{\partial L(\mathbf{x_z}, \lambda_{xz})}{\partial x_{sr}} = \sum_{j=1}^{m}\zeta_{js}^{(l)}\left[\frac{\gamma_{srj1}}{\sigma_{sr1}^{(l+1)^2}} + \frac{\gamma_{srj2}}{\sigma_{sr2}^{(l+1)^2}}\right](y_{srj}-x_{sr}) - \frac{x_{sr}-\mu_{sr}^{(l+1)}}{\sigma_{sr}^{(l+1)^2}} + \sum_{f=1}^{e}\lambda_{xzf}a_{if} = 0 \tag{4.59}
$$

$$
\frac{\partial L(\mathbf{x_z}, \lambda_{xz})}{\partial \lambda_{xzf}} = \sum_{i=1}^{n}\sum_{f=1}^{e}a_{if}x_{zi} = 0 \tag{4.60}
$$

Define $w_{srj}^{(l+1)}$ and $z_{sr}^{(l+1)}$ as:

$$
w_{srj}^{(l+1)} = \zeta_{sj}^{(l)}\left[\frac{\gamma_{srj1}}{\sigma_{sr1}^{(l+1)^2}} + \frac{\gamma_{srj2}}{\sigma_{sr2}^{(l+1)^2}}\right] \tag{4.61}
$$

$$
z_{sr}^{(l+1)} = \frac{-1}{\sigma_{sr}^{(l+1)^2}} \tag{4.62}
$$

Here, $w_{srj}^{(l+1)}$ and $z_{sr}^{(l+1)}$ correspond to the weights for each measured variable for each operating mode. Now, we rewrite the above equation in matrix form such that the resulting expression for rectified values can be compared to the expressions form literature. To this end, we define a diagonal matrices $W_{sj}^{(l+1)} = diag\{w_{s1j}^{(l+1)}, \cdots, w_{snj}^{(l+1)}\}$ and $Z_s^{(l+1)} = diag\{z_{s1}^{(l+1)}, \cdots, z_{sn}^{(l+1)}\}$ for each operating mode, where $diag_{(j=1,\ldots,m)}(a_j)$ denotes the operator that creates the diagonal matrix having the dimension of measured variables (i.e., n) with diagonal elements $w_{srj}^{(l+1)}$ and $z_{sr}^{(l+1)}$. Now, Equation (4.59) can be simplified as

$$\begin{aligned} &\textstyle\sum_{j=1}^{m} W_{zj}^{(l+1)}(\mathbf{y}_j^s - \mathbf{x_z}) + Z_z^{(l+1)}(\mathbf{x_z} - \mu_z^{(l+1)}) + A^T\lambda_x = 0 \\ &\textstyle\sum_{j=1}^{m} \Big(W_{zj}^{(l+1)}\mathbf{y}_j^s\Big) + \Big(Z_z^{(l+1)} - \sum_{j=1}^{m} W_{zj}^{(l+1)}\Big)\mathbf{x_z} - Z_z^{(l+1)}\mu_z^{(l+1)} + A^T\lambda_x = 0 \end{aligned} \tag{4.63}$$

Premultiplying by $A\Big(Z_z^{(l+1)} - \sum_{j=1}^{m} W_{zj}^{(l+1)}\Big)^{-1}$ results in,

$$\begin{aligned} &A\Big(Z_z^{(l+1)} - \textstyle\sum_{j=1}^{m} W_{zj}^{(l+1)}\Big)^{-1} \sum_{j=1}^{m} \Big(W_{zj}^{(l+1)}\mathbf{y}_j^s\Big) - A\Big(Z_z^{(l+1)} - \sum_{j=1}^{m} W_{zj}^{(l+1)}\Big)^{-1} \\ &Z_z^{(l+1)}\mu_z^{(l+1)} + A\Big(Z_z^{(l+1)} - \textstyle\sum_{j=1}^{m} W_{zj}^{(l+1)}\Big)^{-1} A^T\lambda_x = 0 \end{aligned} \tag{4.64}$$

In order to simplify the formulation expression, $B_z^{(l+1)}$ and $C_z^{(l+1)}$ are defined such that:

$$B_z^{(l+1)} = (Z_z^{(l+1)} - \sum_{j=1}^{m} W_{zj}^{(l+1)})^{-1} \tag{4.65}$$

$$C_z^{(l+1)} = \sum_{j=1}^{m} (W_{zj}^{(l+1)}\mathbf{y}_j^s) \tag{4.66}$$

Hence, Equation (4.64) is simplified to,

$$AB_z^{(l+1)}C_z^{(l+1)} - AB_z^{(l+1)}Z_z^{(l+1)}\mu_z^{(l+1)} + AB_z^{(l+1)}A^T\lambda_x = 0 \tag{4.67}$$

Therefore, λ_x is given by,

$$\lambda_x = (AB_z^{(l+1)}A^T)^{-1}AB_z^{(l+1)}(Z_z^{(l+1)}\mu_z^{(l+1)} - C_z^{(l+1)}) \tag{4.68}$$

and as a result, the update equation for rectified values is given by,

$$\mathbf{x_z}^{(l+1)} = B_z^{(l+1)}(I - A^T(AB_z^{(l+1)}A^T)^{-1}AB_z^{(l+1)})(Z_z^{(l+1)}\mu_z^{(l+1)} - C_z^{(l+1)}) \tag{4.69}$$

The resulting expression Equation (4.69) shows rectified estimates of the true state in Equation (4.8). Table 4.1 represents the proposed algorithm for simultaneous gross error detection and data reconciliation under different operating conditions.

Table 4.1: EM Algorithm for Simultaneous Data Reconciliation and Gross Error Detection under Different Operating Conditions

1	Input the raw measurements Y
2	Initialize the parameter $\Theta^{(l)}$
3	**E-step** Evaluate the R function (4.29) using current updated values of parameters
4	**M-step** Update the parameters σ_{zik}, δ_{zik}, η_{zj}, σ_{zi}, and $\mathbf{x_z}$ using Equations (4.42), (4.54), (4.56), (4.48), and (4.69), respectively.
5	Terminate on convergence. Otherwise, proceed to Step 3.

4.5 Simulation Study

In this section, simulated data sets are used to show the performance of the proposed method for simultaneous gross error detection and data reconciliation for a system working under different operating conditions. In order to assess the efficiency of the proposed method, the following performance measure tests are used: (a) Overall power (OP)[10,17], and (b) Average number of type I error (AVTI)[10,17]. They are defined as follows:

$$OP = \frac{Number\ of\ biased\ variables\ correctly\ identified}{Number\ of\ biased\ variables\ simulated} \tag{4.70}$$

$$AVTI = \frac{Number\ of\ unbiased\ variables\ wrongly\ identified}{Number\ of\ simulation\ trials} \tag{4.71}$$

4.5.1 Case 1 - Same Biased Variables in Different Operating Modes

In this example, the performance of the proposed method is illustrated using a water flow network shown in Figure 4.1, while it is working under two operating modes and two variables out of seven contain gross errors.

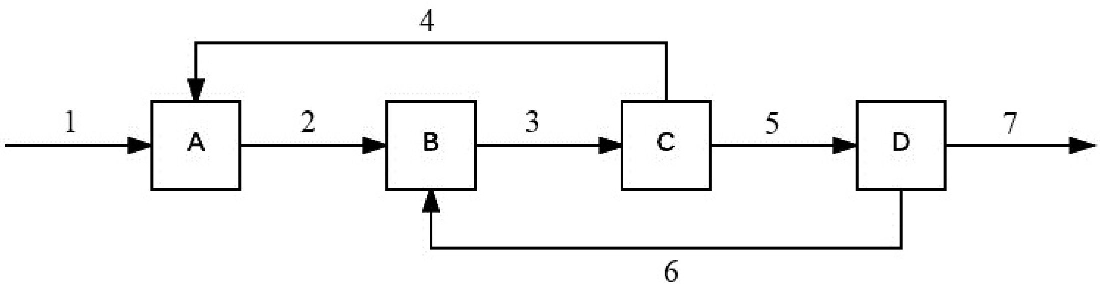

Figure 4.1: Diagram of process network

Based on the above network, the process model equations, which are the constraints of the problem and in this example, are mass balance equations, are as follows,

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

Based on the above equations, the process model coefficient matrix in $A\mathbf{x_z} = 0$ is given by,

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

Assume that x_1 and x_6 are biased, and the magnitude of gross errors in these two variables are +3 and +4, respectively. All the variables are assumed to be measured and the vector of true values of the flow rates for each operating mode are set as $\mathbf{x_1} = \begin{bmatrix} 1 & 2 & 3 & 1 & 2 & 1 & 1 \end{bmatrix}^T$ and $\mathbf{x_2} = \begin{bmatrix} 4 & 5 & 8 & 1 & 7 & 3 & 4 \end{bmatrix}^T$. The measurement data of all the process variables are generated by adding a random noise with zero mean and $0.1I$ variance to the true values. In other words, the standard deviations are around $11 - 32\%$ of the measurements. Also, the probability of occurrence of each operating region is 60% and 40%, respectively. Figure 4.2 represents a set of the simulated data.

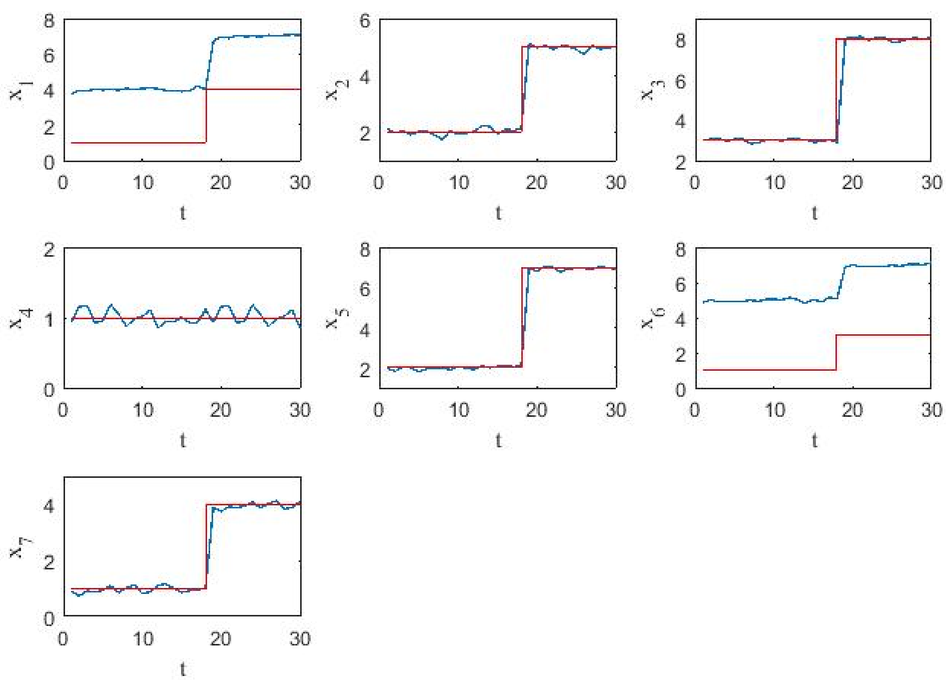

Figure 4.2: Data plot showing gross error in x_1 and x_6, $\sigma_i = 0.1$

Table 4.2 represents the performance of the proposed method using the results of the OP and the AVTI tests.

Table 4.2: Performance measures for multiple gross errors for 50 runs

	Simulation runs	Biased	OP	AVTI
Proposed Method	50	2	0.98	0.08

From Table 4.2 and Figure 4.3, it can be seen that the proposed method can appropriately return rectified values for each operating points.

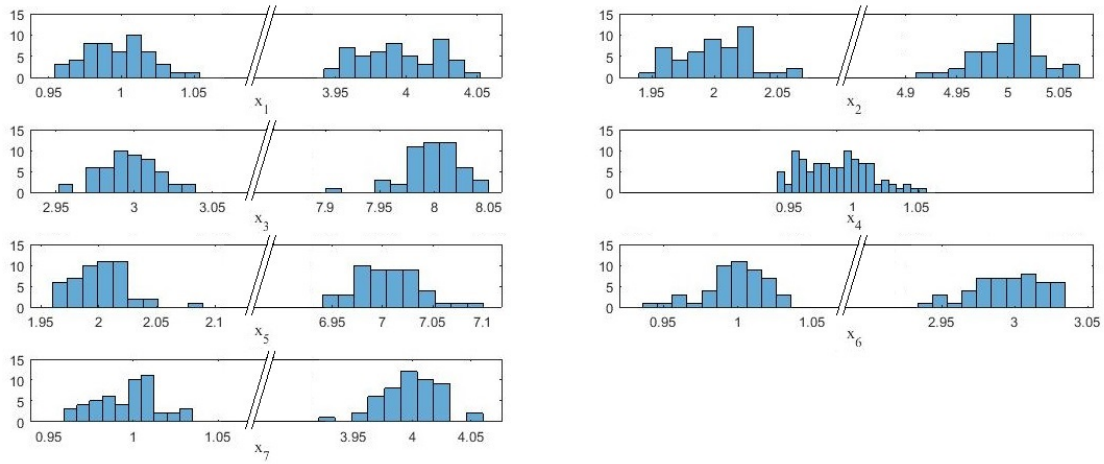

Figure 4.3: Histogram of the rectified values when x_1 and x_6 contain gross errors in both operating modes

Table 4.3 shows the true and rectified values, as well as estimated parameters of MAP estimation.

Table 4.3: Rectified results and estimated distribution parameters for multiple gross errors

Mode	Variables	x	$\hat{x}$	$\hat{\delta}_{i1}$	$\hat{\sigma}_{i1}$	$\hat{\sigma}_{i2}$
	x_1	1	0.9901	1.00	0.0610	3.0074
	x_2	2	1.9738	0.00	0.0922	0.0939
	x_3	3	3.0105	1.00	0.0999	0.0934
First Mode	x_4	1	0.9837	1.00	0.1109	0.0703
	x_5	2	2.0267	1.00	0.0836	0.0924
	x_6	1	1.0366	0.00	0.0100	4.0123
	x_7	1	0.9901	1.00	0.1055	0.0100
	x_1	4	3.9786	1.00	0.1440	3.0057
	x_2	5	4.9988	0.02	0.1083	0.0100
	x_3	8	8.0390	0.98	0.1035	0.0986
Second Mode	x_4	1	1.0202	0.98	0.1086	0.7071
	x_5	7	7.0189	1.00	0.1098	0.0100
	x_6	3	3.0402	0.00	0.0100	3.9864
	x_7	4	3.9786	0.98	0.1140	0.0100

The estimated probability of occurrence of each mode is 56% and 44%, respectively.

4.5.2 Case 2 - Different Biased Variables in Different Operating Modes

In this example, the effectiveness of the proposed method when the biased variables are different for each operating mode is studied. To this end, the data are generated similar to the previous example, and the only difference is in the gross errors of each operating mode. We add gross errors with the magnitude of +3 and +2 to the 2^{nd} and the 7^{th} variables in the first operating mode, respectively, and also a gross error with the magnitude of +4 to the 4^{th} variable in the second operating mode. A set of simulated data are plotted in Figure 4.4.

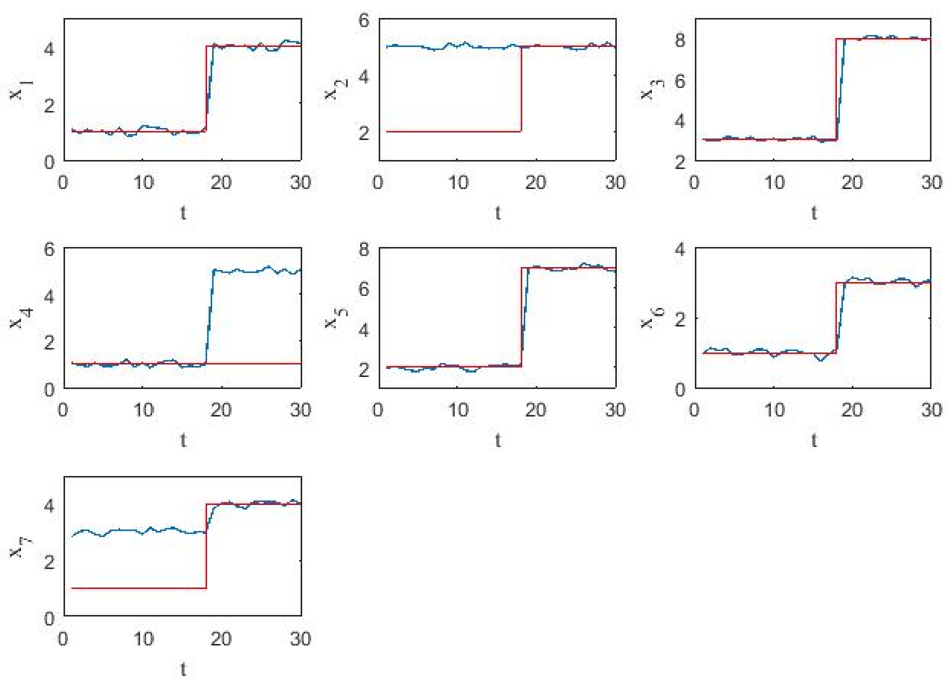

Figure 4.4: Data plot showing gross error in x_1 and x_6 for the first mode and gross error in x_4 for the second mode, $\sigma_i = 0.1$

Table 4.4 represents the performance of the proposed method using the results of the OP and the AVTI tests.

Table 4.4: Performance measures for multiple gross errors for 50 runs

	Simulation runs	Biased	OP	AVTI
Proposed Method	50	2	0.96	0.04

From Table 4.4 and Figure 4.5, it can be seen that the proposed method can appropriately return rectified values for each operating mode.

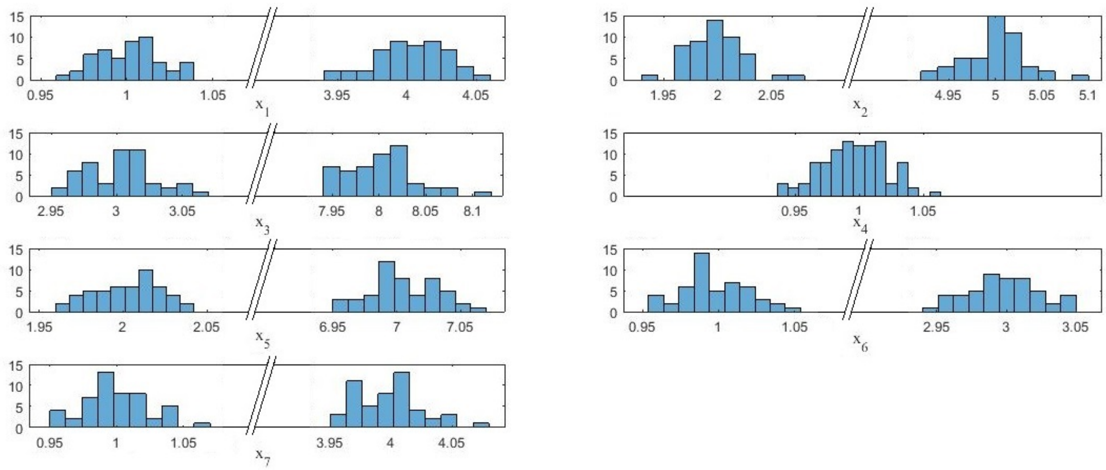

Figure 4.5: Histogram of the rectified values when x_1 and x_6 contain gross error in the first mode and x_4 is corrupted by gross error in the second mode.

Table 4.5 shows the true and rectified values, as well as estimated parameters of MAP estimation.

Table 4.5: Rectified results and estimated distribution parameters for multiple gross errors when the gross errors of each operating are different

Mode	Variables	x	$\hat{x}$	$\hat{\delta}_{i1}$	$\hat{\sigma}_{i1}$	$\hat{\sigma}_{i2}$
	x_1	1	1.0369	1.00	0.1352	0.0369
	x_2	2	2.0731	0.00	0.0855	2.9823
	x_3	3	3.0611	1.00	0.0999	0.0895
First Mode	x_4	1	1.0362	1.00	0.0957	0.0703
	x_5	2	2.0249	1.00	0.0836	0.0924
	x_6	1	0.9880	0.95	0.0931	0.1055
	x_7	1	1.0369	0.00	0.0922	1.9897
	x_1	4	3.9995	1.00	0.0834	0.0120
	x_2	5	5.0407	1.00	0.1083	0.1103
	x_3	8	8.0238	0.98	0.1035	0.1079
Second Mode	x_4	1	1.0477	0.02	0.0100	4.0075
	x_5	7	6.9761	1.00	0.1156	0.0100
	x_6	3	2.9766	0.98	0.1205	0.0733
	x_7	4	3.9995	0.93	0.1140	0.0962

The estimated probability of occurrence of each mode is 66% and 34%, respectively.

4.6 Conclusions

In this chapter the problem of gross error detection and data reconciliation for measurement data which is collected under different operating conditions is studied in a maximum a posteriori framework. The noise model that was used in this study followed a Gaussian mixture distribution with two modes. The formulation of the proposed noise model besides modeling different operating modes resulted in introducing two hidden variables, one for each.

In general, the expectation maximization (EM) algorithm is used to solve an MLE or MAP estimation with hidden variables. Therefore, we used EM approach in order to obtain an estimate of the true state and the error distribution parameters. The proposed method was also capable of returning the rectified values and the magnitude of gross errors for each operating mode even though the biases for each mode are different. Several case studies were presented to illustrate the performance and the different features of the algorithm.

Chapter 5

Conclusions

The main focus of this thesis was to introduce new methods for simultaneous gross error detection and data reconciliation by developing the EM algorithm and utilizing the Gaussian mixture distribution for the noise model. The main contribution of the thesis can be summarized as follows:

Chapter 1 provided the motivation and challenges of data reconciliation and gross error detection and their importance in industrial processes. It also provided the outline and contributions of each Chapter of the thesis.

In Chapter 2, the problem of data reconciliation in the maximum likelihood framework by simultaneously eliminating the gross error in the measurement data was addressed. The contaminated Gaussian distribution model with two modes characterized the error distribution of different sensor mode and hence, the resulting formulation involved solution for hidden variables. The expectation maximization approach was presented which is an iterative algorithm with two expectation and maximization steps. Furthermore, the parameters of the error distribution model were calculated as part of the solution and they were not set in advanced. Moreover, problems with both linear and/or nonlinear constraints were covered in the proposed approach. Finally, several case studies were presented to demonstrate the various features of the proposed method.

Chapter 3 extended the proposed method in Chapter 2 by including the prior knowledge into the objective function. Since the maximum likelihood framework was not able to distinguish between the two gross error sets which have the same effect on the objective function, by adding prior information maximum likelihood was transformed to maximum a posteriori estimation for data reconciliation and gross error detection. By doing so, not only was the more accurate estimate of the true states obtained, but also other possible situations of gross error sets in measurements could be rectified. It is worth to mention that if there was a gross error in a measurement it did not have to be assumed all of the sample points of that variable were corrupted by the gross error, i.e., it could partially occur in a data set. Again, several simulation case studies and examples were provided to show the efficacy of the proposed method.

Chapter 4 expanded the proposed method of Chapter 3 such that it could cover data rectification for the system working under different operating conditions. Therefore, there were multiple operating regions in a data set, and the proposed approach was capable of identifying operating regions, estimating the distribution parameters of each region, and then returning the rectified estimates of each region. Based on these assumptions, the Maximum a Posteriori (MAP) framework was applied and solved using the Expectation Maximization (EM) algorithm. The EM method was used since the proposed mixture distribution involved the noise mode as well as the operating mode as hidden variables. To this end, the data set was divided into several clusters based on the number of operating modes. Then, data rectification, was applied to each operating mode. It is important to note that the proposed approach distinguished between the operating modes in different clusters. As a result, the future model was capable of identifying operating modes and switching between them. Finally, two case studies were presented to demonstrate the efficacy of the proposed method.

All the chapters of this thesis are related to each other, and each chapter extended the proposed method from the previous chapter to cover some other cases besides increasing the accuracy of the rectified estimates. Therefore, the future work could further develop/expand the proposed method and algorithm of Chapter 4.

All the proposed methods in different chapters of the thesis belong to the data reconciliation and gross error detection problems of systems working under steady state conditions, i.e., the constraints of the rectification problem was $f(x, z) = 0$. Extending the proposed methods for unsteady state systems would be a potential future work.

Bibliography

[1] Moustapha Alhaj-Dibo, Didier Maquin, and José Ragot. Data reconciliation: A robust approach using a contaminated distribution. *Control Engineering Practice*, 16(2):159–170, 2008.

[2] Shankar Narasimhan and Cornelius Jordache. *Data reconciliation and gross error detection: An intelligent use of process data*. Gulf Professional Publishing, 1999.

[3] M.J. Fuente, G. Gutierrez, E. Gomez, D. Sarabia, and C. de Prada. Gross error management in data reconciliation. *IFAC-PapersOnLine*, 48(8):623 – 628, 2015.

[4] PM Reilly and RE Carpani. Application of statistical theory of adjustment to material balances. In *Proc. 13th Can. Chem. Eng. Conf. Montreal, Quebec*, 1963.

[5] Ajit C Tamhane and Richard SH Mah. Data reconciliation and gross error detection in chemical process networks. *Technometrics*, 27(4):409–422, 1985.

[6] Richard S Mah, Gregory M Stanley, and Dennis M Downing. Reconcillation and rectification of process flow and inventory data. *Industrial & Engineering Chemistry Process Design and Development*, 15(1):175–183, 1976.

[7] RW Serth and WA Heenan. Gross error detection and data reconciliation in steam-metering systems. *AIChE Journal*, 32(5):733–742, 1986.

[8] CM Crowe, YA Campos, and A Hrymak. Reconciliation of process flow rates by matrix projection. part i: linear case. *AIChE Journal*, 29(6):881–888, 1983.

[9] R. S. H. Mah and A. C. Tamhane. Detection of gross errors in process data. *AIChE Journal*, 28(5):828–830, 1982.

[10] S Narasimhan and RSH Mah. Generalized likelihood ratio method for gross error identification. *AIChE Journal*, 33(9):1514–1521, 1987.

[11] DK Rollins and JF Davis. Unbiased estimation of gross errors in process measurements. *AIChE journal*, 38(4):563–572, 1992.

[12] Hongwei Tong and Cameron M Crowe. Detection of gross erros in data reconciliation by principal component analysis. *AIChE Journal*, 41(7):1712–1722, 1995.

[13] Qiyou Jiang and Miguel J Bagajewicz. On a strategy of serial identification with collective compensation for multiple gross error estimation in linear steady-state reconciliation. *Industrial & engineering chemistry research*, 38(5):2119–2128, 1999.

[14] Joseph Rosenberg, Richard SH Mah, and Corneliu Iordache. Evaluation of schemes for detecting and identifying gross errors in process data. *Industrial & Engineering Chemistry Research*, 26(3):555–564, 1987.

[15] DL Ripps. Adjustment of experimental data. In *Chem. Eng. Prog. Symp. Ser*, volume 61, 1965.

[16] DR Kuehn and Harold Davidson. Computer control ii. mathematics of control. *Chemical Engineering Progress*, 57(6):44–47, 1961.

[17] Tyler A Soderstrom, David M Himmelblau, and Thomas F Edgar. A mixed integer optimization approach for simultaneous data reconciliation and identification of measurement bias. *Control Engineering Practice*, 9(8):869–876, 2001.

[18] Nikhil Arora and Lorenz T. Biegler. Redescending estimators for data reconciliation and parameter estimation. *Computers & Chemical Engineering*, 25(1112):1585 – 1599, 2001.

[19] Yuan Yuan, Shima Khatibisepehr, Biao Huang, and Zukui Li. Bayesian method for simultaneous gross error detection and data reconciliation. *AIChE Journal*, 61(10):3232–3248, 2015.

[20] IB Tjoa and LT Biegler. Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems. *Computers & chemical engineering*, 15(10):679–690, 1991.

[21] Lloyd PM Johnston and Mark A Kramer. Maximum likelihood data rectification: Steady-state systems. *AIChE Journal*, 41(11):2415–2426, 1995.

[22] Diego Martinez Prata, Jos Carlos Pinto, and Enrique Luis Lima. Comparative analysis of robust estimators on nonlinear dynamic data reconciliation. *Computer Aided Chemical Engineering*, 25:501 – 506, 2008.

[23] Derya B Özyurt and Ralph W Pike. Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. *Computers & chemical engineering*, 28(3):381–402, 2004.

[24] Claudia E Llanos, Mabel C Sanchez, and Ricardo A Maronna. Robust estimators for data reconciliation. *Industrial & Engineering Chemistry Research*, 54(18):5096–5105, 2015.

[25] Peter J Huber and EM Ronchetti. Robust statistics wiley. *New York*, 1981.

[26] Geoffrey McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.

[27] Lee W Johnson, Ronald Dean Riess, and Jimmy Thomas Arnold. *Introduction to linear algebra*. Addison-Wesley Longman, 1993.

[28] George B Dantzig and Mukund N Thapa. *Linear programming 2: theory and extensions*. Springer Science & Business Media, 2006.

[29] D Wang and JA Romagnoli. A framework for robust data reconciliation based on a generalized objective function. *Industrial & engineering chemistry research*, 42(13):3075–3084, 2003.

[30] Jose A Romagnoli and Mabel Cristina Sanchez. *Data processing and reconciliation for chemical process operations*, volume 2. Academic Press, 1999.

[31] Cameron M Crowe. Data reconciliation?progress and challenges. *Journal of process control*, 6(2-3):89–98, 1996.

[32] Diego Martinez Prata, Marcio Schwaab, Enrique Luis Lima, and José Carlos Pinto. Simultaneous robust data reconciliation and gross error detection through particle swarm optimization for an industrial polypropylene reactor. *Chemical Engineering Science*, 65(17):4943–4954, 2010.

[33] MJ Leibman, T_F Edgar, and LS Lasdon. Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. *Computers & chemical engineering*, 16(10-11):963–986, 1992.

[34] Miguel J Bagajewicz and Qiyou Jiang. Gross error modeling and detection in plant linear dynamic reconciliation. *Computers & chemical engineering*, 22(12):1789–1809, 1998.

[35] Derrick K Rollins, Yisun Cheng, and Sriram Devanathan. Intelligent selection of hypothesis tests to enhance gross error identification. *Computers & chemical engineering*, 20(5):517–530, 1996.

[36] Mark A Kramer and Richard SH Mah. Model-based monitoring. In *Proc. Second Int. Conf. on Foundations of Computer Aided Process Operations*, pages 45–68. CACHE, 1994.

[37] Halit Eren Bela G Liptak. *Process software and digital networks, Fourth Edition, Volume 3*. CRC Press, 2011.

[38] Derrick K Rollins and Shonda D Roelfs. Gross error detection when constraints are bilinear. *AIChE journal*, 38(8):1295–1298, 1992.

[39] TJ Williams, EJ Schaffer, and A Rohr. 1.10 redundant and voting systems. *Instrument Engineers' Handbook, Volume One: Process Measurement and Analysis*, 1:126, 2003.

[40] Carl Edward Rasmussen. The infinite gaussian mixture model. In *Advances in neural information processing systems*, pages 554–560, 2000.

[41] Y Ramamurthi, PB Sistu, and BW Bequette. Control-relevant dynamic data reconciliation and parameter estimation. *Computers & chemical engineering*, 17(1):41–59, 1993.

[42] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, and Surya N Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers & chemical engineering*, 27(3):293–311, 2003.