



**HAL**  
open science

# Data reconciliation: a robust approach using contaminated distribution. Application in mineral processing

Moustapha Alhaj-Dibo, Didier Maquin, José Ragot

## ► To cite this version:

Moustapha Alhaj-Dibo, Didier Maquin, José Ragot. Data reconciliation: a robust approach using contaminated distribution. Application in mineral processing. 2005. hal-00009005v1

**HAL Id: hal-00009005**

**<https://hal.science/hal-00009005v1>**

Preprint submitted on 22 Sep 2005 (v1), last revised 28 Nov 2006 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Data reconciliation: a robust approach using contaminated distribution. Application in mineral processing

Moustapha Alhaj-Dibo, Didier Maquin<sup>\*</sup>, José Ragot

*Centre de Recherche en Automatique de Nancy – CNRS UMR 7039  
Institut National Polytechnique de Lorraine  
2, Avenue de la forêt de Haye, 54516 Vandœuvre-les-Nancy Cedex, FRANCE*

---

## Abstract

On-line optimisation provides a means for maintaining a process around its optimum operating range. An important component of optimisation relies in data reconciliation which is used for obtaining consistent data. On a mathematical point of view, the formulation is generally based on the assumption that the measurement errors have Gaussian probability density function (pdf) with zero mean. Unfortunately, in the presence of gross errors, all of the adjustments are greatly affected by such biases and would not be considered as reliable indicators of the state of the process. This paper proposes a data reconciliation strategy that deals with the presence of such gross errors. Instead of constructing the objective function to be minimized on the basis of random errors only, the proposed method takes into account both contributions from random and gross errors using a so-called contaminated Gaussian distribution. It is shown that this approach introduces less bias in the estimation due to its natural property to reject gross errors. An academic application to flowrate and concentration data in mineral processing illustrates the efficiency of the proposed method.

*Key words:* Data reconciliation, Robust estimation, Gross error detection, Linear and bilinear mass balances.

---

---

<sup>\*</sup> Corresponding author. Tel.: +33 3 83 59 56 83; fax: + 33 3 83 59 56 44  
*Email address:* Didier.Maquin@ensem.inpl-nancy.fr (Didier Maquin).

## 1 Introduction

The problem of obtaining reliable estimates of the state of a process is a fundamental objective in process control and supervision, these estimates being used to understand the process behaviour. For that purpose, a wide variety of techniques has been developed to perform what is currently known as data reconciliation (Mah, Stanley, Downing, 1976), (Maquin, Bloch, Ragot, 1991). Data reconciliation, which is sometimes referred too as mass and energy balance equilibration, is the adjustment of a set of data so the quantities derived from the data obey physical laws such as material and energy conservation. Since the pionner works devoted to the so-called data rectification (Himmelblau, 1978), the scope of research has expanded to cover other fields such as data redundancy analysis, system observability, optimal sensor positioning, sensor reliability, error characterization, measurement variance estimation. Many applications are related in scientific papers involving various fields in process engineering (Dhurjati, Cauvin, 1999), (Heyen, 1999), (Singh, Mittal, Sen, 2001), (Yi, Kim, Han, 2002).

Unfortunately, the measurement collected on the process may be unknowingly corrupted by gross errors. As a result, the data reconciliation procedure can give rise to absurd results and, in particular, the estimated variables are most often corrupted by these biased data. Several schemes have been suggested to cope with the corruption of normal assumption of the errors, for static systems (Narasimhan, Mah, 1989), (Kim, Kang, Park, Edgar, 1997), (Arora, Biegler, 2001) and also for dynamic systems (Abu-el-zeet, Roberts, Becerra, 2001). Methods for including bounds in process variables to improve gross error detection have also been developed. One major disadvantage of these methods is that they give rise to situations that it may impossible to estimate all the variable by using only a subset of the remaining gross errors free measurements. Alternative approach using constraints both on the estimates and the balance residual equations has been developed for linear system (Ragot, Maquin, Adrot, 1999), (Maquin, Ragot, 2003). There is also an important class of robust estimators whose influence function are bounded allowing to reject outliers (Huber, 1981), (Hampel, Ronchetti, Rousseeuw, Stohel, 1986). Another approach is to take into account the non ideality of the measurement error distribution by using an objective function constructed on contaminated error distribution (Tjoa, Biegler, 1991), (Ozyurt, Pike, 2004). In the following, we adopt and develop this idea for the data reconciliation problem.

The next section is devoted to recall the background of data reconciliation. Robust data reconciliation based on the use of a contaminated error distribution is firstly developed in section 3, for the linear case, and extended to the bilinear

case in the following section. Section 5 shows the adaptation of the method to the frequently encountered case of partial measurements. Finally, the proposed method is implemented on a fictitious but realistic mineral processing plant. Performances of the proposed approach are analyzed and compared with those of a more classical method.

## 2 Data reconciliation background

The classical general data reconciliation problem (Mah, Stanley, Downing, 1976), (Hodouin, Flament, 1989), (Crowe, 1996), deals with a weighted least squares minimisation of the measurement adjustments subject to the model constraints. In order to simplify the presentation, let us first consider a linear steady state process model:

$$Ax = 0, \quad A \in \mathbb{R}^{n.v}, \quad x \in \mathbb{R}^v \quad (1)$$

where  $x$  is the state of the process and  $A$  describes the static constraints with  $\text{rank}(A) = n$ . The measurement devices give the information:

$$\tilde{x} = x + \varepsilon, \quad \varepsilon \sim N(0, V) \quad (2)$$

where  $\varepsilon \in \mathbb{R}^n$  is a vector of random errors characterised by a normal probability density function (pdf) with a diagonal variance matrix  $V$ . For each component  $x_i$  of  $x$ , the following pdf is defined:

$$p(\tilde{x}_i | x_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2} \left(\frac{x_i - \tilde{x}_i}{\sigma_i}\right)^2\right) \quad (3)$$

where  $\sigma_i^2$  are the diagonal elements of  $V$ . From (3) one derives the likelihood function of the observation with the hypothesis of independent realizations. The maximisation of the likelihood function of the observation  $\tilde{x}$ , with regard to  $x$ , and subject to the model constraints (1) leads to the well-known estimate (Maquin, Bloch, Ragot, 1991):

$$\hat{x} = (I - VA^T(AVA^T)^{-1}A)x \quad (4)$$

In fact, the estimates obtained by this method are not always exploitable, the main drawback being the contamination of all estimated values by the outliers which corrupt the measurements. For that reason, robust estimators could be preferred, robustness being the ability to ignore the contribution of extreme data such as gross errors. Two different approaches can be implemented to deal with outliers. The first one consists to sequentially detect, localize and suppress the data which are contaminated and after to reconcile the remaining data. The second approach is a global one and reconcile the data without a

preliminary classification; in fact, weights in the reconciliation procedure are automatically adjusted in order to minimise the influence of the abnormal data. The method presented in this paper is only focused on this last strategy.

### 3 Robust data validation. The linear case.

#### 3.1 Robust estimation

If the measurements contain random outliers, then a single pdf described as in (3) cannot account for the high variance of the outliers. To overcome this problem let us assume that measurement noise is sampled from two pdf, the normal one having a small variance representing regular noise and the abnormal one other having a large variance representing outliers (Wang, Romagnoli, 2002),(Ghosh-Dastider, Schafer, 2003). In a first approach, each measurement  $\tilde{x}_i$  is assumed to have the same normal  $\sigma_1$  and abnormal  $\sigma_2$  standard deviations; this hypothesis will be released later on. Thus, for each observation  $\tilde{x}_i$ , we define the two following pdf ( $j = 1, 2$ ) :

$$p_{j,i}(\tilde{x}_i | x_i, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{x_i - \tilde{x}_i}{\sigma_j}\right)^2\right) \quad (5)$$

The so-called contaminated pdf is then obtained using a combination of these two pdf:

$$p(\tilde{x}_i | x_i, \theta) = wp_{1,i} + (1 - w)p_{2,i}, \quad 0 \leq w \leq 1 \quad (6)$$

where the vector  $\theta$  collects the standard deviations  $\sigma_1$  and  $\sigma_2$ . The quantity  $(1 - w)$  can be seen as an a priori probability of the occurrence of outliers. Assuming the independence of the measurements, the log-likelihood function of the measurement set is then written as:

$$\Phi = \ln \prod_{i=1}^v p(\tilde{x}_i | x_i, \theta) \quad (7)$$

As previously said, the best estimate  $\hat{x}$  (in the maximum likelihood sense) of the state vector  $x$  is obtained by maximizing the log-likelihood function with respect to  $x$  subject to the model constraints:

$$\hat{x} = \arg \max_x \ln \prod_{i=1}^v p(\tilde{x}_i | x_i, \theta) \quad (8a)$$

$$\text{subject to } Ax = 0 \quad (8b)$$

The corresponding Lagrange function associated to this optimization problem is written as follows:

$$\mathcal{L} = \sum_{i=1}^v \ln p(\tilde{x}_i | x_i, \theta) + \lambda^T Ax \quad (9)$$

The partial derivatives of this Lagrange function with regard to the unknown variables  $x$  and  $\lambda$  must therefore be evaluated. It is easy to establish that :

$$\frac{\partial \ln p(\tilde{x}_i | x_i, \theta)}{\partial x_i} = \frac{\frac{w}{\sigma_1^2} p_{1,i} + \frac{1-w}{\sigma_2^2} p_{2,i}}{w p_{1,i} + (1-w) p_{2,i}} (x_i - \tilde{x}_i) \quad (10)$$

Therefore, the estimate  $\hat{x}$  is the solution of the following system:

$$\begin{cases} \left. \frac{\partial \mathcal{L}}{\partial x} \right|_{x=\hat{x}} = W_{\hat{x}}^{-1} (\hat{x} - \tilde{x}) + A^T \lambda = 0 \\ \left. \frac{\partial \mathcal{L}}{\partial \lambda} \right|_{x=\hat{x}} = A \hat{x} = 0 \end{cases} \quad (11)$$

with:

$$W_{\hat{x}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{w}{\sigma_1^2} \hat{p}_{1,i} + \frac{1-w}{\sigma_2^2} \hat{p}_{2,i}}{w \hat{p}_{1,i} + (1-w) \hat{p}_{2,i}} \right) \quad (12a)$$

$$\hat{p}_{j,i} = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left( -\frac{1}{2} \left( \frac{\hat{x}_i - \tilde{x}_i}{\sigma_j} \right)^2 \right) \quad (12b)$$

The notation  $\text{diag}_{i=1..v}(a_i)$  stands for the operator that convert a  $v$ -dimensional vector  $a$ , which entries are  $a_i$ , into a diagonal matrix. Notice that the weighting matrix  $W_{\hat{x}}$  plays a similar role than the variance matrix  $V$  when measurement errors are sampled from a unique (normal) pdf. However, in this last case, the weights are constant as in the proposed approach, the weights depend themselves on the magnitude of the variation  $\hat{x} - \tilde{x}$ . Therefore, it is clear that system (11) is nonlinear with respect to  $\hat{x}$ . However, following the classical approach previously mentioned (see eq. (4)), the solution  $\hat{x}$  is expressed using the following implicit formulation:

$$\hat{x} = (I - W_{\hat{x}} A^T (A W_{\hat{x}} A^T)^{-1} A) \tilde{x} \quad (13)$$

The following direct iterative scheme is proposed for solving this nonlinear system (13):

$$k = 0, \quad x^{(k)} = \tilde{x} \quad (14a)$$

$$\hat{p}_{j,i}^{(k)} = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{\hat{x}_i^{(k)} - \tilde{x}_i}{\sigma_j}\right)^2\right) \quad (14b)$$

$$(W_{\hat{x}}^{(k)})^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{w}{\sigma_1^2} \hat{p}_{1,i}^{(k)} + \frac{1-w}{\sigma_2^2} \hat{p}_{2,i}^{(k)}}{w \hat{p}_{1,i}^{(k)} + (1-w) \hat{p}_{2,i}^{(k)}} \right) \quad (14c)$$

$$\hat{x}^{(k+1)} = \left( I - W_{\hat{x}}^{(k)} A^T (A W_{\hat{x}}^{(k)} A^T)^{-1} A \right) \tilde{x} \quad (14d)$$

A stopping criterion must be chosen for implementing the algorithm; more details will be given in the following section. For sake of simplicity, the proof for the local convergence of the algorithm is omitted.

### 3.2 Weighing function

In order to appreciate how the weights in  $W$ , which should be compared to an influence function as explained in (Hampel, Ronchetti, Rousseeuw, Stohel, 1986), are able to reject the data contaminated by gross errors, figure 1 shows the graph of the function:

$$g(u) = \frac{\frac{w}{\sigma_1^2} p_1 + \frac{1-w}{\sigma_2^2} p_2}{w p_1 + (1-w) p_2}$$

$$p_1 = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{u}{\sigma_1}\right)^2\right)$$

$$p_2 = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{u}{\sigma_2}\right)^2\right)$$

with  $\sigma_1 = 0.5$  and  $\sigma_2 = \{1, 4\}$  and where  $w$  takes the indicated values. For a better comparison, the graphs have been normalized, i.e. we have represented  $\bar{g}(u) = g(u)/g(0)$ . For  $w = 1$ , we naturally obtain a constant weight; thus all the data are equally weighted and, in particular, the optimisation criterion will be sensitive to large magnitude of data, i.e. to outliers. Taking  $w = 0.02$  reduces the influence of outliers. For example, with  $\sigma_2 = 4$ , the weight decreases from 1 for data around the origine to 0.1 for data with large magnitude.

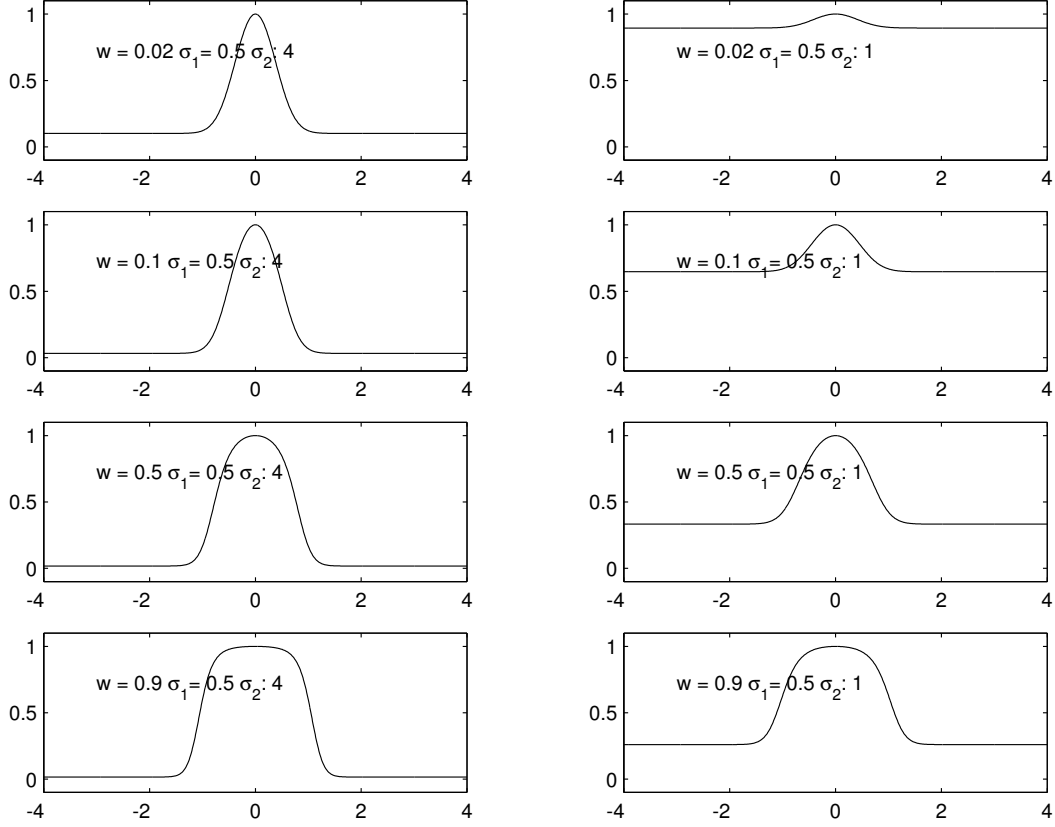


Fig. 1. Influence function

#### 4 Extension to bilinear systems

We consider now the case of a process characterised by two types of variables: macroscopic variables such as flowrates  $x$  and microscopic variables such concentrations or mineral species  $y_c$ ,  $c = 1..q$ . As for the linear case, measurement noise is sampled from two pdf, one having a small variance representing regular noise and the other having a large variance representing outliers. In order to simplify the presentation, each measurement  $x_i$  (resp.  $y_{c,i}$ ) is assumed to have the same normal  $\sigma_{x,1}$  (resp.  $\sigma_{y_c,1}$ ) and abnormal  $\sigma_{x,2}$  (resp.  $\sigma_{y_c,2}$ ) standard-deviations. As previously mentioned, this hypothesis will be withdrawn later on. Thus, for each observation  $\tilde{x}_i$  and  $\tilde{y}_{c,i}$ , we define the following pdf:

$$p(\tilde{x}_i|x_i, \sigma_{x,j}) = \frac{1}{\sqrt{2\pi}\sigma_{x,j}} \exp\left(-\frac{1}{2}\left(\frac{x_i - \tilde{x}_i}{\sigma_{x,j}}\right)^2\right) \quad (15a)$$

$$p(\tilde{y}_{c,i}|y_{c,i}, \sigma_{y_c,j}) = \frac{1}{\sqrt{2\pi}\sigma_{y_c,j}} \exp\left(-\frac{1}{2}\left(\frac{y_{c,i} - \tilde{y}_{c,i}}{\sigma_{y_c,j}}\right)^2\right) \quad (15b)$$

with  $j = 1, 2$ ,  $i = 1..v$ ,  $c = 1..q$ . In the following,  $p_{x,j,i}$  and  $p_{y_c,j,i}$  are shortening notations for  $p(\tilde{x}_i|x_i, \sigma_{x,j})$  and  $p(\tilde{y}_{c,i}|y_{c,i}, \sigma_{y_c,j})$  where indexes  $i$  and  $j$  are



respectively used to point the number of data and the number of the distribution. As for the linear case, the contaminated pdf of the two types of measurements are defined:

$$p_{x,i} = wp_{x,1,i} + (1-w)p_{x,2,i} \quad (16a)$$

$$p_{y_c,i} = wp_{y_c,1,i} + (1-w)p_{y_c,2,i} \quad (16b)$$

In order to simplify the presentation, one used here the same mixture coefficient  $w$  for the two  $x_i$  and  $y_{c,i}$  distributions. Assuming independence of the measurements allows the definition of the global log-likelihood function:

$$\Phi = \ln \prod_{i=1}^v p_{x,i} \prod_{c=1}^q p_{y_c,i} \quad (17)$$

Mass balance constraints for total flowrates and partial flowrates are written using the operator  $\otimes$  used to perform the element by element product of two vectors:

$$Ax = 0 \quad (18)$$

$$A(x \otimes y_c) = 0, \quad c = 1..q \quad (19)$$

Let us now define the optimisation problem consisting in estimating the process variables  $x$  and  $y$ . For that, consider the Lagrange function:

$$\mathcal{L} = \Phi + \lambda^T Ax + \sum_{c=1}^q \mu_c^T A(x \otimes y_c) \quad (20)$$

Constraints are taken into account through the introduction of the Lagrange parameters  $\lambda$  and  $\mu_c, c = 1..q$ . The stationarity conditions of (20) can be expressed as (the estimates are now noted  $\hat{x}$  and  $\hat{y}_c$ ):

$$W_{\hat{x}}^{-1}(\hat{x} - \tilde{x}) + A^T \lambda + \sum_{c=1}^q (A \otimes \hat{y}_c)^T \mu_c = 0 \quad (21a)$$

$$W_{\hat{y}_c}^{-1}(\hat{y}_c - \tilde{y}_c) + (A \otimes \hat{x})^T \mu = 0 \quad (21b)$$

$$A\hat{x} = 0 \quad (21c)$$

$$A(\hat{x} \otimes \hat{y}_c) = 0 \quad (21d)$$

where the weighting matrices  $W_{\hat{x}}$  and  $W_{\hat{y}_c}$  are defined by:

$$W_{\hat{x}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{wp_{\hat{x},1,i} + (1-w)p_{\hat{x},2,i}}{\sigma_{x,1}^2} + \frac{(1-w)p_{\hat{x},2,i}}{\sigma_{x,2}^2}}{wp_{\hat{x},1,i} + (1-w)p_{\hat{x},2,i}} \right) \quad (22a)$$

$$W_{\hat{y}_c}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{wp_{\hat{y}_c,1,i} + (1-w)p_{\hat{y}_c,2,i}}{\sigma_{y_c,1}^2} + \frac{(1-w)p_{\hat{y}_c,2,i}}{\sigma_{y_c,2}^2}}{wp_{\hat{y}_c,1,i} + (1-w)p_{\hat{y}_c,2,i}} \right) \quad (22b)$$

Notice that if each measurement  $x_i$  (resp.  $y_{c,i}$ ) has a particular standard-deviation, formulas (22a) and (22b) still hold by replacing the parameters  $\sigma_{x,1}$  and  $\sigma_{x,2}$  (resp.  $\sigma_{y_{c,1}}$  and  $\sigma_{y_{c,2}}$ ) by  $\sigma_{x,1,i}$  and  $\sigma_{x,2,i}$  (resp.  $\sigma_{y_{c,1,i}}$  and  $\sigma_{y_{c,2,i}}$ ).

System (21) may be directly solved and the solution is expressed as:

$$\hat{x} = (I - W_{\hat{x}} A^T (A W_{\hat{x}} A^T)^{-1} A) (\tilde{x} - W_{\hat{x}} \sum_{c=1}^q A_{\hat{y}_c}^T (A_{\hat{x}} W_{\hat{y}_c} A_{\hat{x}}^T)^{-1} A_{\hat{x}} \tilde{y}_c) \quad (23a)$$

$$\hat{y}_c = (I - W_{\hat{y}_c} A_{\hat{x}}^T (A_{\hat{x}} W_{\hat{y}_c} A_{\hat{x}}^T)^{-1} A_{\hat{x}}) \tilde{y}_c \quad (23b)$$

where the shortening notation  $A_u$  stands for  $A \text{diag}(u)$  (let us remark that  $A(\hat{x} \otimes \hat{y}_c) = A_{\hat{x}} \hat{y}_c = A_{\hat{y}_c} \hat{x}$ ). System (23) is clearly non linear with regard to the unknown  $\hat{x}$  and  $\hat{y}_c$ , the weights  $W_{\hat{x}}$  and  $W_{\hat{y}_c}$  depending on the pdf (15) which themselves depend on the  $\hat{x}$  and  $\hat{y}_c$  estimations (23). In fact (23) is an implicit system in respect to the estimates  $\hat{x}$  and  $\hat{y}_c$  for which we suggest the following iterative scheme:

### Step 1: initialisation

$$k = 0, \hat{x}^{(k)} = \tilde{x}, \hat{y}^{(k)} = \tilde{y}_c$$

Choose  $w$

Adjust  $\sigma_{x,1}$  and  $\sigma_{y_{c,1}}$  from an a priori knowledge about the noise distribution

Adjust  $\sigma_{x,2}$  and  $\sigma_{y_{c,2}}$  from an a priori knowledge about the gross error distribution.

### Step 2: estimation

Compute the quantities (for  $j = 1, 2, i = 1..v$  and  $c = 1..q$ )

$$p_{\hat{x},j,i}^{(k)} = \frac{1}{\sqrt{2\pi}\sigma_{x,j}} \exp\left(-\frac{1}{2} \left(\frac{\hat{x}_i^{(k)} - \tilde{x}_i}{\sigma_{x,j}}\right)^2\right)$$

$$p_{\hat{y}_c,j,i}^{(k)} = \frac{1}{\sqrt{2\pi}\sigma_{y_{c,j}}} \exp\left(-\frac{1}{2} \left(\frac{\hat{y}_{c,i}^{(k)} - \tilde{y}_{ci}}{\sigma_{y_{c,j}}}\right)^2\right)$$

$$W_{\hat{x}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{w p_{\hat{x},1,i}^{(k)}}{\sigma_{x,1}^2} + \frac{(1-w) p_{\hat{x},2,i}^{(k)}}{\sigma_{x,2}^2}}{w p_{\hat{x},1,i}^{(k)} + (1-w) p_{\hat{x},2,i}^{(k)}} \right)$$

$$W_{\hat{y}_c}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{w p_{\hat{y}_c,1,i}^{(k)}}{\sigma_{y_{c,1}}^2} + \frac{(1-w) p_{\hat{y}_c,2,i}^{(k)}}{\sigma_{y_{c,2}}^2}}{w p_{\hat{y}_c,1,i}^{(k)} + (1-w) p_{\hat{y}_c,2,i}^{(k)}} \right)$$

$$A_{\hat{x}}^{(k)} = A \text{diag}(\hat{x}^{(k)}) \quad A_{\hat{y}_c}^{(k)} = A \text{diag}(\hat{y}_c^{(k)})$$

Update the estimates of  $x$  and  $y_c$

$$\hat{x}^{(k+1)} = \left( I - W_{\hat{x}}^{(k)} A^T (A W_{\hat{x}}^{(k)} A^T)^{-1} A \right) \left( \tilde{x} - W_{\hat{x}}^{(k)} \sum_{c=1}^q A_{\hat{y}_c}^{(k)T} (A_{\hat{x}}^{(k)} W_{\hat{y}_c}^{(k)} A_{\hat{x}}^{(k)T})^{-1} A_{\hat{x}}^{(k)} \tilde{y}_c \right)$$

$$\hat{y}_c^{(k+1)} = (I - W_{\hat{y}_c}^{(k)} A_{\hat{x}}^{(k)T} (A_{\hat{x}}^{(k)} W_{\hat{y}_c}^{(k)} A_{\hat{x}}^{(k)T})^{-1} A_{\hat{x}}^{(k)}) \tilde{y}_c$$

### Step 3: convergence test

Compute an appropriate norm of the corrective terms:  $\tau_x^{(k+1)} = \|\hat{x}^{(k+1)} - \tilde{x}\|$  and  $\tau_{y_c}^{(k+1)} = \|\hat{y}_c^{(k+1)} - \tilde{y}_c\|$ . If the variations  $\tau_x^{(k+1)} - \tau_x^{(k)}$  and  $\tau_{y_c}^{(k+1)} - \tau_{y_c}^{(k)}$  are less than a given threshold then stop, else  $k = k + 1$  and go to step 2.

*Remark:* for non linear systems, the initialisation remains a difficult task, convergence of the algorithm being generally sensitive to that choice. In our situation, measurements are a natural choice for initializing the estimates (step 1 of the algorithm). The solution given by classical least squares approach would also provides an acceptable initialization although its sensitivity to gross errors may be sometimes important; the reader should verify that this solution may be obtained by redefining the distributions (16) with  $w = 1$ .

## 5 Extension to partial measurements

Let us now consider the more realistic situation where only some variables are measured. For that purpose, two selection matrices  $H_x$  and  $H_y$  are introduced allowing to define which variables are measured:

$$\tilde{x} = H_x x + \varepsilon_x \quad (24a)$$

$$\tilde{y} = H_y y + \varepsilon_y \quad (24b)$$

Here, only one species or concentration  $y$  was considered. The extension to several species is straightforward. Moreover, in mineral industry, the measurement techniques are such that, generally, all species of a particular stream are analyzed simultaneously leading to the same selection matrix  $H_y$  for all species. Then, the pdf (15a), defined for each measurement  $\tilde{x}_i$ , can be written using a vector form:

$$p_{x,j} = \frac{1}{(2\pi)^{v/2} \sqrt{\det(V_{x,j})}} \exp \left( -\frac{1}{2} (H_x x - \tilde{x})^T V_{x,j}^{-1} (H_x x - \tilde{x}) \right) \quad (25)$$

where  $V_{x,j}$  is the diagonal matrix containing the variances  $\sigma_{x,j}^2$ . A similar expression for  $p_{y,j}$  may be written, that allows the global log-likelihood function

to be expressed:

$$\Phi = \ln (wp_{x,1} + (1 - w)p_{x,2}) (wp_{y,1} + (1 - w)p_{y,2}). \quad (26)$$

Following the same step as previously, the Lagrange function associated to the maximization of (26) subject to the model constraints can be explained. The optimality equations defining the solution are deduced from the derivatives of this Lagrange function with respect to the unknown  $x$ ,  $y$ ,  $\lambda$  et  $\mu$ :

$$H_x^T W_{\hat{x}}^{-1} (H_x \hat{x} - \tilde{x}) + A^T \lambda + (A \otimes \hat{y})^T \mu = 0 \quad (27a)$$

$$H_y^T W_{\hat{y}}^{-1} (H_y \hat{y} - \tilde{y}) + (A \otimes \hat{x})^T \mu = 0 \quad (27b)$$

$$A \hat{x} = 0 \quad (27c)$$

$$A(\hat{x} \otimes \hat{y}) = 0 \quad (27d)$$

In these last expressions, the weight matrices  $W_{\hat{x}}$  and  $W_{\hat{y}}$  were already defined in (22). Of course system (27) can be solved if and only if all the system variables are observable. This observability study is not presented here and we refer the interested reader to the book of Bagajewicz (2000). When this property holds, the system (27) can be transformed into the following implicit system:

$$\hat{x} = \left( G_{\hat{x}} - G_{\hat{x}} A^T (A G_{\hat{x}} A^T)^{-1} A G_{\hat{x}} \right) \left( H_x^T W_{\hat{x}}^{-1} \tilde{x} - A_{\hat{y}}^T (A_{\hat{x}} G_{\hat{y}} A_{\hat{x}}^T)^{-1} A_{\hat{x}} G_{\hat{y}} H_y^T W_{\hat{y}}^{-1} \tilde{y} \right) \quad (28a)$$

$$\hat{y} = (G_{\hat{y}} - G_{\hat{y}} A_{\hat{x}}^T (A_{\hat{x}} G_{\hat{y}} A_{\hat{x}}^T)^{-1} A_{\hat{x}} G_{\hat{y}}) H_y^T W_{\hat{y}}^{-1} \tilde{y} \quad (28b)$$

$$G_{\hat{x}} = (H_x^T W_{\hat{x}}^{-1} H_x + A^T A)^{-1} \quad (28c)$$

$$G_{\hat{y}} = (H_y^T W_{\hat{y}}^{-1} H_y + A_{\hat{x}}^T A_{\hat{x}})^{-1} \quad (28d)$$

Comparing the structures of (28) and (23) allows us to use the iterative scheme of the fourth section. Thus, the same estimation scheme for  $\hat{x}$  and  $\hat{y}$  may be applied either when all the variables or a part of them are measured.

## 6 Example and discussion

It is usual to represent the mass conservation laws of a given phase, species or property using an oriented graph. Such a graph schematizes the flowing of that phase, species or property. A branch in the graph corresponds to a stream and a node to a process equipment or to a group of process equipments. Generally, the basic conservation equations are written in a compact form using the graph incidence matrix.

The method described in section 4 was applied to a fictitious but realistic mineral processing plant which process graph is depicted by figure 2. The considered process involves 16 streams; each of them is characterized by a flowrate and two concentration measurements. All these 48 measurements are corrupted by random noises and some of them are biased by gross errors.

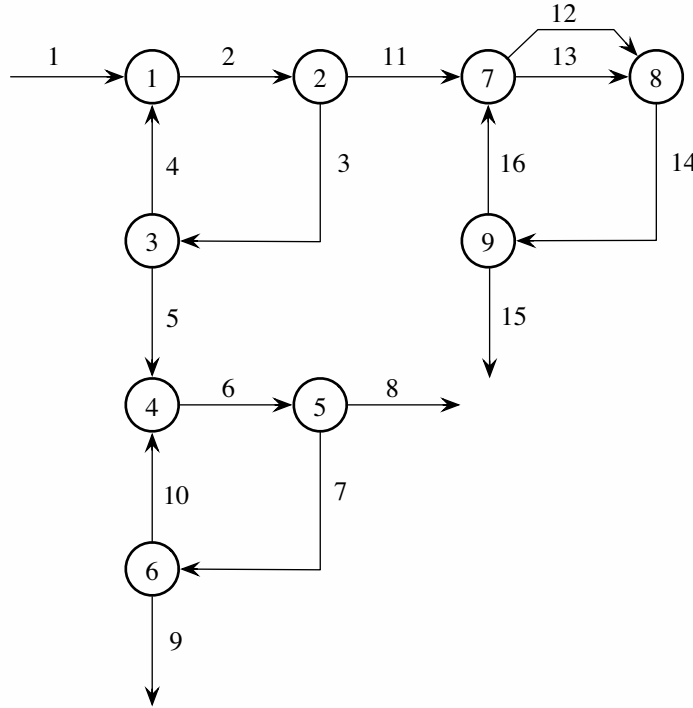


Fig. 2. Oriented graph describing the process flowsheet

The performance results are given when three gross errors (with magnitudes of 6, 8 and 8) affect the measurement 3, 7 and 16; simultaneously, gross errors of magnitude 1.5 affect the measurement of the first concentration for streams 1, 9 and 12, and gross errors of magnitudes 4 and 2.5 affect the measurement of the second concentration for streams 4 and 8. Comparison of the proposed robust least squares algorithm (RLS) with the classical least squares (LS) algorithm is now provided in table 1 where bold characters point out the true values, the RLS and the LS estimates for the variables corrupted by gross errors.

Columns 2 to 4 relate the row measures, columns 5 to 7 show the estimates obtained with RLS and columns 8 to 10 the estimates obtained with LS ; analysing the estimation errors, the RLS estimator clearly allows to suspect variables 3, 7 and 16 for being contaminated by a gross error. Such conclusion is more difficult to express with LS estimator. Table 2 gives explicitly the corrective terms  $\hat{x} - \tilde{x}$  and  $\hat{y}_c - \tilde{y}_c$  for RLS (row 3) and LS (row 4) approaches; for

Table 1  
Measurements and estimates

	Measurement			RLS estimate			LS estimate		
	$x$	$y_1$	$y_2$	$\hat{x}$	$\hat{y}_1$	$\hat{y}_2$	$\hat{x}$	$\hat{y}_1$	$\hat{y}_2$
1	55.88	<b>3.93</b>	3.53	56.50	<b>2.38</b>	3.46	57.30	<b>2.63</b>	3.58
2	65.31	2.73	3.70	65.07	2.71	3.70	65.80	2.89	4.16
3	<b>61.68</b>	2.48	3.54	<b>52.99</b>	2.52	3.58	<b>54.63</b>	2.73	4.16
4	8.38	4.83	<b>9.01</b>	8.57	4.83	<b>5.30</b>	8.50	4.62	<b>8.03</b>
5	44.15	2.13	3.16	44.42	2.07	3.24	46.13	2.38	3.45
6	55.90	2.45	3.57	55.30	2.52	3.51	56.85	2.67	3.68
7	<b>39.05</b>	2.87	3.74	<b>31.44</b>	2.87	3.76	<b>32.97</b>	3.23	3.49
8	23.90	2.05	<b>5.72</b>	23.86	2.05	<b>3.18</b>	23.88	1.91	<b>3.95</b>
9	20.58	<b>3.75</b>	3.34	20.55	<b>2.09</b>	3.32	22.25	<b>2.88</b>	2.91
10	10.33	4.35	4.60	10.89	4.34	4.60	10.72	3.95	4.69
11	12.40	3.57	4.16	12.08	3.54	4.25	11.17	3.68	4.15
12	17.66	<b>5.11</b>	4.32	17.49	<b>3.47</b>	4.27	19.85	<b>3.97</b>	4.30
13	2.66	8.92	6.91	2.27	8.92	6.90	2.71	8.45	6.90
14	19.38	4.18	4.66	19.76	4.10	4.57	22.56	4.51	4.61
15	12.42	3.45	4.24	12.08	3.54	4.25	11.17	3.68	4.15
16	<b>14.77</b>	4.95	5.01	<b>7.68</b>	4.98	5.07	<b>11.39</b>	5.33	5.06

a better comparison, row 2 indicates the true value of the gross error and thus we can appreciate the vicinity of the corrective terms obtained from RLS with the “true” gross errors, that is not always the case when using the LS method.

Table 2  
Corrective terms

	$x_3$	$x_7$	$x_{16}$	$y_{1,1}$	$y_{1,9}$	$y_{1,12}$	$y_{2,4}$	$y_{2,8}$
T	6.00	8.00	8.00	1.50	1.50	1.50	4.00	2.50
RLS	8.69	7.61	7.09	1.55	1.69	1.64	3.71	2.54
LS	7.05	6.08	3.38	1.3	0.87	1.14	0.98	1.77

For another data set, figure 3 visualizes more clearly the estimation errors ( $\hat{x} - \tilde{x}$  and  $\hat{y}_c - \tilde{y}_c$ ) both for RLS (upper part) and LS (lower part). On each graph, horizontal and vertical axis are respectively scaled with the number of the data and the magnitude of the absolute estimation error; the dashed

horizontal line is the threshold chosen to detect abnormal corrective terms. Analysing figure 3 shows two advantages on RLS upon LS approach: first, the corrective terms are more precisely estimated, second, the scattering of the gross errors is less (the corrective terms mainly affect the variables affected by the gross errors and not the others).

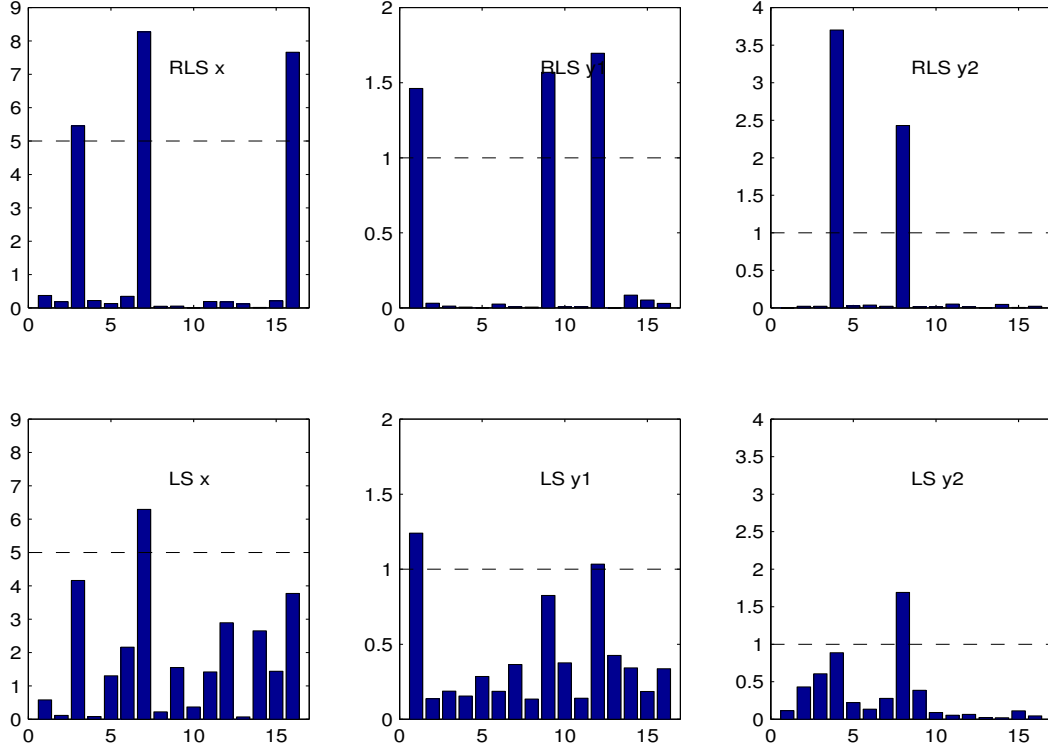


Fig. 3. Corrective terms

Performances of the proposed approach can be also analysed when using a great number of data. For that purpose, the same process has been used with different additive random noise on the data, the gross errors being superposed to the same data as previously. 10000 runs have been performed, allowing to enumerate the cases where the gross errors have been correctly detected or not, both for RLS and LS method. Results, expressed in percentage, are shown in table 3. Roughly speaking, for the given example, the ability of gross error detection for RLS is twice of those of LS. This has been confirmed by many other runs involving various distributions of the measurement errors.

Of course, the choice of the tuning parameters  $w$ ,  $\sigma_{x,i}$  and  $\sigma_{y_c,j}$  of the contaminated distribution affects the detection and the estimation of outliers and therefore requires special attention. In fact, due to the structure of the function defining the weight, we can reduce these parameters to  $w$ ,  $\sigma_{x,1}/\sigma_{x_2}$  and  $\sigma_{y_c,1}/\sigma_{y_c,2}$ . Table 4 presents some results of sensitivity, expressed in percentage of correct detection, using the same process. For each result of detection,

Table 3  
Correct fault detection in %

Var.	RLS gross error detection			LS gross error detection		
	$x$	$y_1$	$y_2$	$x$	$y_1$	$y_2$
$w=0.10$	92.5	99.9	91.4	41.4	57.2	55.2

concerning a particular value of  $w$ ,  $f = \sigma_{x,1}/\sigma_{x_2} = \sigma_{y_c,1}/\sigma_{y_c,2}$ , 10000 runs have been performed, each run having the same outliers but specific random noise affecting the measurements. It should be noted that almost all gross errors were correctly detected with a proper choice of the parameters  $w$  and  $f$ , excepted the error on the flowrate  $x_3$ . For this particular example (but this situation were noticed for almost all the treated examples), one observes that a ‘large’ range of parameter values are acceptable; that implies a low sensitivity of the obtained results with regard to the tuning parameters. However, it is also possible to use an adaptive algorithm for the adjusting of these parameters.

Table 4  
Performance of the approach

$w$	$f$	$x_3$	$x_7$	$x_{16}$	$y_{1,1}$	$y_{1,9}$	$y_{1,12}$	$y_{2,4}$	$y_{2,8}$
0.02	25	1	1	1	100	1	1	1	99
0.05	25	1	1	0	100	0	1	1	100
0.30	25	5	45	3	100	0	2	0	100
0.02	50	72	99	99	100	75	100	41	100
0.05	50	77	100	100	100	96	100	59	100
0.30	50	76	100	100	100	100	100	90	100
0.02	75	76	100	100	100	100	100	99	100
0.05	75	76	100	100	100	100	100	100	100
0.30	75	76	100	99	100	100	100	100	100

A last example is presented to analyse false detection. The same flowsheet is used but there are now three gross errors for each variable, for  $x$  on streams 3, 7 and 16 with magnitudes 6, 8 and 8, for  $y_1$  on streams 1, 9 and 12 with the same magnitudes 1.5, and for  $y_2$  on streams 5, 8 and 14 with the same magnitudes 2. 10000 runs have been performed and the results are summarized in figure 4; detection have been established with the following thresholds : 3, 0.75 and 0.75 respectively for corrective terms on  $x$ ,  $y_1$  and  $y_2$ .

Although the data have been changed, the previous conclusion are still valid



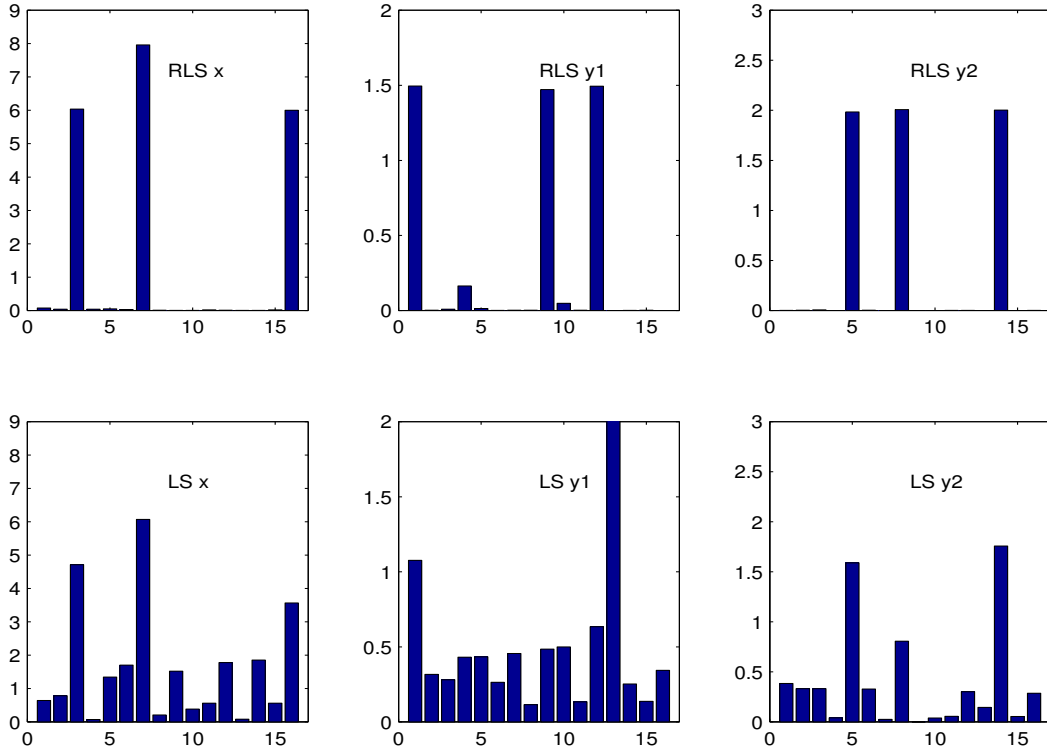


Fig. 4. Corrective terms

and the advantage of RLS against LS is obvious. To be more precise, the number of detected gross errors for the 48 variables are gathered in table 5. Results are expressed in percentages and, for example, 98 percent of the faults affecting variable  $y_{1,9}$  are correctly detected. False detection are also indicated and, for example, the fault-free variable  $y_{2,11}$  has been detected as faulty with a score of 4 percent. The comparison between RLS and LS reveals three major points: fault on variable  $y_{1,12}$  has never been detected with LS, false detection occurs on variable  $y_{1,13}$  with a score of 100 percent, fault on variable  $y_{2,8}$  is detected with a score of 67 percent. Although the results are issued from a particular flowsheet, they clearly show the advantage of the RLS approach.

## 7 Conclusion

To deal with the issues of gross errors influence on data estimation, the paper has presented a robust reconciliation approach. For that purpose, we use a cost function which is less sensitive to the outlying observations than that of least squares. The algorithm can handle multiple biases or outliers at a time and for the given example, 8 outliers have been correctly detected on 48 variables. The implementation of the algorithm is easy, the only parameters to adjust being the variances of the errors (or a ratio between normal and abnormal variance errors), the mixing parameter of the two distributions and

Table 5  
Percentages of detected gross errors

	$x$	$y_1$	$y_2$	$x$	$y_1$	$y_2$
	RLS	RLS	RLS	LS	LS	LS
1	0	<b>99</b>	0	0	<b>100</b>	1
2	0	0	0	0	0	0
3	<b>100</b>	0	0	<b>100</b>	0	2
4	0	10	0	0	11	0
5	0	0	<b>100</b>	0	6	<b>100</b>
6	0	0	0	0	0	19
7	<b>100</b>	1	0	<b>100</b>	28	0
8	0	1	<b>99</b>	0	1	<b>67</b>
9	0	<b>98</b>	0	0	<b>40</b>	0
10	0	1	0	0	55	0
11	0	0	4	0	0	0
12	0	<b>100</b>	2	0	<b>0</b>	6
13	0	0	0	0	100	0
14	0	0	<b>100</b>	3	0	<b>100</b>
15	0	0	0	0	0	1
16	<b>100</b>	0	0	<b>100</b>	0	7

the detection threshold.

The results of reconciliation will clearly depend not only on the data, but also on the model of the process itself. As a perspective of development of robust reconciliation strategies, there is a need for taking account of model uncertainties and optimise the balancing parameter  $w$ . Moreover, for process with unknown parameters, it should be important to jointly estimate the reconciled data and the process parameters.

## References

Abu-el-zeet, Z.H., Roberts, P.D. and Becerra, V.M. (2001). Bias detection and identification in dynamic data reconciliation. In *European Control Conference*, Porto, Portugal, September 4-7.

- Arora, N. and Biegler, L.T. (2001). Redescending estimator for data reconciliation and parameter estimation. *Computers and Chemical Engineering*, 25 (11/12), 1585-1599.
- Bagajewicz, M.J. *Process plant instrumentation: design and upgrade*. Lancaster, PA: Technomic Publishing Company, 2000.
- Crowe, C.M. (1996). Data reconciliation - progress and challenges. *Journal of Process Control*, 6 (2/3), 89-98.
- Dhurjati, P. and Cauvin, S. (1999). On-line fault detection and supervision in the chemical process industries. *Control Engineering Practice*, 7 (7), 863-864.
- Ghosh-Dastider, B. and Schafer, J.L. (2003). Outlier detection and editing procedures for continuous multivariate data. Working paper 2003-07, RAND, Santa Monica OPR, Princeton University.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stohel, W.A. (1986). *Robust statistic: the approach on influence functions*. Wiley, New-York.
- Heyen, G. (1999). Industrial applications of data reconciliation: operating closer to the limits with improved design of the measurement system. In *Workshop on Modelling for Operator Support*, Lappeenranta University of Technology, June 22.
- Himmelblau, D.M. (1978). *Fault detection and diagnosis in chemical and petrochemical processes*. Elsevier Scientific Pub. Co.
- Hodouin, D. and Flament, F. (1989). New developments in material balance calculations for mineral processing industry. In *Society of Mining Engineers annual meeting*, Las Vegas, February 27 - March 2.
- Huber, P.J. (1981). *Robust statistic*. John Wiley & Sons, New York.
- Kim, I., Kang, M.S., Park, S. and Edgar T.F. (1997). Robust data reconciliation and gross error detection: the modified MIMT using NLP. *Computers and Chemical Engineering*, 21, 775-782.
- Mah, R.S.H., Stanley, G.M. and Downing, D. (1976). Reconciliation and rectification of process flow and inventory data. *Ind. Eng. Chem., Process Des. Dev.*, 15 (1), 175-183.
- Maquin, D., Bloch, G. and Ragot, J. (1991). Data reconciliation for measurements. *European Journal of Diagnosis and Safety in Automation*, 1 (2), 145-181.
- Maquin, D. and Ragot, J. (2003). Validation de données issues de systèmes de mesure incertains. *Journal Européen des Systèmes Automatisés*, 37 (9), 1163-1179.
- Narasimhan, S. and Mah, R.S.H. (1989). Treatment of general steady state process models in gross error identification. *Computers and Chemical Engineering*, 13 (7), 851-853.
- Ozyurt, D.B. and Pike R.W. (2004). Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. *Computers and Chemical Engineering*, 28, 381-402.
- Ragot, J., Maquin, D. and Adrot, O. (1999). LMI approach for data reconciliation. In *38th Conference of Metallurgists, Symposium Optimization and*

- Control in Minerals, Metals and Materials Processing*, Quebec, Canada, August 22-26.
- Singh, S.R., Mittal, N.K. and Sen, P.K. (2001). A novel data reconciliation and gross error detection tool for the mineral processing industry. *Minerals Engineering*, 14 (7), 808-814.
- Tjoa, I.B. and Biegler L.T. (1991). Simultaneous strategy for data reconciliation and gross error analysis. *Computer and Chemical Engineering*, 15 (10), 679-689.
- Wang, D. and Romagnoli, J.A. (2002). Robust data reconciliation based on a generalized objective function. In *15th IFAC World Congress on Automatic Control*, Barcelona, Spain, July 21-26.
- Yi, H.-S., Kim, J. H. and Han, C. (2002). Industrial application of gross error estimation and data reconciliation to byproduction gases in iron and steel making plants. In *International Conference on Control, Automation and Systems*, Muju, Korea, October 16-19.