DETECTION AND DIAGNOSIS OF DATA RECONCILIATION PROBLEMS IN AN INDUSTRIAL CHEMICAL INVENTORY SYSTEM

Zhengang Han * **Sirish Shah** *,[1]
Shankar Narasimhan [2] **Hasna Zaknoun** [3]

* *Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada*

Abstract: According to one industrial engineer, "reconciliation of mass balances to monitor chemical tank inventories is a struggle even at the best of times". This paper is concerned with the design of an offline and online fault detection and diagnosis monitoring system for the caustic tank inventory process at a pulp and paper company. Even with limited instrument redundancy, the offline monitoring analysis was able to correctly detect and diagnose sensor calibration errors. The scheme is now undergoing online implementation tests.

Keywords: Data reconciliation, gross error detection, process monitoring, sensor calibration error, fault detection and diagnosis

1. INTRODUCTION

Modern chemical plants are usually equipped with hundreds of instruments and many complex processing units. Therefore, even if only one single element of the whole plant does not function normally, the whole process performance and operation can degrade. In order to avoid this and run the whole process more efficiently, a scheme which can monitor the "health" of all the instruments and processing units is necessary from operation point of view.

The basic idea in process monitoring is to timely check the newly measured process variables and to see if they are consistent with the process models under the fault free situation. If there is a large deviation between these two then a fault is said to be detected. This is usually referred as "fault detection". Further, by manipulating the deviation, the location of the fault can be obtained by comparing it with the preset fault signatures. This is known as "fault isolation".

Two elements are essential for fault detection and isolation (FDI): process models and process measurement. For the purpose of robust and yet accurate FDI, one has to obtain appropriate process models. Once the process measurements are found to be inconsistent with respect to these models, one can then infer that the instruments or the process equipment are abnormal without bothering to consider the possibility of model inaccuracy. For example physical relationships, such as mass and energy balances can be used as a process model for FDI. For fault diagnosis, process measurements need to have a level of redundancy or the models need have extra degrees of freedom, i.e. each measured process variable can be inferred from the other measured variables as well as process

[1] Author to whom correspondence should be addressed. Email: Sirish.Shah@UAlberta.ca, Tel: 1-780-4925162
[2] Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai, India
[3] Millar Western Forest Products Ltd., Whitecourt, Canada

models. This type of redundancy is also referred as "analytical redundancy" (Chow and Willsky, 1984).

Data reconciliation (DR), which has been well studied in recent years (Mah 1990, Madron 1992, Crowe 1996, Bagajewicz 2000, Narasimhan and Jordache 2000), is concerned with the validation of process variables that satisfy physical constraints of the process. However, to obtain accurate estimates, some action has to be taken to eliminate the influence of gross errors. Hypothesis testing has been extensively used for detection, isolation and identification of gross errors. Therefore, gross error detection (GED) is closely related to DR. Literature surveys of GED can be found in Narasimhan and Mah (1987), Mah (1990), Madron (1992), Sánchez and Romagnoli (2000) and Narasimhan and Jordache (2000).

This paper is organized as follows: section 2 provides the theoretical background on offline and online process monitoring system design; the process of interest is introduced in section 3; the detailed description of offline and online analysis and results are presented in section 4; and the paper ends with concluding remarks in section 5.

2. SENSOR CALIBRATION ERROR DETECTION AND ISOLATION

Suppose the fault-free process measurement vector $\mathbf{y}(k) \in \Re^m$ can be described by:

$$\mathbf{y}(k) = \mathbf{x}(k) + \boldsymbol{\varepsilon}(k) \tag{1}$$

where $\mathbf{x}(k) \in \Re^m$ is the vector of true values at time k and $\boldsymbol{\varepsilon}(k) \in \Re^m$ is a zero mean normal distributed random vector representing the measurement noise, i.e. $\boldsymbol{\varepsilon}(k) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} \in \Re^{m\times m}$ is the covariance matrix of noise $\boldsymbol{\varepsilon}(k)$. In most cases, the noise vector is mutually independent, i.e. $\boldsymbol{\Sigma}$ is a diagonal matrix.

Assume that the process model is given by

$$\mathbf{A}\mathbf{x}(k) = \mathbf{0} \tag{2}$$

where $\mathbf{A} \in \Re^{q\times m}$ is a known matrix and q is the number of constraints. If sensor(s) are suffering from calibration problems, their measurements may be different from $\mathbf{x}$. Different types of calibration errors as illustrated in Fig. 1 are possible. In this paper, the most general case - a bias plus slope error - is considered and described mathematically as:

$$y_i(k) = \alpha_i x_i(k) + \beta_i + \varepsilon_i(k) \tag{3}$$

where α_i and β_i are calibration parameters associated with slope and bias errors respectively, and x_i and y_i denote the i^{th} element of vector $\mathbf{x}$ and $\mathbf{y}$. In the fault diagnosis literature (Gertler, 1988), the slope error is classified as a multiplicative fault, while the bias error is referred to as an additive fault. In the fault-free case, $\alpha_i = 1$ and $\beta_i = 0$.

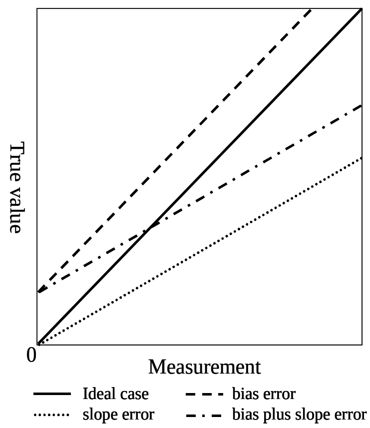

Fig. 1. Illustration of different types of calibration errors

To detect the sensor calibration error, one can directly project the constraint matrix $\mathbf{A}$ on the measured variables $\mathbf{y}(k)$ and check the residuals $\mathbf{r}(k) = \mathbf{A}\mathbf{y}(k)$. Obviously, if the sensors do not have any calibration error, the residuals will be only related to the measurement noise, i.e. $\mathbf{r}(k) = \mathbf{A}\boldsymbol{\varepsilon}(k)$. Therefore, the residuals will be zero-mean, i.e. $\mathbf{r}(k) \sim \mathcal{N}(0, \Sigma_r)$. However, if some of the sensors suffer from calibration errors, the residuals will not be zero-mean. Thus, one can easily detect possible sensor calibration error(s) by checking the mean of residuals.

Once a calibration problem is detected, the next step is to proceed to the diagnosis stage, i.e. to determine the exact location and type of the sensor calibration error. In this step, one has to categorize the process variables into two sets: suspected $\mathbf{x}^s(k) \in \Re^{m_1}$ and unsuspected $\mathbf{x}^u(k) \in \Re^{m_2}$. Thus, the process model $\mathbf{A}$ can be rearranged as $\mathbf{A} = [\mathbf{A}^s \;\; \mathbf{A}^u]$ where $\mathbf{A}^s$ and $\mathbf{A}^u$ correspond to $\mathbf{x}^s$ and $\mathbf{x}^u$ respectively. Therefore, Eqn. 2 can be re-written as:

$$\mathbf{A}^s\mathbf{x}^s(k) + \mathbf{A}^u\mathbf{x}^u(k) = \mathbf{0} \tag{4}$$

with the assumption of

$$\begin{aligned}\mathbf{y}^s(k) &= \Lambda\mathbf{x}^s(k) + \Delta + \boldsymbol{\varepsilon}^s(k)\\ \mathbf{y}^u(k) &= \mathbf{x}^u(k) + \boldsymbol{\varepsilon}^u(k)\end{aligned}$$

where $\Lambda \in \Re^{m_1\times m_1}$ is an unknown diagonal matrix and $\Delta \in \Re^{m_1}$ is an unknown vector.

When one applies the above assumption on the process constraint Eqn. 4, the following can be obtained:

Table 1. Measured variables of chemical tank inventory system

Names	Descriptions	Units
L	Tank level	%
F1	Flow rate in outflow #1	L/min
F2	Flow rate in outflow #2	kg/min
F3	Flow rate in outflow #3	kg/min
F4	Flow rate in outflow #4	L/min
F5	Flow rate in outflow #5	kg/min
F6	Flow rate in outflow #6	kg/min
P1	Valve position of outflow #1	%
P2	Valve position of outflow #2	%
P3	Valve position of outflow #3	%
P4	Valve position of outflow #4	%
P5	Valve position of outflow #5	%
P6	Valve position of outflow #6	%

$$\begin{aligned}\mathbf{r}(k) &= \mathbf{A}^u\mathbf{y}^u(k) + \mathbf{A}^s\Lambda^{-1}\mathbf{y}^s(k) - \mathbf{A}^s\Lambda^{-1}\Delta \\ &= \mathbf{A}^s\Lambda^{-1}\boldsymbol{\varepsilon}^s(k) + \mathbf{A}^u\boldsymbol{\varepsilon}^u(k) \end{aligned} \tag{5}$$

Then, the problem can be formulated in the form of a *least-squares optimization* problem as follows:

$$J = \min_{\Lambda,\Delta} \mathbf{r}^T(k)\mathbf{r}(k) \tag{6}$$

Please note that only $2m_1$ parameters need to be determined by the optimization routine as Λ is a diagonal matrix. By observing the estimated value of the diagonal elements of Λ and Δ, one can determine which sensor(s) are mostly likely to suffer from calibration errors by comparing the estimates with the fault-free case.

3. PROCESS DESCRIPTION

The process of interest is a caustic tank inventory system at Millar Western Forest Products Ltd., Whitecourt, Canada. The process serves as a buffer to supply the caustic usage for the whole plant. The simplified schematics of the tank inventory system is illustrated in Fig. 2.

In the process, caustic is delivered by transportation trucks at periodic intervals depending on the plant chemical usage. Therefore, the inlet flow rate is not continuous and hence not measured. There are 6 independent outlet flow control loops. Each pipeline supplies caustic to different parts of the plant and the flow rates in the outlet pipes are controlled independently.

The level of the tank is not controlled but measured. The tank is a vertical cylinder tank with a diameter of 6.8 meters and a height of 6 meters. Thus the nominal volume is 218 cubic meters. The level sensor is located 0.45 meters above the tank bottom. Therefore, the measurable volume of the tank is about 200 cubic meters. All the related measured variables are listed in Table 1.

Notice from Table 1, that the six flow rate sensors do not use the same measurement units. Two of them (F1 and F4) are volume flow sensors in L/min, while the other four are mass flow sensors in kg/min. In this paper, it is assumed the delivered caustic having a constant density: $\rho = 1.52kg/L$.

Furthermore, all six valves are equal percentage valves, which are essentially nonlinear valves. The characteristic of equal percentage valve are described in the literature (Johnson 1988). All flow rate, level and valve position data are sampled every 1 minute, i.e. $T_s = 1$ min.

4. MONITORING SYSTEM DESIGN AND DIAGNOSIS RESULTS

4.1 Process model description

4.1.1. Mass Balance The relationship between the volume of caustic in the tank and the inlet and outlet flow rates can be represented by the following equation:

$$\rho\frac{dV(t)}{dt} = \rho F_{in}(t) - \rho F_{out}(t) \tag{7}$$

where $V(t)$ is the chemical volume in the tank, and $F_{in}(t)$ and $F_{out}(t)$ are the inlet and outlet volume flow rates respectively.

In discrete time domain, the above equation can be written as:

$$\begin{aligned} r(k) &= V(k) - V(k-1) - (F_{in}(k) - F_{out}(k))\,T_s \\ &= 0 \end{aligned} \tag{8}$$

where $r(k)$ is defined as the "mass balance residual". This discrete-time representation implies that the inlet and outlet flow rates do not change in between samples. In this process, since the sampling interval is relatively large, this assumption does not hold most of the time. Therefore, when this discrete-time model is used, small deviation of the residuals from zero can be expected due to this systematic model-plant mismatch (MPM).

$F_{out}(k)$ can be easily calculated by summing all the outlet flow rates. $V(k)$ can also be obtained from the tank level sensor L and *a priori* knowledge of the total volume of the storage tank. However, $F_{in}(k)$ is not measured. Therefore, according to the essential elements of process monitoring mentioned in the introduction section, the process measurements are not "redundant".

Nevertheless, this does not mean that one can not design a monitoring scheme on the current system. Once we take a closer look at the current system, the operation can be categorized into two stages - "during delivery" and "zero delivery" - with different F_{in}s as below:

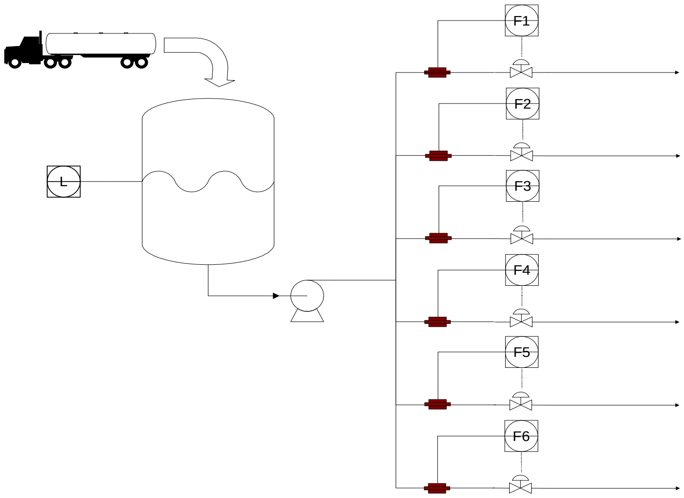

Fig. 2. The schematics of chemical tank inventory process at Millar Western Forest Products Ltd.

$$F_{in}(k) = \begin{cases} 0, & \text{zero delivery} \\ \text{unknown}, & \text{during delivery} \end{cases} \quad (9)$$

Thus, during the period of "zero delivery", one exactly knows F_{in} even though there is no measurement. Therefore, when there is no delivery, the "redundancy" condition can be satisfied. One can make use of this "redundancy" for the purpose of process monitoring. The details are shown later in this paper.

4.1.2. Valve characteristics All control valves in this process are equal percentage valves. Theoretically, the flow rate and valve position should obey the following relationship (Johnson 1988):

$$F = F_{min} R^{S/S_{max}} \quad (10)$$

where $R = \dfrac{F_{max}}{F_{min}}$ is referred to as *rangeability*, F_{max} and F_{min} are the maximum and minimum flow rates respectively, F is flow rate passing through the valve, and S and S_{max} are the stem position and maximum stem position.

By taking logarithm of Eqn. 10, we obtain:

$$\log F = \frac{S}{S_{max}} \cdot \log R + \log F_{min} \quad (11)$$

By doing this, the nonlinear equal percentage valve characteristic has been transformed into a linear framework. Both the flow rate F and the valve position S/S_{max} are measured in this system. Therefore, there are some measurement redundancies with respect to the valve characteristic. The estimates of the unknown parameters $\log R$ and $\log F_{min}$ can be obtained through *least squares* estimation by using archived process data of flow and valve position measurements, provided these measurement do not contain any bias. Once the parameter estimates have been obtained, the valve equations provide the necessary redundancy for fault diagnosis even during delivery periods. Due to limited space, the details of the parameter estimates are omitted.

However, the models of control valves are different from the mass balance model in the following aspects. Firstly, unlike mass balance, valve characteristic cannot be guaranteed to be valid for ever. Due to wear and tear, the parameters of the valve model may change with time. This can be treated as a parametric fault and distinguished from sensor biases using the available data. The diagnosis problem becomes more complex, but it is solvable. Secondly, the mass balance equation is known exactly. But for the valve characteristics, the unknown parameters, such as $\log R$ and $\log F_{min}$, need to be determined from the data. Because of these differences, these two types of models need to be treated differently in the later analysis.

4.2 Offline analysis for detection and diagnosis of sensor calibration errors

After selecting the data portion corresponding to the "zero delivery" case, Eqn. 8 becomes:

$$r(k) = V(k) - V(k-1) + F_{out}(k)T_s \quad (12)$$

where $F_{out}(k) = \sum_{i=1,4} F_i(k) + \frac{1}{\rho} \sum_{i=2,3,5,6} F_i(k)$ and $V(k) = 200,000 \times L(k)$.

Thus, the mass balance residual at each time instant can be obtained. The mass balance residual for a period of 40,000 data points is shown in Fig. 3. From this figure, note that the mean of the residual is negative, and it is not stationary throughout the range. After consulting with the process engineer and confirming that no leakage is taking place, we suspected that the problem may

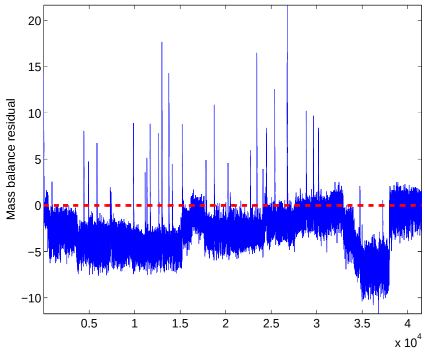

Fig. 3. Mass balance residual before calibration correction

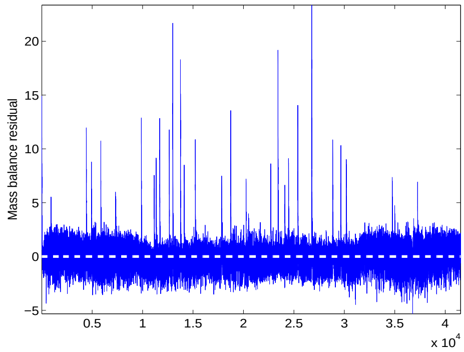

Fig. 4. Mass balance residual after calibration correction

possibly be due to flow sensor calibration error. The target was then to find out which sensor(s) have the calibration problem and what is the magnitude of this calibration problem.

In order to confirm the suspicion, it is assumed that the level sensor is accurate, while all the flow sensors may suffer from bias plus slope calibration errors, i.e. $F_i^f(k) = \alpha_i F_i^\circ(k) + \beta_i + \varepsilon_i(k),\ \forall\ i = 1, \cdots, 6$, where $F_i^f(k)$ and $F_i^\circ(k)$ denote the i^{th} measured and true flow rates respectively, α_i and β_i are constant coefficients, and $\varepsilon_i(k)$ are zero mean normally distributed measurement noise. If the i^{th} flow rate sensor works fine, then ideally $\alpha_i = 1$ and $\beta_i = 0$. Therefore, the mass balance equation can be further written as

$$r(k) = V(k) - V(k-1) + \sum_{i=1}^{6} \frac{T_s}{\alpha_i} F_i^f(k) - \sum_{i=1}^{6} \frac{\beta_i \cdot T_s}{\alpha_i} = \sum_{i=1}^{6} \frac{T_s}{\alpha_i} \varepsilon_i(k) \tag{13}$$

Since all the β_i are included in the same summation operator, one cannot distinguish the β_is from each other. Thus, define $\beta = \sum_{i=1}^{6} \frac{\beta_i}{\alpha_i}$. From the objective function shown in Eqn. 6, it is simply a least-squares problem. Then, from Eqn. 13 through the least-squares estimation, the estimated coefficients and their confidence limits are shown in Table 2.

Since the estimated value of β is small in comparison with the flows, it can be deduced that there are no additive biases in any of the flow sensors (in the absence of bias cancellations). Also from the estimated values of αs for all sensors except F2, it can be inferred that the corresponding flow sensors do not have a slope error. The estimated value of α_2 which is significantly greater than unity, indicates a slope error in the corresponding flow sensor.

The afore-mentioned analysis results were further confirmed by the process control engineer. The problem was traced to inconsistency in the sensor ranges in the DCS system and the sensor. It was determined that the flow rate sensor F2 was changed and re-calibrated to the range of 0 - 15 L/M. Whereas, the conversion range in the DCS system still remained at the original 0 - 25 L/M. Therefore, this gives the scaling factor to be $25/15 = 1.67$, which is close to the estimate 1.73. The residual after calibration correction shown in Fig. 4 is more or less zero mean.

4.3 Online monitoring design and implementation

In order to provide online monitoring scheme for the caustic tank inventory process, the possible abnormal situations that are considered in this paper are listed below:

- *Sensor faults*: the whole process is equipped with 13 sensors. All of these sensors could potentially malfunction during the operation.
- *Process faults*: this type of fault include any malfunction of the process equipment other than instrument errors, e.g. tank leakage.

To correctly detect and further isolate the afore-mentioned faults, analytical redundancy has to be used. In this project, there are totally 7 equations representing various relationships among measured variables. They consist of one mass balance and six valve characteristic equations. For detection, as long as one of these equations becomes invalid, one can infer that a fault has occurred, i.e. a fault is detected. For isolation, since different types of faults affect different equations, one can determine the root cause by observing which equations are affected. Table 3 shows how different faults affect different equations. The following remarks regard to Table 3:

- The Y-axis denotes possible faults.
- The X-axis denotes different equations - MB stands for mass balance and V_i represents the valve characteristic of valve i.
- '$\times$' denotes the intersection where a fault affects the corresponding equation.

Table 2. Estimated calibration coefficients

	α_1	α_2	α_3	α_4	α_5	α_6	β
Estimates	1.0410	1.7354	1.0437	1.0532	1.0084	1.0478	-0.2708
Upper bound	1.2808	2.0551	1.2129	1.3327	1.1965	1.3356	1.2430
Lower bound	0.8011	1.4157	0.8745	0.7737	0.8203	0.7600	-1.7846

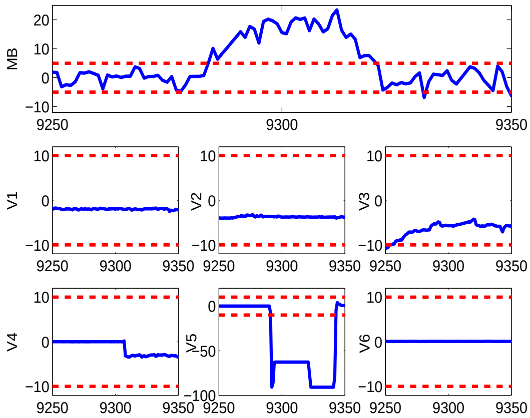

Fig. 5. Residuals when an alarm was generated

Table 3. Fault isolation logic

	MB	V_1	V_2	V_3	V_4	V_5	V_6
L	×						
F1	×	×					
F2	×		×				
F3	×			×			
F4	×				×		
F5	×					×	
F6	×						×
P1		×					
P2			×				
P3				×			
P4					×		
P5						×	
P6							×
Process	×						

- If an intersection is blank then there is no fault affecting the corresponding equation.
- Different faults affect different combinations of equations. Thus, if one or more equations are not valid, one can look up Table 3 to find out the root cause of the fault.
- One cannot distinguish the difference between level sensor and process faults because they affect the process model in the same way due to the limited degrees of freedom.

On May 29 2003, one alarm was generated by the afore mentioned monitoring scheme. Fig. 5 shows the 7 residuals during that period of time. From Fig. 5, the valve characteristic equation of flow loop #5 was found to be invalid as well as the mass balance residual, while the other valve characteristic equations are still valid. Therefore, according to the fault isolation logic presented in Table 3, the flow sensor F5 was suspected to have measurement problem during that period of time. This online analysis result was also confirmed by process control engineer.

5. CONCLUDING REMARKS

In this paper, the detection and diagnosis of sensor calibration error has been proposed and experimentally evaluated on an industrial caustic tank inventory monitoring process at Millar Western Forest Products Ltd. The proposed method uses a first order model to represent the calibration error and applied optimization routine to perform diagnosis of this problem. The effectiveness of this method has been proved by successfully isolating a mis-calibrated sensor. It is shown that even if additive faults cannot be isolated with the given level of redundancy in a system, it is possible to isolate multiplicative faults, by suitably exploiting the nonlinear effect of these faults. In addition, an online monitoring scheme has been designed and implemented by making use of the process model structure presented in Table 3. The effectiveness of this online monitoring scheme has also been verified.

REFERENCES

Bagajewicz, M. (2000). *Process Plant Instrumentation: Design and Upgrade.* CRC.

Chow, E. and A. Willsky (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE Trans. Auto. Cont.* **29**, 603–614.

Crowe, C. (1996). Data reconciliation - progress and challenges. *J. Proc. Cont.* **6**, 89–98.

Gertler, J. (1988). Survey of model-based failure detection and isolation in complex plants. *IEEE Cont. Sys. Mag.* **12**, 3–11.

Johnson, C. (1988). *Process control instrumentation technology.* 3rd ed.. John Wiley and Sons

Madron, F. (1992). *Process Plant Performance, Measurement Data Processing for Optimization and Retrofits.* Ellis Horwood. West Sussex, England.

Mah, R. (1990). *Chemical Process Structures and Information Flows.* Butterworths. Stoneham, MA, USA.

Narasimhan, S. and C. Jordache (2000). *Data Reconciliation and Gross Error Detection. An Intelligent Use of Process Data.* Gulf Publishing Company.

Narasimhan, S. and R. Mah (1987). Generalized likelihood ratio method for gross error identification. *AIChE J.* **33**, 1514–1521.

Sánchez, M. and J. Romagnoli (2000). *Processing and Reconciliation for Chemical Process Operations.* Academic Press.