



**HAL**  
open science

## Data reconciliation: a robust approach using contaminated distribution.

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.. 2006. hal-00009005v2

**HAL Id: hal-00009005**

**<https://hal.science/hal-00009005v2>**

Preprint submitted on 18 Sep 2006 (v2), 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 a contaminated distribution

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

*Institut National Polytechnique de Lorraine  
Centre de Recherche en Automatique de Nancy – UMR 7039 CNRS-UHP-INPL  
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. This optimisation heavily relies on process measurements and accurate process models. However, these measurements often contain random and possibly gross errors as a result of miscalibration or failure of the measuring instruments. Thus, these measurements do not generally satisfy the process constraints and they need to be reconciled. In the data reconciliation process, data are adjusted to satisfy the process constraints while minimizing the error in the least squares sense. On a mathematical point of view, data reconciliation is generally based on the assumption that the measurement errors have Gaussian probability density function (pdf) with zero mean and known variance. The elimination of the less frequent gross errors is achieved by so called gross error detection. Therefore, simultaneous data reconciliation and gross error detection have emerged as a key issue of on-line optimization. Indeed, in the presence of gross errors, all of the estimates provided by classical data reconciliation methods 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 extracted 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).

Process measurements are inevitably corrupted by errors during the measurement itself but also during its processing and transmission stages. The total error in a measurement, which is the difference between the measured value and the (definitely unknown) value of a variable, can be conveniently represented as the sum of the contributions from two types of errors : random and gross errors. Random errors which are inherent to the measurement process are usually small in magnitude and are most often described by the use of probability distributions. On the other hand, gross errors are caused by non random events such as instrument malfunctioning, miscalibration, wear or corrosion of sensors and so on. The nonrandom nature of these errors implies that at any given time they have a certain magnitude and sign which may be unknown. Thus, if the measurement is repeated with the same instrument under identical conditions, the contribution of a systematic gross error to the measurement value will be the same (Narasimhan, Jordache, 2000). It is the reason why gross errors are also called systematic errors or biases. In the sequel of the paper all these terms will be employed without distinction. However, despite this strict definition of gross errors, a means for elaborating robust estimators of process variables consists to consider gross errors as random variables. For Veverka, Madron (1997), "Statistically, a gross error is an error whose occurrence as realization of a random variable is highly unlikely". This trick will be clearly justified in the section 3.2 dedicated to the use of a contaminated distribution.

As previously said the measurements 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 lead to situations where it may be 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). Johnston, Kramer (1995) established an analogy between maximum likelihood estimation and robust regression and they reported the feasibility and better performance of the robust estimators as the objective function in the data reconciliation problem when the data contain gross errors. These studies have been based on robust statistics and their ability to reject outliers (Huber, 1981), (Hampel, Ronchetti, Rousseeuw, Stohel, 1986). This kind of approach will be recalled in the first part of section 3. 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 based on the assumption that all the measurements are normally distributed. Robust data reconciliation method is described in section 3. In a first part, the main idea is introduced based on the use of the so-called generalized maximum likelihood function. The second part presents the proposed method which make use of a contaminated error distribution. The previous results are established in the context of linear models, so the next section extends the method to the bilinear case. 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  $\text{rank}(A) = n$  according to the fact that the constraints are independent. 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, in that particular case, to minimize a classical least squares objective function:

$$\min_x \phi = \frac{1}{2} \sum_{i=1}^v \left(\frac{x_i - \tilde{x}_i}{\sigma_i}\right)^2 = \frac{1}{2} \|x - \tilde{x}\|_{V^{-1}}^2 \quad (4)$$

Using Lagrange multiplier method for solving this quadratic optimization problem subject to equality constraints leads to the following system:

$$\begin{cases} V^{-1}(\hat{x} - \tilde{x}) + A^T \lambda = 0 \\ A\hat{x} = 0 \end{cases} \quad (5)$$

where  $\lambda \in \mathbb{R}^n$  is the so-called Lagrange parameter vector. Solving this system of equations gives the well-known following estimate (Kuehn, Davidson, 1961), (Maquin, Bloch, Ragot, 1991):

$$\hat{x} = (I - VA^T(AVA^T)^{-1}A)\tilde{x} \quad (6)$$

In fact, the estimates obtained by this method are not always exploitable, the main drawback being the contamination of all estimated values by the gross errors 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 gross errors. The first one consists to sequentially detect, localize and suppress the data which are contaminated and after to reconcile the remaining data. Most of these methods involve the use of statistical tests

based on the assumption that the random errors in the data are normally distributed. In one of the simplest methods, the set of residuals from the least squares procedure is tested for outliers and any measurement for which the corresponding residual fails the test is considered to contain a gross error. Serth, Heenan (1986) showed that superior performance is obtained if the test for outliers is applied in an iterative way. At each stage of the iteration, only the measurement corresponding to the greater outlier is identified as containing a gross error and removed from the data set. The procedure is stopped when all remaining residuals satisfy the test for outliers. 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 Generalized maximum likelihood estimation

Clearly, the least squares objective function in the previous formulation comes from the assumption that all the measurements are normally distributed, without taking into account gross errors that may be present. The influence of these gross errors on the estimates can be minimized by defining robust objective functions. For example, let us consider the following so-called generalized maximum likelihood objective function proposed by Huber (Huber, 1981):

$$\phi = \sum_{i=1}^v \rho \left( \frac{x_i - \tilde{x}_i}{\sigma_i} \right) \quad (7)$$

In this last expression,  $\rho$  is any reasonable monotone function of the standard error provided that the gross errors have reduced effect on the estimation of process variables. We will examine later how such objective functions can be built, but now let us establish the solution of the optimization problem with this new objective function.

Let us define the influence function  $\Psi(u)$  and the weight function  $w(u)$ :

$$\Psi(u) = \frac{\partial \rho(u)}{\partial u} \quad \text{and} \quad w(u) = \frac{\Psi(u)}{u} \quad (8)$$

The minimization of the objective function (7) with regard to  $x$  and subject

to (1) leads to following system of equations:

$$\begin{cases} \left. \frac{\partial \mathcal{L}}{\partial x_i} \right|_{x_i=\hat{x}_i} = w_i(\hat{x}_i - \tilde{x}_i) + \sum_{j=1}^n a_{ji}\lambda_j = 0 & i = 1, \dots, v \\ \left. \frac{\partial \mathcal{L}}{\partial \lambda_j} \right|_{x_i=\hat{x}_i} = \sum_{i=1}^v a_{ji}\hat{x}_i = 0 & j = 1, \dots, n \end{cases} \quad (9)$$

where  $w_i$  is a simplified notation for  $w(\hat{x}_i - \tilde{x}_i)$ . Let us now define the following matrix:

$$W_{\hat{x}} = \text{diag}_{i=1..v} \left( \frac{1}{w_i} \right) \quad (10)$$

where 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.

Using this notation, the system (9) can be written in the following more compact form:

$$\begin{cases} W_{\hat{x}}^{-1}(\hat{x} - \tilde{x}) + A^T \lambda = 0 \\ A\hat{x} = 0 \end{cases} \quad (11)$$

The two systems (5) and (11) can be easily compared; in (5) the weight matrix  $V^{-1}$  is constant though, in (11),  $W_{\hat{x}}^{-1}$  is depending on the adjustments  $\hat{x} - \tilde{x}$ . The equation system (5) is linear with regard to the unknown  $\hat{x}$ , that's why the solution (6) is expressed by a closed form. Unfortunately, it is not the case for the system (11) which is nonlinear. However, following the classical approach previously mentioned (see eq. (6)), the solution  $\hat{x}$  is expressed using the following implicit formulation:

$$\hat{x} = (I - W_{\hat{x}} A^T (A W_{\hat{x}} A^T)^{-1} A) \tilde{x} \quad (12)$$

The reader will notice that (12) has the general form  $\varphi = f(\varphi)$  and therefore may be solved by a fixed point iteration algorithm for which the fixed point theorem forecasts the convergence conditions (Border, 1985), (Hoffman, 2001). The solution can therefore be obtained in an iterative way, updating, at each step, the weight matrix  $W_{\hat{x}}$ . For the  $k^{\text{th}}$  step of this iterative calculus, let us denote  $\hat{x}^{(k)}$  and  $W_{\hat{x}}^{(k)}$  the estimate and the corresponding weight matrix. Starting with  $\hat{x}^{(0)} = \tilde{x}$ , the solution is then expressed, at the  $(k+1)^{\text{th}}$  step, as:

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

The stopping rule of this calculus can be conditioned by a test on the norm of the difference between two consecutive estimates. If  $\|\hat{x}^{(k+1)} - \hat{x}^{(k)}\| \leq \varepsilon$  where  $\varepsilon$  is fixed by the user, the calculus is stopped and the final estimate is equal to  $\hat{x}^{(k)}$ .

Of course the choice of the objective function introduced in (7) completely conditions the results of the proposed method. As previously said, if  $\rho$  is a quadratic function of the standard error corresponding to a least squares objective function, all the adjustments influence linearly the criterion (7); that explains the lack of robustness with regard gross errors of the classical least squares estimation. A robust estimator should not be influenced a lot when gross errors occur. Therefore, an important property can be deduced from that remark: the influence function must be bounded. In fact, the choice for the influence function (or equivalently for the corresponding objective function) is very large. The reader is referred to the paper by Özyurt (Ozyurt, Pike, 2004) for a comparison of their respective performances.

To explain more the role played by this function, let us consider now a family of Cauchy functions parameterized by  $\alpha$ :

$$\rho(u) = \frac{\alpha^2}{2} \log \left( 1 + \left( \frac{u}{\alpha} \right)^2 \right) \quad (14)$$

From (8), it is easy to obtain the associated weight function:

$$w(u) = \frac{1}{1 + \left( \frac{u}{\alpha} \right)^2} \quad (15)$$

This weight function can be generalized in order to control the slope of its decreasing by substituting the power of the term  $u/\alpha$  by any positive integer  $p$ :

$$w(u) = \frac{1}{1 + \left( \frac{u}{\alpha} \right)^p} \quad (16)$$

In that case, the associated objective function is much more complex. However, the reader will notice that its expression is not necessary in the considered context. Figure 1 shows the look of the influence function and the weight function for some value of  $p$  with  $\alpha = 2$ .

For  $p = 1$ , the influence function is increasing but for  $p > 1$ , that function increases until a particular value and next decreases. This property explains the ability to limit the influence of gross errors onto the estimates. The  $\alpha$  parameter determines the intermediate zone between increasing and decreasing of the influence function. The parameter  $\alpha$  must be chosen by the user depending on an a priori knowledge about the magnitude of the errors (considered as abnormal) which corrupt the measurements.

The ability to limit the influence of gross errors can also be explained looking at the weight function. Indeed, the more the estimation procedure applies a



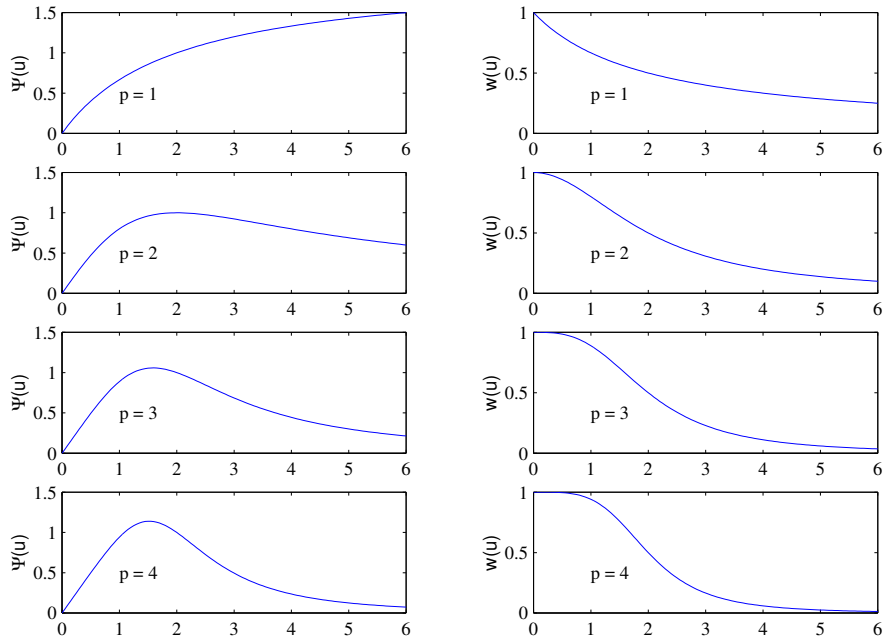


Fig. 1. Influence and weight functions.

large adjustment to the measurement of a variable in order to satisfy the process model, the less the weight of this adjustment influences the criterion to be optimized. In fact, this idea is very close to the classical so-called Iterative Measurement Test (IMT) (Serth, Heenan, 1986) dedicated to the serial elimination of measurements containing gross errors. Indeed this strategy identifies, at each stage of an iterative procedure, the measurement corrupted by a gross error (generally based on a threshold applied on the adjustment) before removing it from the data set. Of course, after that elimination, the suspected measurement do not influence the residual criterion at all as the corresponding variable is then considered as unknown. In the proposed method, the elimination of measurements corrupted by gross errors is done through the adaptation of their weights in the criterion to be optimized. The elimination is also done through an iterative procedure, but at a particular step of calculus, the weights of all the adjustments are updated taking into account their magnitudes.

### 3.2 Estimation based on a contaminated distribution

Another means to be unaware of the presence of gross errors consists to take them into account a priori in the error probability distribution. A distribution based on the additive combination of two Gaussian distributions can be used. Indeed, if the measurements contain gross errors, then a single pdf described as in (3) cannot account for their occurrences. 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 ( $j = 1, 2$ ):

$$p_{j,i}(\tilde{x}_i | x_i, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{x_i - \tilde{x}_i}{\sigma_j}\right)^2\right) \quad (17)$$

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

$$p(\tilde{x}_i | x_i, \theta) = \eta p_{1,i} + (1 - \eta) p_{2,i}, \quad 0 \leq \eta \leq 1 \quad (18)$$

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

$$\Phi = \ln \prod_{i=1}^v p(\tilde{x}_i | x_i, \theta) \quad (19)$$

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:

$$\hat{x} = \arg \max_x \ln \prod_{i=1}^v p(\tilde{x}_i | x_i, \theta) \quad (20a)$$

$$\text{subject to } Ax = 0 \quad (20b)$$

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

$$\mathcal{L} = \sum_{i=1}^v \ln p(\tilde{x}_i | x_i, \theta) + \lambda^T Ax \quad (21)$$

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:

$$\frac{\partial \ln p(\tilde{x}_i | x_i, \theta)}{\partial x_i} = \frac{\frac{\eta}{\sigma_1^2} p_{1,i} + \frac{1-\eta}{\sigma_2^2} p_{2,i}}{\eta p_{1,i} + (1-\eta) p_{2,i}} (x_i - \tilde{x}_i) \quad (22)$$

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

$$\begin{cases} \left. \frac{\partial \mathcal{L}}{\partial x} \right|_{x=\hat{x}} = W_{\hat{x}}^{-1}(\hat{x} - \tilde{x}) + A^T \lambda = 0 \\ \left. \frac{\partial \mathcal{L}}{\partial \lambda} \right|_{x=\hat{x}} = A\hat{x} = 0 \end{cases} \quad (23)$$

with:

$$W_{\hat{x}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{\eta}{\sigma_1^2} \hat{p}_{1,i} + \frac{1-\eta}{\sigma_2^2} \hat{p}_{2,i}}{\eta \hat{p}_{1,i} + (1-\eta) \hat{p}_{2,i}} \right) \quad (24a)$$

$$\hat{p}_{j,i} = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left( -\frac{1}{2} \left( \frac{\hat{x}_i - \tilde{x}_i}{\sigma_j} \right)^2 \right) \quad (24b)$$

Clearly, the obtained system of equations defining the estimate (23) is exactly the same as that obtained previously using the generalized maximum likelihood objective function (11). The only difference concerns the definition of the weight matrix  $W_{\hat{x}}^{-1}$ . Following the resolution method introduced in the previous section, the following direct iterative scheme is proposed for solving this nonlinear system (23):

$$k = 0, \quad x^{(k)} = \tilde{x} \quad (25a)$$

$$\hat{p}_{j,i}^{(k)} = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left( -\frac{1}{2} \left( \frac{\hat{x}_i^{(k)} - \tilde{x}_i}{\sigma_j} \right)^2 \right) \quad (25b)$$

$$(W_{\hat{x}}^{(k)})^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{\eta}{\sigma_1^2} \hat{p}_{1,i}^{(k)} + \frac{1-\eta}{\sigma_2^2} \hat{p}_{2,i}^{(k)}}{\eta \hat{p}_{1,i}^{(k)} + (1-\eta) \hat{p}_{2,i}^{(k)}} \right) \quad (25c)$$

$$\hat{x}^{(k+1)} = (I - W_{\hat{x}}^{(k)} A^T (A W_{\hat{x}}^{(k)} A^T)^{-1} A) \tilde{x} \quad (25d)$$

Clearly, representing gross errors by random variables with centered probability density functions can appear as a complete contradiction with the definition of what a gross error is. If so, gross errors are then only considered as normal errors with high variance and low probability. However, this trick that has been already used by other authors, even if it don't really describe the situation and it allows the construction of a robust estimator. The main idea is already to give a small weight, in the criterion to be minimized, to the large adjustments (differences between estimate and measurement). Indeed, this approach limits the ‘‘smearing’’ caused by the occurrence of a gross error on the estimates of other process variables.

Let us now analyze the behavior of the weight function introduced in (22) and explain how the proposed estimation method is able to reject the data

contaminated by gross errors. Figure 2 shows the graph of the function:

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

with:

$$p_i = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2} \left(\frac{u}{\sigma_i}\right)^2\right) \quad i = 1, 2$$

The weight functions  $w(u)$  were evaluated for  $\sigma_1 = 0.5$  and  $\sigma_2 \in \{1, 5\}$  (corresponding to the ratios  $r = \sigma_1/\sigma_2 \in \{2, 10\}$ ) and  $\eta \in \{0.5, 0.8, 0.9, 0.99\}$ . For a better comparison, the graphs have been normalized, i.e. we have represented  $g(u) = w(u)/w(0)$ . The behaviour of this function is very similar to that derived from generalized likelihood objective function (see fig. 1) therefore the proposed method will have the same ability to reject gross errors. Analysing these functions, one can observe that, provided the standard deviation ratio  $r$  is greater than 10, the weight function tends to zero for sufficiently large adjustments. One can also notice that the contamination ratio  $\eta$  influences the width of the zone for which the normalized weight function  $g(u)$  is approximately equal to one.

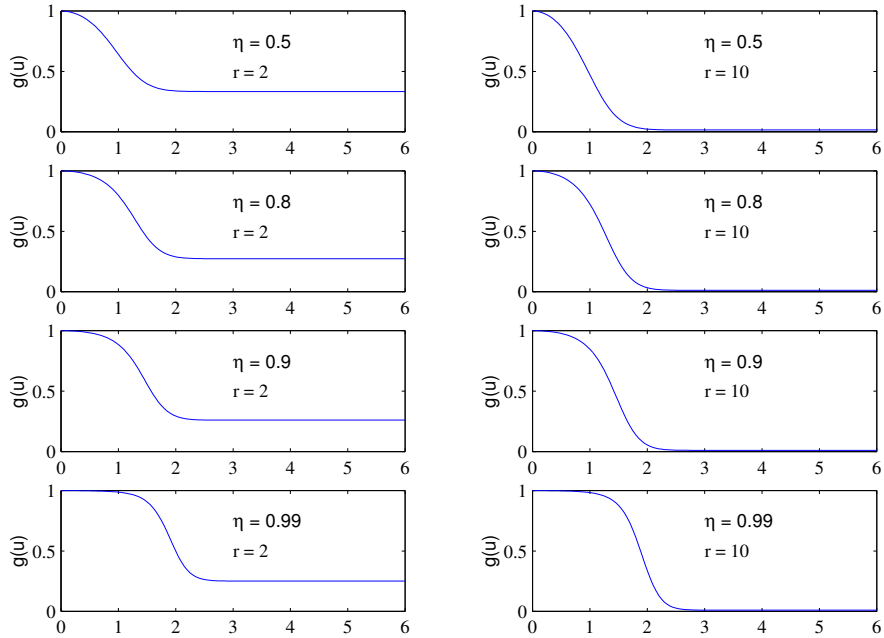


Fig. 2. Weight functions

## 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 biases. 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:

$$p(\tilde{x}_i|x_i, \sigma_{x,j}) = \frac{1}{\sqrt{2\pi}\sigma_{x,j}} \exp\left(-\frac{1}{2}\left(\frac{x_i - \tilde{x}_i}{\sigma_{x,j}}\right)^2\right) \quad (26a)$$

$$p(\tilde{y}_{c,i}|y_{c,i}, \sigma_{y_{c,j}}) = \frac{1}{\sqrt{2\pi}\sigma_{y_{c,j}}} \exp\left(-\frac{1}{2}\left(\frac{y_{c,i} - \tilde{y}_{c,i}}{\sigma_{y_{c,j}}}\right)^2\right) \quad (26b)$$

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:

$$p_{x,i} = \eta_1 p_{x,1,i} + (1 - \eta_1) p_{x,2,i} \quad (27a)$$

$$p_{y_{c,i}} = \eta_2 p_{y_{c,1},i} + (1 - \eta_2) p_{y_{c,2},i} \quad (27b)$$

In order to simplify the presentation, all the distributions describing the  $y_{c,i}$  variables have been chosen identical i.e. with the same contamination ratio  $\eta_2$ . Assuming independence of the measurements allows the definition of the global log-likelihood function:

$$\Phi = \ln \prod_{i=1}^v \left( p_{x,i} \prod_{c=1}^q p_{y_{c,i}} \right) \quad (28)$$

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:

$$Ax = 0 \quad (29)$$

$$A(x \otimes y_c) = 0, \quad c = 1..q \quad (30)$$

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

$$\mathcal{L} = \Phi + \lambda^T Ax + \sum_{c=1}^q \mu_c^T A(x \otimes y_c) \quad (31)$$

Constraints are taken into account through the introduction of the Lagrange parameters  $\lambda$  and  $\mu_c, c = 1..q$ . The stationarity conditions of (31) can be expressed as (the estimates are now noted  $\hat{x}$  and  $\hat{y}_c$ ):

$$W_{\hat{x}}^{-1}(\hat{x} - \tilde{x}) + A^T \lambda + \sum_{c=1}^q (A \otimes \hat{y}_c)^T \mu_c = 0 \quad (32a)$$

$$W_{\hat{y}_c}^{-1}(\hat{y}_c - \tilde{y}_c) + (A \otimes \hat{x})^T \mu_c = 0 \quad (32b)$$

$$A\hat{x} = 0 \quad (32c)$$

$$A(\hat{x} \otimes \hat{y}_c) = 0 \quad (32d)$$

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

$$W_{\hat{x}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{\eta_1 p_{\hat{x},1,i}}{\sigma_{x,1}^2} + \frac{(1-\eta_1) p_{\hat{x},2,i}}{\sigma_{x,2}^2}}{\eta_1 p_{\hat{x},1,i} + (1-\eta_1) p_{\hat{x},2,i}} \right) \quad (33a)$$

$$W_{\hat{y}_c}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{\eta_2 p_{\hat{y}_c,1,i}}{\sigma_{y_c,1}^2} + \frac{(1-\eta_2) p_{\hat{y}_c,2,i}}{\sigma_{y_c,2}^2}}{\eta_2 p_{\hat{y}_c,1,i} + (1-\eta_2) p_{\hat{y}_c,2,i}} \right) \quad (33b)$$

Notice that if each measurement  $x_i$  (resp.  $y_{c,i}$ ) has a particular standard-deviation, formulas (33a) and (33b) 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  $\sigma_{y_c,2,i}$ ).

Using the same resolution scheme as previously, system (32) may be solved and the solution is expressed as:

$$\hat{x} = (I - W_{\hat{x}} A^T (A W_{\hat{x}} A^T)^{-1} A) (\tilde{x} - W_{\hat{x}} \sum_{c=1}^q A_{\hat{y}_c}^T (A_{\hat{x}} W_{\hat{y}_c} A_{\hat{x}}^T)^{-1} A_{\hat{x}} \tilde{y}_c) \quad (34a)$$

$$\hat{y}_c = (I - W_{\hat{y}_c} A_{\hat{x}}^T (A_{\hat{x}} W_{\hat{y}_c} A_{\hat{x}}^T)^{-1} A_{\hat{x}}) \tilde{y}_c \quad (34b)$$

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 (34) 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 (26) which themselves depend on the  $\hat{x}$  and  $\hat{y}_c$  estimations (34). In fact (34) 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 the contamination ratios  $\eta_1$  and  $\eta_2$

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  $c = 1..q$ )

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

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

$$W_{\hat{x}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{\eta_1 p_{\hat{x},1,i}^{(k)}}{\sigma_{x,1}^2} + \frac{(1-\eta_1) p_{\hat{x},2,i}^{(k)}}{\sigma_{x,2}^2}}{\eta_1 p_{\hat{x},1,i}^{(k)} + (1-\eta_1) p_{\hat{x},2,i}^{(k)}} \right)$$

$$W_{\hat{y}_c}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{\eta_2 p_{\hat{y}_c,1,i}^{(k)}}{\sigma_{y_c,1}^2} + \frac{(1-\eta_2) p_{\hat{y}_c,2,i}^{(k)}}{\sigma_{y_c,2}^2}}{\eta_2 p_{\hat{y}_c,1,i}^{(k)} + (1-\eta_2) p_{\hat{y}_c,2,i}^{(k)}} \right)$$

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

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 can also provide an acceptable initialization although its sensitivity to gross errors may be sometimes important; the reader will be able to verify that this solution may be obtained by redefining the distributions (27) with  $\eta_1 = \eta_2 = 1$ .

## 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:

$$\tilde{x} = H_x x + \varepsilon_x \quad (35a)$$

$$\tilde{y} = H_y y + \varepsilon_y \quad (35b)$$

Here, only one species or concentration  $y$  is 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 (26a), defined for each measurement  $\tilde{x}_i$ , can be written using a vector form:

$$p_{x,j} = \frac{1}{(2\pi)^{v/2} \sqrt{\det(V_{x,j})}} \exp\left(-\frac{1}{2}(H_x x - \tilde{x})^T V_{x,j}^{-1} (H_x x - \tilde{x})\right) \quad (36)$$

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:

$$\Phi = \ln(\eta_1 p_{x,1} + (1 - \eta_1) p_{x,2}) (\eta_2 p_{y,1} + (1 - \eta_2) p_{y,2}) \quad (37)$$

Following the same steps as previously, the Lagrange function associated to the maximization of (37) 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 variables  $x$ ,  $y$ ,  $\lambda$  and  $\mu$ :

$$H_x^T W_{\hat{x}}^{-1} (H_x \hat{x} - \tilde{x}) + A^T \lambda + (A \otimes \hat{y})^T \mu = 0 \quad (38a)$$

$$H_y^T W_{\hat{y}}^{-1} (H_y \hat{y} - \tilde{y}) + (A \otimes \hat{x})^T \mu = 0 \quad (38b)$$

$$A \hat{x} = 0 \quad (38c)$$

$$A(\hat{x} \otimes \hat{y}) = 0 \quad (38d)$$

In these last expressions, the weight matrices  $W_{\hat{x}}$  and  $W_{\hat{y}}$  were already defined in (33). Of course system (38) 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 (38) can be transformed into the following implicit system:



$$\hat{x} = \left( G_{\hat{x}} - G_{\hat{x}} A^T (A G_{\hat{x}} A^T)^{-1} A G_{\hat{x}} \right) \left( H_x^T W_{\hat{x}}^{-1} \tilde{x} - A_{\hat{y}}^T (A_{\hat{x}} G_{\hat{y}} A_{\hat{x}}^T)^{-1} A_{\hat{x}} G_{\hat{y}} H_y^T W_{\hat{y}}^{-1} \tilde{y} \right) \quad (39a)$$

$$\hat{y} = (G_{\hat{y}} - G_{\hat{y}} A_{\hat{x}}^T (A_{\hat{x}} G_{\hat{y}} A_{\hat{x}}^T)^{-1} A_{\hat{x}} G_{\hat{y}}) H_y^T W_{\hat{y}}^{-1} \tilde{y} \quad (39b)$$

$$G_{\hat{x}} = (H_x^T W_{\hat{x}}^{-1} H_x + A^T A)^{-1} \quad (39c)$$

$$G_{\hat{y}} = (H_y^T W_{\hat{y}}^{-1} H_y + A_{\hat{x}}^T A_{\hat{x}})^{-1} \quad (39d)$$

Comparing the structures of (39) and (34) 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. An arc 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 3. The considered process involves 16 streams; each of them is characterized by a flowrate and two concentration variables. Except the flowrates of streams number 1, 4 and 11, all the other variables are measured (45 measurements) and are corrupted by random noises. Moreover, 7 measurements are also biased by gross errors. Table 1 indicates the biased measurements and the magnitude of the biases.

Table 1  
Biased measurements

	$x_3$	$x_7$	$x_{16}$	$y_{1,1}$	$y_{1,9}$	$y_{2,8}$	$y_{2,14}$
Magnitude	8.00	8.00	5.00	1.50	1.50	2.00	2.00

Table 2 gathers the measurements and their standard deviations as well as the obtained estimates. For this example, the ratio between “normal” and “abnormal” standard deviations, for all the measurements, has been fixed to  $r = 10$  and the contamination ratios  $\eta_1 = \eta_2 = 0.95$ .

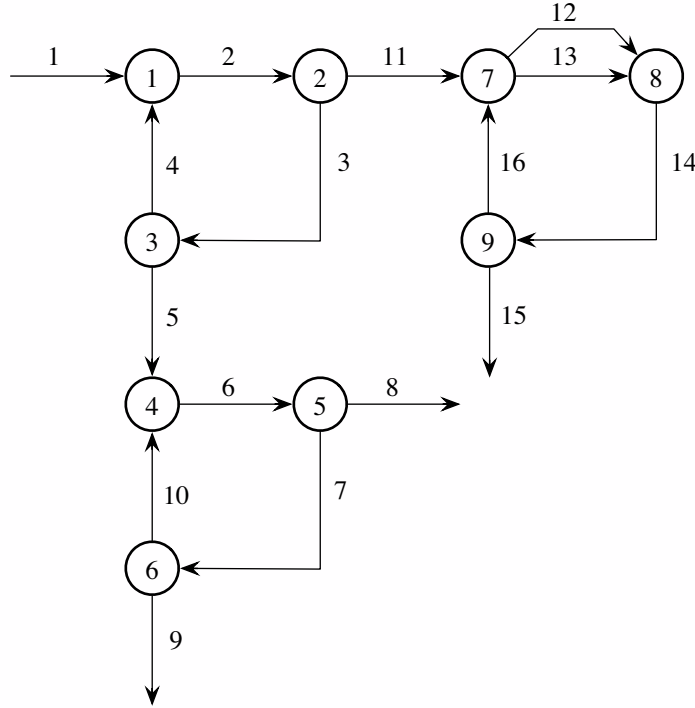


Fig. 3. Oriented graph describing the process flowsheet

It is important to note that the presence of the seven biased measurements had only a weak influence on the obtained estimated values. Indeed, Table 3 shows the corrective terms related to biased measurements. Clearly, the proposed robust data reconciliation method has been able to “compensate” the biases by applying adequate corrective terms to the raw measurements. The magnitude of the corrective terms (table 3) can be compared to the magnitude of the biases (table 1).

In order to analyze the efficiency of the proposed data reconciliation method, corrective terms have been expressed in terms of a percentage of the measurement. Therefore, for each measured variable  $z$ ,  $z \in \{x, y_1, y_2\}$ ,  $c(z) = 100 \frac{\hat{z} - \tilde{z}}{\tilde{z}}$ . Figure 4 shows the relative corrections  $c(z)$  made by the proposed method while figure 5 depicts the same relative corrections when using the classical method of least squares.

The analysis of these figures clearly shows the robustness to the presence of gross errors in the data. The contrast between the relative corrections of the biased measurements and the others is considerably enhanced by the proposed method. Only the biased measurements are strongly corrected and the smearing effect classically observed when using least squares method is considerably minimized.

The main goal of the paper is to provide a robust data reconciliation method

Table 2  
Measurements, standard deviations and estimates

	Measurement			Standard deviation			Robust 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	22.580	2.430	3.460
2	26.500	2.700	3.530	1.325	0.270	0.353	25.540	2.700	3.640
3	29.200	2.520	3.550	1.460	0.252	0.355	20.670	2.500	3.510
4		4.750	5.000		0.475	0.500	2.960	4.760	4.980
5	18.320	2.090	3.290	0.916	0.209	0.329	17.710	2.120	3.270
6	22.020	2.460	3.510	1.101	0.246	0.351	21.870	2.500	3.510
7	20.800	2.900	3.740	1.040	0.290	0.374	12.290	2.900	3.720
8	9.430	2.010	5.200	0.472	0.201	0.520	9.570	2.000	3.250
9	8.010	3.710	3.290	0.401	0.371	0.329	8.140	2.270	3.290
10	4.140	4.150	4.550	0.207	0.415	0.455	4.150	4.140	4.560
11		3.490	4.190		0.349	0.419	4.870	3.550	4.160
12	6.560	3.630	4.340	0.328	0.363	0.434	6.650	3.500	4.280
13	1.040	8.710	6.320	0.052	0.871	0.632	1.040	8.590	6.300
14	7.380	4.240	6.650	0.369	0.424	0.665	7.690	4.190	4.550
15	4.990	3.490	4.100	0.250	0.349	0.410	4.870	3.550	4.160
16	7.690	5.150	5.180	0.385	0.515	0.518	2.820	5.290	5.230

Table 3  
Corrective terms

	$x_3$	$x_7$	$x_{16}$	$y_{1,1}$	$y_{1,9}$	$y_{2,8}$	$y_{2,14}$
Corrective term	8.52	8.51	4.87	1.46	1.45	1.95	2.10

which results are the more insensitive as possible to the presence of gross errors in the measurement set. However, it is clear that the obtained results, in particular the corrective terms, can be analyzed in order to detect and locate these gross errors. Indeed, the magnitude of the corrective terms, i.e.  $\hat{x} - \tilde{x}$ ,  $\hat{y}_1 - \tilde{y}_1$  and  $\hat{y}_2 - \tilde{y}_2$ , can give some information about the quality of the raw measurements. In the linear case, it is easy to deduce, from the statistical properties of the measurements, that of the corrective terms ( $\hat{x} - \tilde{x}$ ). The detection test of gross errors can then be done using a classical hypothesis test. For the nonlinear case (even for the considered bilinear case), the statistical properties of the corrective terms are more difficult to establish. In all the cases, the variance of the estimates cannot be expressed in a closed form

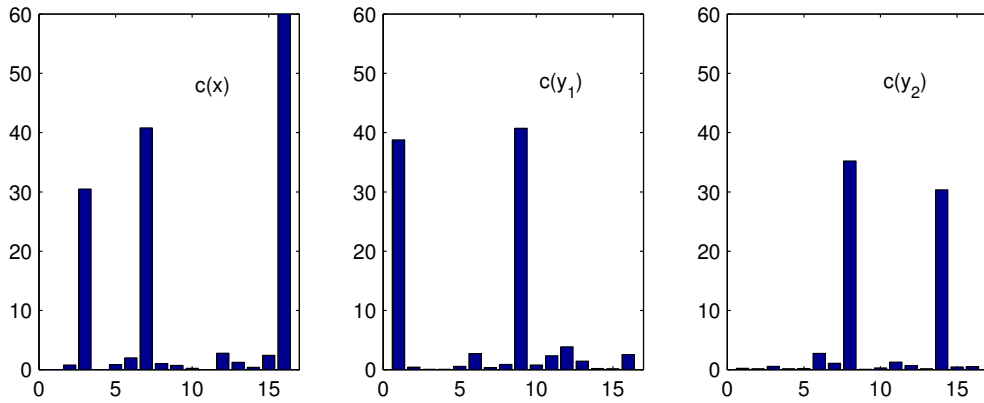


Fig. 4. Relative corrections – RLS case

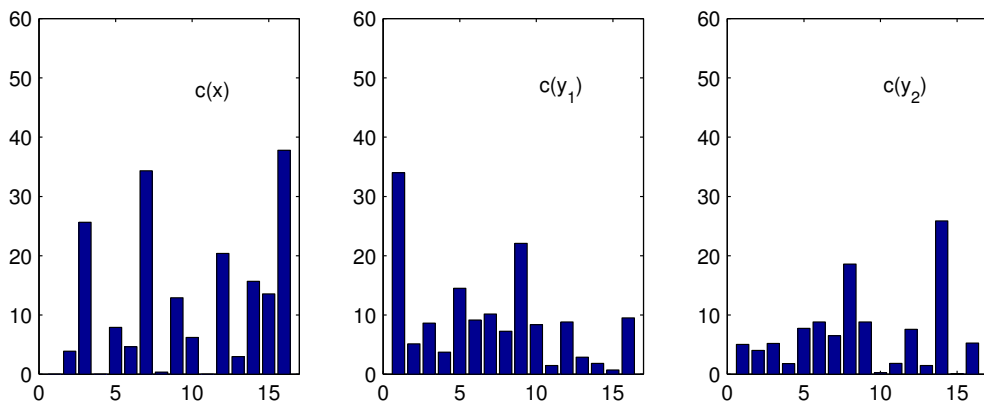


Fig. 5. Relative corrections – LS case

and always relies on some approximations. These approximations are often to blame, in particular in the presence of gross errors.

It is important to note that almost all the existing methods of gross error detection and localization are based on the crossing of a threshold related to the corrective terms (or any standardized corrective terms). When the statistical distribution of these terms is known (and/or calculable), this threshold can be determined on statistical considerations as, for example, the type I error of hypothesis testing. In the other cases, a similar methodology can be applied; however, the threshold must be fixed on an empirical way. For example, one can decide that a measurement contains a gross errors if its corrective term is greater than three or four times its standard deviation. This kind of analysis can also be done, of course, using the relative corrections.

Another technique, described by Tjoa, Biegler (1991) consists to test each corrective term against the contaminated distribution. If the probability associated with a corrective term related to a measurement that is suspected to contain a gross error is greater than that of random error, then the measurement is identified as containing a gross error.

As previously mentioned, the ratio between “normal” and “abnormal” standard deviations of the contaminated distribution do not really constitute a design parameter of the proposed method. Provided the standard deviation ratio  $r$  is greater than 10, the weight function tends to zero for sufficiently large adjustments. In order to evaluate the sensitivity of the results with regard to the choice of the contamination ratios  $\eta$ , 100 different sets of data have been simulated changing, for each set, the additive random noise on the measurements while keeping the seven biases mentioned in table 1. Table 4 shows the mean of the relative corrective terms for the 100 data sets and for three distinct values of the contamination ratio  $\eta \in \{0.8, 0.9, 0.95\}$  (the contamination ratios  $\eta_1$  and  $\eta_2$  have been chosen identical and equal to  $\eta$ ).

Table 4  
Mean of the relative corrective terms

	$\eta = 0.80$			$\eta = 0.90$			$\eta = 0.95$		
	$c(x)$	$c(y_1)$	$c(y_2)$	$c(x)$	$c(y_1)$	$c(y_2)$	$c(x)$	$c(y_1)$	$c(y_2)$
1	–	38	6	–	38	3	–	37	1
2	1	0	4	2	0	0	2	0	1
3	28	0	5	28	1	1	28	1	1
4	–	0	2	–	0	0	–	0	0
5	0	1	8	0	0	5	0	0	2
6	0	2	6	0	1	3	0	1	0
7	38	2	6	40	1	3	40	1	1
8	1	1	24	0	0	31	0	0	38
9	1	39	8	1	40	3	0	39	0
10	0	2	1	0	1	0	0	1	0
11	–	1	1	–	1	0	–	1	0
12	2	3	1	3	3	0	8	2	0
13	0	1	0	0	1	0	0	1	0
14	2	1	31	2	0	31	7	0	31
15	4	1	1	3	1	0	4	1	0
16	60	2	1	61	2	0	55	1	0

If the proposed method functioned perfectly (not any propagation of biases to the healthy measurements), these averages of the relative corrections should be null except for the biased measurements. The results presented in that table are closed to this ideal case and one can notice that the results are not very sensitive to the choice of the contamination ratio. This important result

was foreseeable taking into account the shape of the weight functions shown on figure 2.

## 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 gross errors at a time. For the given example, 7 measurements among 45 have been biased and the proposed method has been able to provide a good estimation of the true values of the different variables of the process. The implementation of the algorithm is easy; moreover many simulations carried out showed a relative low sensitivity of the results to the design parameters, i.e. the standard deviation ratio  $r$  and the contamination ratio  $\eta$  (provided, of course, these parameters are not completely randomly chosen).

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.
- Border, K.C. (1985). *Fixed point theorems with applications to economics and game theory*. Cambridge University Press.
- 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.
- Hoffman, J.D. (2001). *Numerical methods for engineers and scientists*. Marcel Dekker.
- Huber, P.J. (1981). *Robust statistic*. John Wiley & Sons, New York.
- Johnston, L.P.M. and Kramer, M.A. (1995). Maximum likelihood data rectification: steady state systems. *AIChE Journal*, 41(11), 2415-2426.
- 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.
- Kuehn, D.R. and Davidson H. (1961). Computer control. II: Mathematics of control. *Chemical Engineering Progress*, 57, 44.
- 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 Jordache, C. (2000). *Data reconciliation and gross error detection – an intelligent use of process data*, Gulf Publishing Company, Houston, Texas, USA.
- 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.
- Serth, R. and Heenan, W. (1986). Gross error detection and data reconciliation in steam metering systems. *AIChE Journal*, 32, 733-742.
- 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.
- Veverka, V.V. and Madron F. (1997). Material and energy balancing in the process industries: from microscopic balances to large plants. *Computer-aided chemical engineering series*, vol. 7, Amsterdam, The Netherlands, Elsevier.
- 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.