

# DATA RECONCILIATION: A ROBUST APPROACH USING CONTAMINATED DISTRIBUTION. APPLICATION TO A PETROCHEMICAL PROCESS

José Ragot \* Didier Maquin \*

\* *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*

{Jose.Ragot, Didier.Maquin}@ensem.inpl-nancy.fr

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. Application to total flowrate and concentration data in a petroleum network transportation is provided  
*Copyright*© 2005 *IFAC*.

Keywords: Data reconciliation, Robust estimation, Gross error detection, Linear and bilinear mass balances.

## 1. INTRODUCTION

The problem of obtaining reliable estimates of the state of a process is a fundamental objective, 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, *et al.*, 1976), (Maquin, *et al.*, 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 and Cauvin, 1999), (Heyen, 1999), (Singh, *et al.*, 2001), (Yi, *et al.*, 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 will be corrupted by this bias. Several schemes have been suggested

to cope with the corruption of normal assumption of the errors, for static systems (Narasimhan and Mah, 1989), (Kim, *et al.*, 1997), (Arora, 2001) and also for dynamic systems (Abu-el-zeet, *et al.*, 2001). Methods to include bounds in process variables to improve gross error detection have been developed. One major disadvantage of these methods is that they give rise to situations that it may be impossible to estimate all the variables by using only a subset of the remaining free gross errors measurements. Alternative approach using constraints both on the estimates and the balance residual equations has been developed for linear system (Ragot, *et al.*, 1999), (Maquin and Ragot, 2004). There is also an important class of robust estimators whose influence functions are bounded allowing to reject outliers (Huber, 1981), (Hampel, *et al.*, 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. In the following, we adopt and develop this idea for the data reconciliation problem.

Section 2 will be 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. Finally, in section 5, the proposed method is implemented on a fictitious but realistic petroleum network transportation.

## 2. DATA RECONCILIATION BACKGROUND

The classical general data reconciliation problem (Mah, *et al.*, 1976), (Hodouin and Flament, 1989), (Crowe, 1996), deals with a weighted least squares minimisation of the measurement adjustments subject to the model constraints. Here, the process model equations are taken as linear for sake of simplicity:

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

where  $x$ , with components  $x_i$  is the state of the process. The measurement devices give the information:

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

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

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

where  $\sigma_i^2$  are the diagonal elements of  $V$ . From (3) one derives the likelihood function of the ob-

servations with the hypothesis of independent realisations. Maximisation of the likelihood function leads to the estimate  $\hat{x} = (I - VA^T(AVA^T)^{-1}A)x$  (Maquin, *et al.*, 1991). In fact, the method doesn't work in any situation, the main drawback being the contamination of all estimated values by the outliers. For that reason robust estimators could be preferred, robustness being the ability to ignore the contribution of extreme data such as gross errors. There are two approaches to deal with outliers. The first one consists to sequentially detect, localise and suppress the data which are contaminated and after to reconcile the remaining data. The second approach is global and reconciles 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, one having a small variance representing regular noise and the other having a large variance representing outliers (Wang and Romagnoli, 2002), (Ghosh and 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 ( $j = 1, 2$ ):

$$p_{j,i}(\tilde{x}_i | x_i, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2} \left(\frac{x_i - \tilde{x}_i}{\sigma_j}\right)^2\right) \quad (4)$$

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

$$p(\tilde{x}_i | x_i, \theta) = wp_{1,i} + (1-w)p_{2,i} \quad 0 \leq w \leq 1 \quad (5)$$

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:

$$\Phi = \ln \prod_{i=1}^v p(\tilde{x}_i | x_i, \theta) \quad (6)$$

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:

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

$$\text{subject to } Ax = 0 \quad (7b)$$

Using the classical Lagrange method leads to the following estimate  $\hat{x}$ :

$$\hat{x} = (I - W_{\hat{x}} A^T (A W_{\hat{x}} A^T)^{-1} A) \tilde{x} \quad (8a)$$

$$W_{\hat{x}}^{-1} = \text{diag} \left( \frac{w}{\sigma_1^2} \hat{p}_{1,i} + \frac{1-w}{\sigma_2^2} \hat{p}_{2,i} \right)_{i=1..v} \quad (8b)$$

$$\hat{p}_{j,i} = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left( -\frac{1}{2} \left( \frac{\hat{x}_i - \tilde{x}_i}{\sigma_j} \right)^2 \right) \quad (8c)$$

where the diag operator allows one to define a diagonal matrix from the elements (pointed by  $i$ ) of a vector. Thus system (8) is clearly non linear and we suggest to solve it using the following direct iterative scheme:

$$k = 0, \quad x^{(k)} = \tilde{x} \quad (9a)$$

$$\hat{p}_{j,i}^{(k)} = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left( -\frac{1}{2} \left( \frac{\hat{x}_i^{(k)} - \tilde{x}_i}{\sigma_j} \right)^2 \right) \quad (9b)$$

$$(W_{\hat{x}}^{(k)})^{-1} = \text{diag} \left( \frac{w}{\sigma_1^2} \hat{p}_{1,i}^{(k)} + \frac{1-w}{\sigma_2^2} \hat{p}_{2,i}^{(k)} \right)_{i=1..v} \quad (9c)$$

$$\hat{x}^{(k+1)} = (I - W_{\hat{x}}^{(k)} A^T (A W_{\hat{x}}^{(k)} A^T)^{-1} A) \tilde{x} \quad (9d)$$

A stopping criterion must be chosen for implementing the algorithm. For sake of simplicity, the proof for the local convergence of the algorithm is omitted and the reader is invited to refer to the specialized literature for obtaining more details about fixed point theory (Border, 1985).

### 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, *et al.*, 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 origin to 0.1 for data with large magnitude.

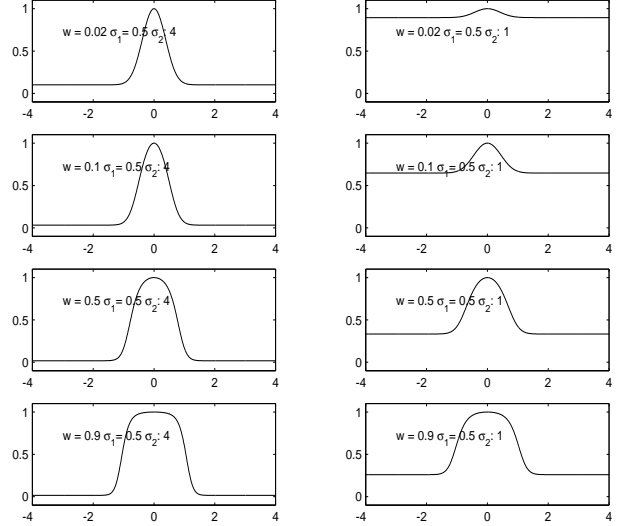


Fig. 1. Influence function

## 4. EXTENSION TO BILINEAR SYSTEMS

We consider now the case of a process characterised by two types of variables such as flowrates  $x$  and concentrations  $y$ . 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.  $y_i$ ) is assumed to have the same normal  $\sigma_{x,1}$  (resp.  $\sigma_{y,1}$ ) and abnormal  $\sigma_{x,2}$  (resp.  $\sigma_{y,2}$ ) standard-deviation. This hypothesis will be withdrawn later on. Thus, for each observation  $\tilde{x}_i$  and  $\tilde{y}_i$ , we define the following pdf:

$$p(\tilde{x}_i | x_i, \sigma_{x,j}) = \frac{1}{\sqrt{2\pi}\sigma_{x,j}} \exp \left( -\frac{1}{2} \left( \frac{x_i - \tilde{x}_i}{\sigma_{x,j}} \right)^2 \right) \quad (10a)$$

$$p(\tilde{y}_i | y_i, \sigma_{y,j}) = \frac{1}{\sqrt{2\pi}\sigma_{y,j}} \exp \left( -\frac{1}{2} \left( \frac{y_i - \tilde{y}_i}{\sigma_{y,j}} \right)^2 \right) \quad (10b)$$

with  $j = 1, 2, i = 1..v$ . In the rest of the paper,  $p_{x,j,i}$  and  $p_{y,j,i}$  respectively are shortening notations for  $p(\tilde{x}_i | x_i, \sigma_{x,j})$ , and  $p(\tilde{y}_i | y_i, \sigma_{y,j})$  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:

$$p_{x,i} = w p_{x,1,i} + (1-w) p_{x,2,i} \quad i = 1..v \quad (11a)$$

$$p_{y,i} = w p_{y,1,i} + (1-w) p_{y,2,i} \quad i = 1..v \quad (11b)$$

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

$$\Phi = \ln \prod_{i=1}^v p_{x,i} p_{y,i} \quad (12)$$

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

$$Ax = 0 \quad (13)$$

$$A(x \otimes y_c) = 0 \quad (14)$$

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

$$L = \Phi + \lambda^T Ax + \mu^T A(x \otimes y) \quad (15)$$

Constraints are taken into account through the introduction of the Lagrange parameters  $\lambda$  and  $\mu$ . A sequence of very elementary matrix algebra leads to the stationarity conditions of (15) (the estimates are now noted  $\hat{x}$  and  $\hat{y}_c$ ):

$$W_{\hat{x}}^{-1}(\hat{x} - \tilde{x}) + A^T \lambda + (A \otimes \hat{y})^T \mu = 0 \quad (16a)$$

$$W_{\hat{y}}^{-1}(\hat{y} - \tilde{y}) + (A \otimes \hat{x})^T \mu = 0 \quad (16b)$$

$$A\hat{x} = 0 \quad (16c)$$

$$A(\hat{x} \otimes \hat{y}) = 0 \quad (16d)$$

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

$$W_{\hat{x}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{wp_{\hat{x},1,i}}{\sigma_{x,1}^2} + \frac{(1-w)p_{\hat{x},2,i}}{\sigma_{x,2}^2}}{wp_{\hat{x},1,i} + (1-w)p_{\hat{x},2,i}} \right) \quad (17a)$$

$$W_{\hat{y}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{wp_{\hat{y},1,i}}{\sigma_{y,1}^2} + \frac{(1-w)p_{\hat{y},2,i}}{\sigma_{y,2}^2}}{wp_{\hat{y},1,i} + (1-w)p_{\hat{y},2,i}} \right) \quad (17b)$$

Notice that if each measurement  $x_i$  (resp.  $y_i$ ) has a particular standard-deviation, formulas (17a) and (17b) still hold by replacing the parameters  $\sigma_{x,1}$  and  $\sigma_{x,2}$  (resp.  $\sigma_{y,1}$  and  $\sigma_{y,2}$ ) by  $\sigma_{x,1,i}$  and  $\sigma_{x,2,i}$  (resp.  $\sigma_{y,1,i}$  and  $\sigma_{y,2,i}$ ). System (16) may be directly solved and the solution is expressed as:

$$\hat{x} = (I - W_{\hat{x}} A^T (A W_{\hat{x}} A^T)^{-1} A) \dots (\tilde{x} - W_{\hat{x}} A_{\hat{y}}^T (A_{\hat{x}} W_{\hat{y}} A_{\hat{x}}^T)^{-1} A_{\hat{x}} \tilde{y}) \quad (18a)$$

$$\hat{y} = (I - W_{\hat{y}} A_{\hat{x}}^T (A_{\hat{x}} W_{\hat{y}} A_{\hat{x}}^T)^{-1} A_{\hat{x}}) \tilde{y} \quad (18b)$$

where the shortening notations  $A_x$  and  $A_y$  respectively stand for  $A \text{diag}(x)$  and  $A \text{diag}(y)$ . System (18) is clearly non linear with regard to the unknown  $\hat{x}$  and  $\hat{y}$ , the weights  $W_{\hat{x}}$  and  $W_{\hat{y}}$  depending on the pdf (10) which themselves depend on the  $\hat{x}$  and  $\hat{y}$  estimations (18). In fact (18) is an implicit system in respect to the estimates  $\hat{x}$  and  $\hat{y}$  for which we suggest the following iterative scheme:

### Step 1: initialisation

$k = 0$

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

Choose  $w$

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

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

### Step 2: estimation

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

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

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

$$W_{\hat{x}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{wp_{\hat{x},1,i}^{(k)}}{\sigma_{x_1}^2} + \frac{(1-w)p_{\hat{x},2,i}^{(k)}}{\sigma_{x_2}^2}}{wp_{\hat{x},1,i}^{(k)} + (1-w)p_{\hat{x},2,i}^{(k)}} \right)$$

$$W_{\hat{y}}^{-1} = \text{diag}_{i=1..v} \left( \frac{\frac{wp_{\hat{y},1,i}^{(k)}}{\sigma_{y_1}^2} + \frac{(1-w)p_{\hat{y},2,i}^{(k)}}{\sigma_{y_2}^2}}{wp_{\hat{y},1,i}^{(k)} + (1-w)p_{\hat{y},2,i}^{(k)}} \right)$$

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

Update the estimation of  $x$  and  $y$

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

$$\dots \left( \tilde{x} - W_{\hat{x}}^{(k)} A_{\hat{y}}^{(k)T} (A_{\hat{x}}^{(k)} W_{\hat{y}}^{(k)} A_{\hat{x}}^{(k)T})^{-1} A_{\hat{x}}^{(k)} \tilde{y}_c \right)$$

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

### Step 3: convergence test

Compute an appropriate norm of the corrective terms:  $\tau_x^{(k+1)} = \|\hat{x}^{(k+1)} - \tilde{x}\|$  and  $\tau_y^{(k+1)} = \|\hat{y}^{(k+1)} - \tilde{y}\|$ . If the variations  $\tau_x^{(k+1)} - \tau_x^{(k)}$  and  $\tau_y^{(k+1)} - \tau_y^{(k)}$  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 square approach would also provide 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 (11) with  $w = 1$ .

## 5. EXAMPLE AND DISCUSSION

The method described in section 4 is applied to the system (a part of a petroleum network

transportation) depicted by figure 2, for which 11 streams are considered; each stream is characterized by total flowrate (oil plus water) and water percentage (ratio water/oil). Random errors were added to the 11 variables but the gross errors were added only on some of them.

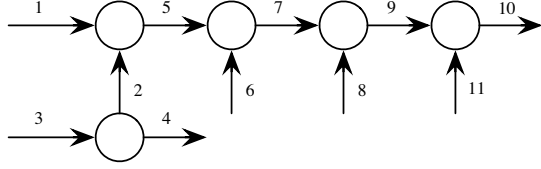


Fig. 2. Flowsheet

The performance results are given when two gross errors of magnitudes 8 affect the measurements 3 and 7, and simultaneously, two gross errors of magnitude 10 affect the measurements of the concentration for streams 1 and 9. Comparison of the proposed robust least square algorithm (RLS) with the classical least square (LS) algorithm is now provided in table 2.

Table 1. Measurements and estimations

Measurement		RLS estimate		LS estimate		
$x$	$y$	$\hat{x}$	$\hat{y}$	$\hat{x}$	$\hat{y}$	
1	16.9	<b>19.9</b>	17.06	<b>14.41</b>	16.61	<b>12.89</b>
2	13.6	9.5	13.54	8.20	14.97	8.29
3	<b>19.6</b>	25.8	<b>16.49</b>	24.98	<b>15.26</b>	31.64
4	3.0	100	2.98	100.52	0.33	46.86
5	30.7	9.7	30.63	11.39	31.76	11.24
6	3.8	7.6	3.84	7.41	3.55	7.51
7	<b>38.2</b>	9.5	<b>34.49</b>	10.81	<b>35.27</b>	10.80
8	5.9	32.9	5.87	32.49	6.12	33.27
9	40.4	<b>22.9</b>	40.36	<b>13.92</b>	41.28	<b>14.72</b>
10	36.4	2.2	36.32	2.70	33.57	2.26
11	3.9	112.0	4.04	113.77	7.71	128.82

Columns 2 and 3 relate the row measures, columns 4 and 5 show the estimations obtained with RLS and columns 6 and 7 the estimations obtained with LS. Analysing the RLS estimation errors clearly allows to suspect variables 3 and 7 for being contaminated by a gross error. Such conclusion is more difficult to express with LS estimator. Indeed, stream 4 for  $x$  variable has a measurement of 3.0 and respective RLS and LS estimates are 2.98 and 0.33; it means that LS corrects data which are not a priori corrupted by errors. The same effect may be seen on several other streams such as stream 11 for variable  $y$  for which LS also proposes a important correction of a free gross error variable.

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 axes 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 of 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 measurements corrupted by the gross errors and not the others). Figure 4 shows mean values of the corrective terms obtained from 100 runs.

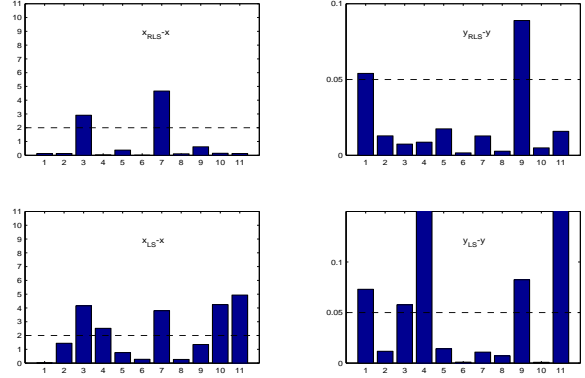


Fig. 3. Corrective terms

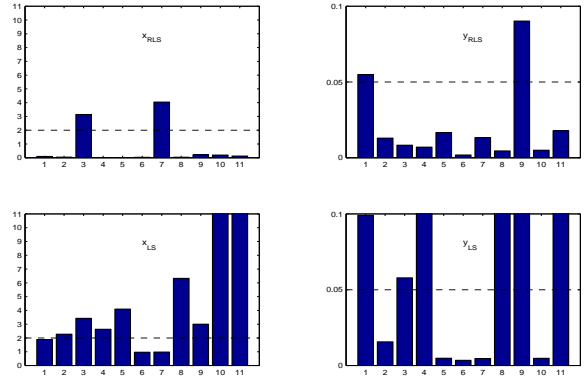


Fig. 4. Mean corrective terms (100 runs)

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 the enumeration of the cases where the gross errors have been correctly detected and isolated, both for RLS and LS methods. Results, expressed in percentage of correct fault detection, are shown in table 2. 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.

Table 2. Correct fault detection in %

Var.	RLS gross error detection		LS gross error detection	
	$x$	$y$	$x$	$y$
$w=0.10$	92.5	95.3	41.4	56.6

Of course, the choice of the tuning parameters  $w$ ,  $\sigma_{x,i}$  and  $\sigma_{y,i}$  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  $\sigma_{y,1}/\sigma_{y,2}$ . Therefore, it is relatively easy to adjust manually the parameters of the method and a “large” range of acceptable values may be found. However, it is also possible to use an adaptive algorithm for this adjusting.

## 6. CONCLUSION

To deal with the issues of gross errors influence on data estimation, the paper has presented a robust reconciliation approach. For that purpose, a cost function which is less sensitive to the outlying observations than that of least squares was introduced. The algorithm can handle multiple biases or outliers at a time and for the given example, 4 outliers have been correctly detected on 22 variables.

The results of reconciliation 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 into account the model uncertainties and optimise the balancing parameter  $w$ . Moreover, for process with unknown parameter, it should be important to jointly estimate the reconciled data and the process parameters.

## REFERENCES

- Abu-el-zeet, Z.H., P.D. Roberts and V.M. Becerra (2001). Bias detection and identification in dynamic data reconciliation. European Control Conference, Porto, Portugal, September 4-7.
- Arora, N. and L.T. Biegler (2001). Redescending estimator for data reconciliation and parameter estimation. *Computers and Chemical Engineering*, 25 (11/12), 1585-1599.
- 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., S. Cauvin (1999). On-line fault detection and supervision in the chemical process industries. *Control Engineering Practice*, 7 (7), 863-864.
- Ghosh-Dastider, B. and J.L. Schafer (2003). Outlier detection and editing procedures for continuous multivariate data. Working paper 2003-07, RAND, Santa Monica OPR, Princeton University.
- Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stohel (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. 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 F. Flament (1989). New developments in material balance calculations for mineral processing industry. Society of Mining Engineers annual meeting, Las Vegas, February 27 - March 2.
- Hubert, P.J. (1981). *Robust statistic*. John Wiley & Sons, New York.
- Kim, I., M.S. Kang, S. Park. and T.F. Edgar (1997). Robust data reconciliation and gross error detection: the modified MIMT using NLP. *Computers and Chemical Eng.*, 21, 775-782.
- Mah, R.S.H., G.M. Stanley and D. Downing (1976). Reconciliation and rectification of process flow and inventory data. *Ind. Eng. Chem., Process Des. Dev.*, 15 (1), 175-183.
- Maquin, D., G. Bloch and J. Ragot (1991). Data reconciliation for measurements. *European Journal of Diagnosis and Safety in Automation*, 1 (2), 145-181.
- Maquin, D. and J. Ragot (2003). Validation de données issues de systèmes de mesure incertains. *Journal Européen des Systèmes Automatisés*, 37 (9), 1163-1179.
- Narasimhan, S. and R.S.H. Mah (1989). Treatment of general steady state process models in gross error identification. *Computers and Chemical Engineering*, 13 (7), 851-853.
- Ragot, J., D. Maquin and O. Adrot. LMI approach for data reconciliation (1999). 38th Conference of Metallurgists, Symposium Optimization and Control in Minerals, Metals and Materials Processing, Quebec, Canada, August 22-26.
- Singh, S.R., N.K. Mittal and P.K. Sen (2001). A novel data reconciliation and gross error detection tool for the mineral processing industry. *Minerals Engineering*, 14 (7), 808-814.
- Wang, D. and J.A. Romagnoli (2002). Robust data reconciliation based on a generalized objective function. 15th IFAC World Congress on Automatic Control, Barcelona, Spain, July 21-26.
- Yi, H.-S., J. H. Kim and C. Han (2002). Industrial application of gross error estimation and data reconciliation to byproduction gases in iron and steel making plants. International Conference on Control, Automation and Systems, Muju, Korea, October 16-19.