



**HAL**  
open science

## Data reconciliation for process flow

José Ragot, Didier Maquin, Gérard Bloch

► **To cite this version:**

José Ragot, Didier Maquin, Gérard Bloch. Data reconciliation for process flow. European Control Conference, ECC'91, Jul 1991, Grenoble, France. pp.1430-1435. hal-00302871

**HAL Id: hal-00302871**

**<https://hal.science/hal-00302871>**

Submitted on 29 Mar 2014

**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.

J. RAGOT, D. MAQUIN, G. BLOCH

Centre de Recherche en Automatique de Nancy  
Rue du doyen Marcel Roubault  
BP 40 - 54501 Vandoeuvre Cedex - France

## SUMMARY

Before improving the control of a plant, we must make sure of information coherency issued from instrument lines or sensors. In fact, this information can be corrupted by errors and can also deviate from the optimum functioning range. Various methods for the detection and location of gross errors in process data have been proposed in recent years including the parity space approach, the standardized least square residuals approach and the standardized imbalance residuals approach.

In this survey, the authors try to summarize the various aspects of data reconciliation, to point out the main difficulties and to present the state of the art in this field. The authors present the steps of the data reconciliation problem in the following order : techniques of data reconciliation, classification of the data by the observability concept, gross error detection and localisation, variance of measurement error estimation and sensor positioning. The whole presentation is applied to linear systems nevertheless some extensions are given for non linear systems.

## KEYWORDS

data reconciliation, redundancy, observability, error detection, variance estimation, model based fault detection

## INTRODUCTION

The validation of signals is a technique which integrates information from redundant and from functionally diverse sensors to provide highly reliable information to operating crews and to automatic controllers. Signal validation is generally performed by like-sensor comparisons (direct redundancy). When the increase of sensors is impossible, we prefer to use analytical redundancy ; analytical redundancy refers to the physical relationships, such as conservation of mass or conservation of energy, that exist among the many variables being measured in a system.

### 1 - DATA RECONCILIATION

The data reconciliation problem can be formulated in very simple terms. A set of measurements does not check process functioning equations. How may one correct (or reconcile) the measurements in order to force them to verify this set of equations which are supposed to be exact ?

In practise, the formulation of the problem is not so easy. Indeed one is faced with the choice of hypothesis about the statistical distribution of the measurement errors, the system dimension (number of variables and constraints), the nature of the constraints (static or not, linear or not), the process state (steady or not), and incomplete or imperfect knowledge of the model structure or of the parameters.

Reconciliation techniques have been applied in various fields, including chemical processes (see Clair (1975), Holly (1989)) and mineralurgy (Cutting (1976)).

#### 1.1 - Problem formulation

Generally, the systems in question can be described in two parts. The first is the set of static constraints and the second is the measurement equation :

$$F(X^*, \theta^*) = 0 \quad (1.1)$$

$$Z = H(X^*) + \varepsilon \quad (1.2)$$

If the measurement errors are normally distributed with zero mean and known diagonal variance matrix  $V$ , the probability density function of the measurements  $Z$  can be written as :

$$P(Z) = \frac{1}{(2\pi)^{m/2} |V|^{1/2}} \exp\left(-\frac{1}{2} \|Z - H(X^*)\|_{V^{-1}}^2\right) \quad (1.3)$$

The maximum likelihood estimator  $(\hat{X}, \hat{\theta})$  of the true values  $(X^*, \theta^*)$  maximises the probability density function  $P$ , subject to the model constraints.

The difficulty in solving this problem is closely related to the structure and the dimension of model equations as well as to the number of observations. In the case of linear constraints and measurement equations, analytical solutions can be found. In the other cases, one must use techniques of hierarchical calculus, of linearization, of variable substitution or even approximate solutions.

#### 1.2 - Linear models case

The linear models case can seem to be very restrictive. Nevertheless, industrial applications frequently call for this class of systems, which represents systems described physically by balance equations in total material or energy flows. Moreover, an approximate solution to problems described by multilinear equations can be found, using linear formulation, by decoupling multilinear equations in linear equations. Generally, a process is made up of interconnected subsets and can be represented by an oriented graph, where the  $v$  streams represent the flows and the  $n$  nodes processing units or junction points or several streams. Each node is associated to a balance equation, drawn from the conservation laws of material or energy. Thus, the complete set of the balance equations can be written as :

$$M X^* = b \quad (1.4)$$

where  $M$  is the  $(n.v)$  incidence matrix, associated to the system.

If the variables are measured with linear devices, the measurement equation is simply :

$$Z = H X^* + \varepsilon \quad (1.5)$$

A global resolution of problem (1.6) by the Lagrange multipliers technique leads to the unbiased estimator :

$$\hat{X} = P X = (I - G^{-1} M^T (M G^{-1} M^T)^{-1} M) X \quad (1.7)$$

$$G^{-1} = H^T V^{-1} H + M^T M \quad (1.8)$$

In practise, for low dimension and well-conditioned systems, the computation of  $\hat{X}$  by equation (1.7) does not pose any numerical difficulty. For large scale systems (for instance, more than hundred nodes), others techniques must be employed, including constraints elimination, recursive estimation, decentralized calculus, sparse matrices techniques, orthogonal transformations... For instance, the search for the minimum of a function subject to constraints can be replaced by the search for the minimum of a function not subject to constraints, by reducing the number of variables. Another technique consists of progressively incorporating the constraints. The measurements are projected in the first constraint subspace ; the estimation obtained is then projected in the first two constraints subspace ...

### 1.3 - Bilinear models case

In this case, the system is defined by an incidence matrix M and two kinds of values  $X^*$  and  $Y_i^*$  ( $i = 1, \dots, nc$ ) to be estimated (for instance, in mineralurgical processes, material flow rates and concentrations or contents, respectively, linked by total and partial material balance equations) :

$$M X^* = 0 \quad (1.9a)$$

$$M X^* * Y_i^* = 0 \quad i = 1, \dots, nc \quad (1.9b)$$

where the symbol \* denotes the Hadamard product : the jth component of the vector  $X^* Y$  is the product of the jth components of  $X$  and  $Y$ .

The validation problem consists of seeking the minimum with respect to  $\hat{X}$  and  $\hat{Y}_i$  ( $i = 1, \dots, nc$ ) of the probability density function in  $X$  and  $Y$ .

Several methods can be used to solve this optimisation problem : an approximate suboptimal method which neglects the coupling between balance equations, a method based on the linearization of balance equations, a method using estimation by relaxation. In the case of multistage operations, the constraint equations take the form of a tridiagonal matrix and this particular sparse form is used to reduce the number of equations to be solved simultaneously in an iterative procedure. We note that the bilinear formulation can be extended to multilinear cases. But even in the bilinear case, the linearized systems become rapidly of large scale. So, for multilinear systems, one is well advised to solve the optimisation problem by a hierarchical decomposition of the calculations, for instance by variable type. The resulting direct iterative scheme, which is rather difficult to present with conciseness, can be found in Ragot (1990).

## 2 - OBSERVABILITY

### 2.1 - Introduction

The different techniques which have been previously presented can be applied to systems described by algebraic linear or non linear equations when all the variables are measured. Technological and economical constraints generally prevent the measurement of all the variables. The fragmentary character of information prohibits the immediate usage of the previous methods. It is then necessary to do a preliminary quantitative analysis of all the available information in order to determine, before any calculus, those which can

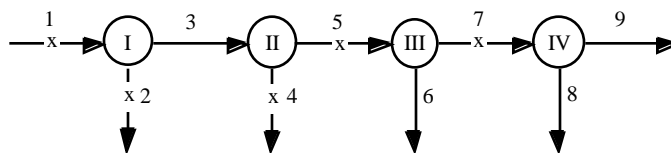


Figure 1 : a simple network

In this commonplace example, we easily find that variables 1, 2, 4 and 5 are bound by a redundancy equation (an equation with only measured variables). These variables are said to be overdetermined and will be corrected by the previous techniques. On the contrary, variable 7 could not be estimated. Variables 3 and 6 can be easily deduced from equation of nodes I and II for the first one and III for the second one. The lack of information prohibits any estimation of variables 8 and 9. In fact, there is only one equation to determine two unknown quantities.

In general the variables can be classified into two essential categories (i) the observable variables : this group contains the measured and the deducible variables, (ii) the unobservable variables : the lack of information prevents their estimation.

### Observability

We consider a linear system under steady-state condition defined by equations (1.4) and (1.5). This system is said to be globally observable if the knowledge of the measurements  $Z$  and of the model constraints allow one to find a single estimation  $\hat{X}$  when  $\varepsilon = 0$ . This system of equations has a single solution only if :

$$\text{rank} \begin{pmatrix} H \\ M \end{pmatrix} = \text{dim}(X^*) = v \quad (2.1)$$

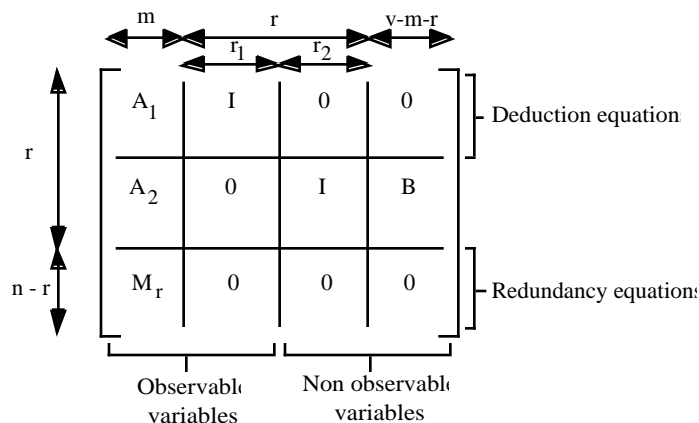
If all the variables are not observable, it is necessary to determine the observable subsystem to which data reconciliation can be applied.

### Redundancy

The linear system described by (1.4) and (1.5) is said to be redundant when the data exceeds the minimum necessary for a single determination of the system state  $X^*$ . This definition can be translated by an inequality on the dimensions of the different matrices. The system is redundant if  $n + m > v$  and its degree of redundancy  $r$  is equal to  $n + m - v$ .

### Practical method of decomposition

Several methods can be used to perform this decomposition (Darouach, 1986) singular values decomposition, reduction of the incidence matrix to an echelon form etc... This last method is the simplest and the most efficient because it uses the particular structure of the incidence matrices which contain only  $\pm 1$  and 0 elements (Maquin, 1987). After some elementary row and column operations, the incidence matrix can be written as :



The examination of this matrix shows : redundancy equations which are described by  $M_r$  deduction equations described by the

to null columns of  $M_T$  and cannot be corrected), unmeasured and deducible variables (these variables are deduced from those of the previous classes), unmeasured and non estimable variables : these variables do not appear in the deduction equations and cannot be corrected ; further measurements are necessary to make these variables observable.

## 2.2 - The bilinear case

In the case of only one component, the bilinear system can be described by equations (1.9) with  $i=1$ . Equation (1.9b) can also be written as :

$$M X * Y = (M \otimes X) Y = (M \otimes Y) X$$

where the  $\otimes$  operator is applied between a matrix and a vector in order to form a matrix in which the line  $k$  is obtained by multiplying each of the terms of the line  $k$  of the matrix by each term of the vector.

The two measurement vectors  $X$  and  $Y$  can be partitioned into measured and unmeasured parts :

$$X = \begin{bmatrix} X_m \\ X_{\bar{m}} \end{bmatrix} \quad Y = \begin{bmatrix} Y_m \\ Y_{\bar{m}} \end{bmatrix} \quad (2.2)$$

These decompositions allow the classification of the variables into four distinct groups. According to the above partition (2.2), the incidence matrix  $M$  is partitioned as well :

$$M = [M_1 \quad M_2 \quad M_3 \quad M_4] \quad (2.3)$$

The system (1.9) can then be written as :

$$M_1 X_{\bar{m}1} + M_2 X_{\bar{m}2} = -(M_3 X_{m1} + M_4 X_{m2}) = d_1 \quad (2.4a)$$

$$M_1 X_{\bar{m}1} * Y_{m1} + M_2 X_{\bar{m}2} * Y_{m2} + M_3 X_{m1} * Y_{\bar{m}1} - M_4 X_{m2} * Y_{\bar{m}2} = d_2 \quad (2.4b)$$

or in matrix notation :

$$O_{(X_m, Y_m)} \begin{bmatrix} X_{\bar{m}2} \\ X_{\bar{m}1} \\ X_{\bar{m}2} * Y_{\bar{m}2} \\ Y_{\bar{m}1} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \quad (2.5)$$

with

$$O_{(X_m, Y_m)} = \begin{bmatrix} M_2 & M_1 & O & O \\ O & M_1 \otimes Y_{m1} & M_2 & M_3 \otimes X_{m1} \end{bmatrix} \quad (2.6)$$

The system will be observable if :

$$\text{rank } O_{(X_m, Y_m)} = \dim \begin{bmatrix} X_{\bar{m}} \\ Y_{\bar{m}} \end{bmatrix} \quad (2.7)$$

If the system is not globally observable, a decomposition is necessary. We will show that the study of observability of bilinear systems can be reduced to a sequential study of observability of linear systems (Maquin, 1987).

For the study of the observability of  $X_{\bar{m}1}$ , let us introduce two matrices  $Q$  and  $R$  which allow the elimination of the unmeasured terms  $X_{\bar{m}2}$  and  $Y_{\bar{m}2}$  in equation (2.4). Let us consider  $Q$ , an

$$(Q M_1 \otimes Y_{m1}) X_{\bar{m}1} + (Q M_3 \otimes X_{m1}) Y_{\bar{m}1} = Q d_2 \quad (2.8b)$$

With  $R$ , a regular matrix defined by :  $R Q M_3 = 0$ , the system (2.8) may be transformed to :

$$Q M_1 X_{\bar{m}1} = Q d_1 \quad (2.9a)$$

$$(R Q M_1 \otimes Y_{m1}) X_{\bar{m}1} = R Q d_2 \quad (2.9b)$$

Four steps are then necessary to study the observability :

- System (2.9) enables one to extract the observable part of  $X_{\bar{m}1}$ .
- Using the already known observable part of  $X_{\bar{m}1}$ , equation (2.8) allows the determination of the observable part of  $Y_{\bar{m}1}$ .
- The observable part of  $X_{\bar{m}2}$  is obtained from the knowledge of observable parts of  $X_{\bar{m}1}$  and  $Y_{\bar{m}1}$ , and equation (2.4a).
- Similarly, equation (2.4b) is used to extract the observable part of  $Y_{\bar{m}2}$ .

Crowe et al. (1986) have developed a quite similar approach to solve this problem. They proposed the construction of two projection matrices in order to decompose the problem into subproblems to be solved in sequence. The first matrix eliminates all unmeasured component flow rates and concentrations from the equations ; the second then removes the unmeasured total flow rates. To solve this problem, Romagnoli et al. (1980) introduced a structural representation of the system. Using the properties of the structural matrices, which only take into account the occurrences of a variable in the equations, they proposed a classification of the variables. Their analysis uses the generic rank of the structural matrices and an algorithm to assign one equation to any unmeasured variables. It is important to note that these previous works require the composition of each stream to be either completely measured or not measured at all.

More recently, Kretsovalis et al. (1988) presented a classification algorithm which does not impose any restriction on the number and the localization of measurements. They included in their treatment chemical reactions, flow splitters and pure energy flows. They make use of graph theoretical concepts and proceed in a layered approach through the various graphs derived from the process graph.

## 3 - GROSS ERROR DETECTION

This part of the paper presents methods of detection and location of measurement failures which are also called gross or large errors. Some of the techniques presented use linear models but some others can be applied to non-linear models. We assume that all variables are measured (redundant system). The general procedure of error detection is divided into two parts :

- the generation of so-called residuals, which are functions of measurements that are accentuated by the errors,
- the detection, the isolation and the estimation of the error.

In the following, we present the balance residuals analysis (known as method of pseudonodes), the parity space approach, the analysis of residuals (known as method of measurement test), the analysis of the sum of squares deviations. Most of the methods which have been developed for gross error detection involve the use of statistical tests based on the assumption that the measurement errors are gaussian. After generating the set of residuals, tests for outliers are applied ; any measurement for which the residual fails the test is considered to contain a gross error.

With the previous hypothesis of a gaussian distribution of the measurement errors, one shows that the vector  $R$  follows a normal distribution with zero mean and covariance  $V_R$  :  $V_R = M V M^T$ . In order to compare the components of the  $R$  vector, let us define a normalized imbalanced vector  $R_N$  whose component  $R_N(i)$  is defined by :

$$R_N(i) = \frac{R(i)}{\sqrt{V_R(i,i)}} \quad \text{for } i = 1, \dots, n \quad (3.1)$$

Each component  $R_N(i)$  has a normal distribution with a zero mean and unity variance. Then, a simple statistical test criterion of data inconsistency can be used. From a cumulative normal distribution table the probability of  $R_N(i)$  being, for example, in the interval of 1.96 to 1.96 is read to be 0.95. Therefore, when  $|R_N| > 1.96$ , we might say that the inconsistency is significant with a probability of 0.95.

If  $R_N(i)$  exceeds the critical value, this denotes that node  $i$  is a bad node. If we assume the presence of only one gross error which affects only one stream of the process, it has been established that the node which contains this stream has the bigger residual. Then, a simple examination of the components of  $R$  shows the suspect stream.

When several gross errors are present, their location is more complicated. It has been proposed (Mah, 1982) to apply the preceding test to each node and also to the aggregates of two or more nodes, which are called pseudonodes. Note that the main assumption underlying the method is that the gross errors in two or more measurements do not cancel each other. The location of the suspect stream necessitates the fusion of not necessarily linear equations. Notice that these fusions, which correspond to the elimination of one variable between two equations, are not easy from an analytical point of view and, what is more, these fusions can not always be achieved, nor are they unique. See Kratz (1990) for an example.

### 3.2 - Parity space approach

In the absence of gross errors, the measurements depend on the true values following the linear relation (1.5). The so-called parity space is orthogonal to the space taken from the columns of  $H$ . The parity vector is related to the measurement vector  $Z$  through a projection matrix  $W$  (Potter and Suman, 1977) of dimension  $n.v$  ( $n=v-m$ ) :

$$P = W Z \quad (3.2)$$

From (1.5), written with direct measurement ( $H=I$ ), the expression (3.2) yields :

$$P = W \varepsilon \quad (3.3)$$

Consequently, the parity vector is independent of the measurement vector and contains only the errors due to the faults. As an extension it is possible to consider the constrained case, where the true values are constrained by the linear model.

Equation (3.3) shows that for normal functioning, the magnitude of parity vector is small (presence of measurement noise). If a failure occurs in only one of the sensors, then the parity vector may grow in a fixed direction associated with the failed sensor.

Moreover the components of the parity vector have the same probability distribution as the measurement errors which are independent gaussian with a zero mean value. As the variable  $c^2 = P^T P$  is the sum of the square of  $(v-m)$  normally distributed variables, it has a chi-square probability distribution with  $(v-m)$  degrees of freedom and may be compared to the threshold  $c_{1-\alpha}^2$  where  $c_{1-\alpha}^2$  is the value

$$P_j = \frac{W_j^T P}{\|W_j\|} \quad (3.4)$$

When the  $k$ th sensor is faulty, then the parity vector  $P$  follows the direction of the  $k$ th column of  $W$  which is the greatest projection of  $P$  on the axis. Then after the detection of one or several failed sensors, we locate those with the greatest projection. Next, we delete the suspect sensor and calculate the detection test after the deletion of each sensor. We stop the procedure when the magnitude of the parity vector  $P$  corresponding to the remaining sensors no longer satisfies the detection test.

### 3.3 - Standardized least square residuals analysis

The so-called measurement test is based on statistical tests for outliers applied to the residuals between measurement and estimations. The different steps are :

*Step 1.* Apply the reconciliation procedure to the data of the process and compute the estimate vector  $\hat{X}$ , the least square residuals vector  $E$  and their variance matrix  $V_E$ .

*Step 2.* Compute the standardized least square residuals by :

$$E_N = \text{diag}(V_E)^{-1} E \quad (3.5)$$

Under the hypothesis that the measured values contain no gross error,  $E_N$  has a unit variance matrix.

*Step 3.* Detect and locate the failed measurement : as for the residuals of the imbalance, each  $E_N(i)$  is compared with a critical test value (defined by the overall probability of type I error). If at least one component of  $E_N$  is out of the confidence interval, this denotes the presence of a bad stream. Locating the failed sensor one proves that it corresponds to the greatest standardized least square residual.

For the linear case, instead of (3.5), Tamhane (1982) has shown that for a non diagonal covariance matrix  $V$ , a vector of test statistics with the maximal power for detecting a single gross error is obtained by premultiplying  $R$  by  $V^{-1}$ . The power of the test (the probability of correctly detecting and identifying gross errors when they are present in the process data) has been established and discussed by Iordache (1985) under different conditions (various networks, position of the error, variance values ...) using a Monte Carlo simulation. Note that Jongenelen (1988) pointed out the case where the variance  $V$  depends on an unknown scale factor  $\sigma^2$ .

Some of the preceding results can be applied for non linear systems.

### 3.4 - Residual criterion analysis

In 1988, Crowe (1988) identified the gross errors (in the linear case) by examining the objective function. A recursive algorithm has been proposed which results from the deletion of suspect measurements. It is shown that if the deletion of a single measurement decreases the objective function (in respect to a statistical test), this measurement corresponds to a gross error.

As was pointed out by Reilly (1963), the quantity  $\Phi_r = E^T V^{-1} E$  has a chi-square distribution with the number of degrees of freedom equal to the rank  $ok M$ . Thus the imbalances can be globally tested against tabulated values of chi2.

A difficulty with this global test is that while it indicates well the presence of gross errors it is not able to identify the source of these errors. However, the use of a sequential procedure allows one to locate the streams which contain a gross error. Rinns (1962)

Rosenberg (1987) which also proposes two composite tests (extended measurement test, dynamic measurement test).

A comparison of the different methods and with several others can be found in Serth (1986). Another approach has been formulated by Yamamura (1988) also using a serial elimination ; a new criterion test has been tested which is based on the Akaike's information function and in order to minimize the amount of computation a branch-and-bound algorithm is used.

#### 4 - ESTIMATION OF MEASUREMENT VARIANCE

Most of the reconciliation techniques are based on the assumption that the measurement errors are random variables obeying a known statistical distribution. Almost without exception the techniques start with a given known covariance matrix of measurement errors. Almasy (1984) has proposed a method of estimating this matrix which makes use of available data and takes into account the reconciliation point of view. Recently, Ragot (1990) proposed also a method which allows simultaneously the estimation of the variance matrix and the reconciliation of the data on several time intervals.

This given procedure is referred as a direct method due to the fact that the covariance matrix and the reconciliation of data are estimated at the same time. For a series of p steady state point  $X_j^*$  :

$$X_{ij} = X_j^* + \varepsilon_{ij} \quad j = 1, \dots, p \quad i = 1, \dots, m_j \quad (4.1)$$

$$M X_j^* = 0 \quad j = 1, \dots, p \quad (4.2)$$

$$m = \sum_{j=1}^p m_j$$

The maximisation of the likelihood function of the measurements gives the optimality equations :

$$V = \frac{1}{m} \text{diag} \left( \sum_{j=1}^p \sum_{i=1}^{m_j} (\hat{X}_j - X_{ij}) (\hat{X}_j - X_{ij})^T \right) \quad (4.3)$$

$$\hat{X}_j = (I_v - VM^T(MVM^T)^{-1}M) \frac{1}{m_j} \sum_{i=1}^{m_j} X_{ij} \quad j = 1, \dots, p \quad (4.4)$$

A hierarchical solution for estimating V and  $\hat{X}_j$  has been proposed (Ragot 1990) : it uses a structure with two levels of hierarchy :

- The first level calculates  $\hat{X}_j$  by taking V to be known (evaluated in the second level of calculation).
- The second level calculates V as a function of the estimation  $\hat{X}_j$  transmitted by the first level.

The two levels of calculation are repeated until the estimations converge.

#### 5 - SENSOR POSITIONING

Data validation, as previously described, is considered as an analysis stage and it gives the user coherent statistical information. Additionally, the above may be completed by a more ambitious study introducing modifications of the instrumental scheme. Firstly, it is necessary to analyse the lack of information and to make the user aware of the instrumental inadequacy of certain parts of the process. Secondly, in order to make the process information fuller, an understanding of its deficiency enables the satisfactory localization of sensors. However, in practice, the introduction of such supplementary sources of information must take into account economic and technical constraints.

fying which of the measurable parameters have to be measured. Another objective is to select an optimal measurement structure to enhance estimation accuracy.

More recently, Kretsovalis (1987) developed a new approach for steady-state systems with measurement noise only. This work point out the problem on the accuracy of the estimates obtained by data reconciliation. Quantitative relations have been developed for the effects of adding and removing a single measurement on the estimation accuracy. These relationships have been utilized in developing evolutionary strategies for selecting an optimal measurement structure. For the sake of brevity, all these different approaches cannot be covered. We restricted ourselves to the presentation of an original method which can be used for the placement of sensors in systems described by linear and, by extension, bilinear equations.

#### 5.1 - The linear case

Two distinct questions can be formulated : (i) knowing that different system measurement points exist and that some streams are unmeasurable, is it possible to attain total observability by adding further sensors ? and (ii) fixing a cost function that attributes to each streams a given weight, which group of streams then must be measured to obtain total observability by minimizing the measurement cost ?

The first problem is easily solved by fixing boolean constraints which specify absolutely whether or not measurements can be made. The second one requires the introduction of continuous constraints.

The vector of true values  $X^*$  is decomposed into three distinct categories :  $X_m^*$ : vector of measured streams,  $X_f^*$ : vector of unmeasured but measurable streams (free stream),  $X_{\bar{m}}^*$ : vector of unmeasured and unmeasurable streams.

We try to solve the following problem : is the knowledge of the measurements  $X_m^*$  sufficient to obtain total observability of the process ? Elsewhere, can this observability be obtained in placing supplementary sensors on streams defined by  $X_f^*$  ?

According to the  $X^*$  decomposition, the incidence matrix of the network may also be decomposed as :

$$M = (M_m \mid M_f \mid M_{\bar{m}}) \quad (5.1)$$

The basic idea (Maquin, 1986) is to apply the transformation described in the second part of this paper independently to  $M_f$  and to  $M_{\bar{m}}$ . After extracting regular parts from  $M_{\bar{m}}$  and applying a suitable transformation to the result, in order to obtain an identity matrix, the incidence matrix can be written as :

$$M' = \left( \begin{array}{cc|cc|c} A_{1m} & A_{1f} & O & A_{1\bar{m}} & I \\ A_{2m} & A_{2f} & I & O & O \\ A_{3m} & 0 & O & O & O \end{array} \right) \quad (5.2)$$

This matrix can be read in the following manner. Remembering that the subscripts m, f and  $\bar{m}$  indicate respectively the measured streams, the free one and the unmeasurable one, line 3 of  $M'$  corresponds to the redundancy equations. Line 2 does not contain unmeasured variables. Therefore it is sufficient to fix the "f" variables intervening in the second column, to make the variable in the third column deducible. Line two is then only composed by deduction equation. Then, the first line only contains unmeasurable variables : if the matrix  $A_{1\bar{m}}$  exists, there is no possible deduction. The

## 5.2 - The bilinear case

The optimal location of measurements in linear systems is based on the counting of equations which contain more than two unknown quantities ; according to the nature of these unknown quantities (measurable or not measurable), supplementary sensors can be placed to render the system observable.

The bilinear case can be treated according to this principle. However, this is more complicated in practise because of the interaction of X and Y. This means that the observability of X or Y variables are simultaneously modified by the addition of a sensor for X or Y variables. This interaction between the different variables can be summarized by the following three rules :

1- An equation which contains only one unknown X quantity is observable. The X unknown quantity is directly deduced from a linear equation.

2- An equation which contains only one unknown Y quantity (with all the X variables known) is observable. The unknown Y quantity can be deduced from a bilinear equation.

3- An equation with all the X variables known except two of them and all the Y variables known is observable. The two unknown quantities can be deduced from the set of linear and bilinear equation.

The principle of the location of measurements is directly derived from this analysis. The addition of a new sensor is carried out in order to obtain one of the three previous situations. As the addition of a sensor modifies the observability of the system, this procedure had to be repeated until total observability is obtained.

## REFERENCES

[ALMA 84] G.A. ALMASY, R.S.H. MAH, "Estimation of measurement error variances from process data", *Ind. Eng. Chem. Proc. Des. Dev.*, vol. 23, p. 779-784, 1984.

[CLAI 75] R. CLAIR, F. LEBOURGEOIS, J.P. CAUJOLLE, G. BORNARD, "Measurement data set improvement through material balance calculation. Application to an industrial chemical plant", IFAC congress, Boston 1975.

[CROW 86] C.M. CROWE, "Reconciliation of process flow rates by matrix projection, part 2 : the non linear case", *AICHE J.*, vol. 32, n° 4, p. 616-623, 1986.

[CROW 88] C.M. CROWE, "Recursive identification of gross error in linear data reconciliation", *AICHE J.*, vol. 34, n° 4, p. 541-550, 1988.

[CUTT 76] G.W. CUTTING, "Estimation of interlocking mass balances on complex mineral beneficiation plants", *Intern. J. of Min. Proces.*, vol. 3, n° 3, p. 207-218, 1976.

[DARO 86] M. DAROUACH, "Observabilité et validation des données des systèmes de grande dimension, application à l'équilibrage de bilans de mesures", Thèse de doctorat d'Etat, Nancy, 1987.

[HOLL 89] W. HOLLY, R. COOK, C.M. CROWE, "Reconciliation of mass flow rate measurement in a chemical extraction plant", *Can. J. Chem. Eng.*, vol. 67, August, p. 595-601, 1989.

[IORD 85] C. IORDACHE, R.S.H. MAH, A.C. TAMHANE, "Performance studies of the measurement test for detection of gross errors in process data". *AICHE J.*, vol. 31, n° 7, p. 1187-1200, 1985.

[JONG 88] E.M. JONGENELEN, C.D. HEVER, G.A. VAN ZEE, "Detection of gross errors in process data using studentized residuals" *Comp Chem Eng.* vol. 12 n° 8 p. 845-847 1988

[KRET 87] A. KRETISOVALIS, R.S.H. MAH, "Observability and redundancy classification in multicomponent process networks", *AICHE J.*, vol. 33, n° 1, p. 70-82, 1987.

[KRET 88] A. KRETISOVALIS, R.S.H. MAH, "Observability and redundancy classification in generalized process networks, 1 : theorems, 2 : algorithms", *Comp. Chem. Eng.*, vol. 12, n° 7, p. 671-703, 1988.

[MACD 88] R.J. MACDONALD, C.S. HOWAT, "Data Reconciliation and Parameter Estimation in Plant Performance Analysis", *AICHE J.*, vol. 34, n°1 p. 1-9, 1988.

[MAH 82] R.S.H. MAH, A.C. TAMHANE, "Detection of gross errors in process data", *AICHE J.*, vol. 28, n° 5, p. 828-830, 1982.

[MAQU 86] D. MAQUIN, J. FAYOLLE, M. DAROUACH, J. RAGOT, "Le positionnement de capteurs dans les systèmes industriels de grande dimension", *Proc. of IMACS cong.*, Villeneuve d'Ascq, 1986.

[MAQU 87] D. MAQUIN, "Observabilité, diagnostic et validation de données des procédés industriels", Thèse de l'Université de Nancy I, novembre 1987.

[NARA 87] S. NARASIMHAN, R.S.H. MAH, "Generalized likelihood ratio method for gross error identification", *AICHE J.*, vol. 33, n° 9, p. 1514-1521, 1987.

[NOGI 72] S. NOGITA, "Statistical test and adjustment of process data", *Ind. Eng. Chem. Proc. Des. Dev.*, vol. 11, n° 2, p. 197-200, 1972.

[PATT 89] R.J. PATTON, P.M. FRANK, R.N. CLARK, "Fault diagnosis in dynamic systems : theory and application". Prentice Hall, 1989.

[POTT 77] J.E. POTTER, M.C. SUMAN, "Thresholds redundancy management with arrays of skewed instruments". *Electronic Flight Control systems*, AGARDOGRAPH 224, p. 2115-21, 1977.

[RAGO 90] J. RAGOT, M. DAROUACH, D. MAQUIN, G. BLOCH, *Validation de données et diagnostic. Traité des nouvelles technologies, série diagnostic et maintenance*, Hermès, 431 p., 1990.

[RAGO 90] J. RAGOT, D. MAQUIN, H. CHAFOUK, A. AITOUICHE, "Data reconciliation with unknown variance", 7th intern. symp. IMEKO on techn. diagnostics, Helsinki, 1990.

[REIL 63] P.M. REILLY, R.E. CARPANI, "Application of statistical theory of adjustment to material balances", 13th canadian chemical engineering conference, Montreal, October 1963.

[RIPP 62] D.L. RIPPS, "Adjustment of experimental data", *Chem. Eng. Prog.*, vol. 58, n° 10, p. 120, 1962.

[ROMA 80] J.A. ROMAGNOLI, G. STEPHANOPOULOS, "On the rectification of measurement errors for complex chemical plants, steady-state analysis", *Chem. Eng. Sc.*, vol. 35, n° 5, p. 1067-1081, 1980.

[ROSE 87] J. ROSENBERG, R.S.H. MAH, C. IORDACHE, "Evaluation of schemes for detecting and identifying gross errors in process data", *Ind. Eng. Chem. Res.*, vol. 26, n° 3, p. 555-564, 1987.

[SERT 86] J.E. SERTH, W.A. HEENAN, "Gross error detection and data reconciliation in steam-metering system". *AICHE J.*, vol. 32, n° 5, p.733-742, 1986.

[TAMH 82] A.C. TAMHANE, "A note on the use of residuals for detecting an outlier in linear regression", *Biometrika*, vol. 69, n° 2, p. 488-489, 1982.