



Data validation in large scale steady state linear systems

Mohamed Darouach, José Ragot, Jean Fayolle, Didier Maquin

► To cite this version:

Mohamed Darouach, José Ragot, Jean Fayolle, Didier Maquin. Data validation in large scale steady state linear systems. 12th IMACS World Congress on Scientific Computation, IMACS'88, Jul 1988, Paris, France. hal-00293576

HAL Id: hal-00293576

<https://hal.science/hal-00293576>

Submitted on 16 Apr 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.

DATA VALIDATION IN LARGE SCALE STEADY STATE LINEAR SYSTEMS

M. DAROUACH*, J. RAGOT**, J. FAYOLLE**, D. MAQUIN**

(*) L.A.R.A., C.R.A.N.
I.U.T. DE LONGWY
54400 LONGWY

(**) L.A.R.A., C.R.A.N., C.R.V.M.
rue du Doyen Marcel Roubault - BP 40
54500 VANDOEUVRE

Université de Nancy I et Institut National Polytechnique de Lorraine

Keywords : Data validation, observability, redundancy, mass balances, recurrent estimation technique.

ABSTRACT : The problem of data validation in large scale steady state linear system is addressed. The system is described by algebraic equations. The conditions for observability and redundancy are defined and a procedure is presented for classifying the system variables into observable, unobservable, redundant and no redundant variables. Reconciliation of the redundant process measurements is developed using a recurrent estimation technique which is well suited in real time data processing. An application of the method to material balances is presented.

INTRODUCTION

Data reliability is of fundamental importance for process diagnosis, identification and control. Measurements having undetected large, random or biased errors result in false control of process. Data validation is therefore situated between the process measurements acquisition and the decision to be taken.

This validation poses several problems amongst which are the following :

- Process models are non linear and inexact
- Measurements are subject to errors
- Not all variables are measured.

This study is limited to the case of linear steady state systems. Moreover the process models are exact as they are based on material balances. However, measurements and data available are subject to errors which may be divided into two types :

- random errors are generally taken to be independent, gaussian with a zero mean value,
- gross errors are caused by non-random events such as instrument biases, incomplete or imprecise process models.

For reasons of cost, inconvenience or technical unfeasibility, not all of the variables are measured. But we may want to determine if possible, the unmeasured process variables in term of measured ones.

Correct data validation can only be carried out if the variables have been classified. This allows us to determine the redundant and no-redundant measurements, the determinable and no-determinable unmeasured variables.

Once this classification has been made it is possible to :

- detect and locate gross errors
- diagnose errors
- validate measurements.

The complete data validation strategy can be schematised by the block diagram of figure 1.

Before presenting a strategy for data validation we first present some definitions and concepts for linear systems under steady state conditions.

CONCEPTS AND DEFINITIONS

A linear system under steady state conditions is defined by :

- A set of constraints

$$A = \{x \in R^v / Ax - b = 0\} \quad (1)$$

where A is a $n \cdot v$ matrix ($n < v$), b is a n column-vector, x is the state vector.

- A set of m measurement equations

$$Z = Hx + \varepsilon \quad (2)$$

where H is a $m \cdot v$ matrix

ε is an additive vector of measurement errors.

Observability

The system described by (1) and (2) is said to be globally observable if the state x is uniquely determined by the observations Z (equations 2) and the constraints (1).

The condition of global observability can be summarized by the following theorem [1], [2], [3].

Theorem :

The system described by (1), (2) is completely or globally observable if and only if :

$$r = \text{rank} \begin{pmatrix} H \\ A \end{pmatrix} = \dim x = v$$

If $r < v$ the system described by (1), (2) is not globally observable. In this case it can be decomposed on the basis of the observability concept. In fact, if $r < v$ there exist a non singular matrix T such that :

$$\begin{pmatrix} H \\ A \end{pmatrix}^T = \begin{pmatrix} H_1 & 0 \\ A_1 & 0 \end{pmatrix} \quad (3)$$

where $\begin{pmatrix} H_1 \\ A_1 \end{pmatrix}$ is a $(m+n) \cdot r$ matrix and $\text{rank} \begin{pmatrix} H_1 \\ A_1 \end{pmatrix} = r$.

This means that in the new coordinates y ($x = T y$) the subsystem

$$\begin{aligned} b &= A_1 y_1 \\ Z &= H_1 y_1 + \epsilon \end{aligned} \quad (4)$$

(where $y_1 \in \mathbb{R}^r$ are the first r components of y)

is observable and the remaining subsystem of dimension $(v-r)$ is unobservable.

This result can be shown by orthogonal transformation or decomposition into singular values [5].

Redundancy [1], [2], [3]

The system described by (1) and (2) is said to be redundant when the available data exceed the minimum necessary for unique determination of the system state. This definition indicates that measurements are redundant if their deletion causes no loss of observability.

The result for the decomposition by observability can be extended to that of measurements redundancy.

In this case we say that :

the system described by (1), (2) such that $n + m > v$ and $r < v$ (globally unobservable) has a redundancy $\alpha = n + m - r$ and can be decomposed into an observable and an unobservable part. The observable part can itself be decomposed into redundant and non-redundant parts. These results will be used later in the procedure for classifying the system variables in data validation by mass balances equilibration.

DATA VALIDATION

Estimation

Let us consider a steady state linear system described by (1), (2), for which we want to estimate the state x .

We make use of the following hypothesis :

- 1 - The system is globally observable : $r = v$
- 2 - Rank $A = n$
- 3 - The errors ϵ are normally distributed with a zero mean and known covariance matrix V .

Now, with these hypothesis the maximum likelihood estimator reduces to the least squares estimator subject to the constraint (1), that is :

$$\min \Phi = \| Z - H \hat{x} \|^2_{V^{-1}} \quad (5)$$

$$\text{subject to the constraint } A \hat{x} = b \quad (6)$$

Forming the lagrangian :

$$L = \Phi + \lambda^T (A \hat{x} - b)$$

where the λ is lagrangian multiplier associated with (6), the first order necessary conditions yield the $(n+v)$ simultaneous equations :

$$(H^T V^{-1} H) \hat{x} + A^T \lambda = H^T V^{-1} Z \quad (7)$$

$$A \hat{x} = b \quad (8)$$

In general $(H^T V^{-1} H)^{-1}$ does not exist ; however as the system is observable:

$$R^{-1} = H^T V^{-1} H + A^T A$$

is invertible and by multiplying (8) by A^T and by adding the result to (7), we obtain :

$$\hat{x} = R (H^T V^{-1} Z + A^T b - A^T \lambda) \quad (9)$$

substituting (9) in (8) yields :

$$\lambda = (ARA^T)^{-1} A R H^T V^{-1} Z + (I - (ARA^T)^{-1}) b \quad (10)$$

(note that (ARA^T) is regular since rank $A = n$, hypothesis 2).

Finally we obtain:

$$\hat{x} = P_r R H^T V^{-1} Z + R A^T (ARA^T)^{-1} b \quad (11)$$

$$\text{with } P_r = I - R A^T (ARA^T)^{-1} A \quad (12)$$

Statistical properties of the estimator

Under the hypothesis (3), we have the following results :

$$E(\hat{x}) = P_r R H^T V^{-1} H x + R A^T (ARA^T)^{-1} b \quad (13)$$

$$\text{Var}(\hat{x}) = P_r R H^T V^{-1} H R P_r$$

After some reductions we obtain:

$$\text{Var}(\hat{x}) = \Sigma = P_r R \quad (14)$$

From (11) and (14) we can express \hat{x} as :

$$\hat{x} = \Sigma H^T V^{-1} Z + R A^T (ARA^T)^{-1} b \quad (15)$$

Supplementary linear constraints

In the present section we consider the estimation of the state variables in linear steady state system subject to additional linear constraints. From the knowledge of the estimate without the constraints, we establish a recurrence formula which gives a new estimation of the state variables.

Consider the globally observable linear system :

$$(S_1) \quad \begin{cases} Z = H x + \epsilon \\ A_1 x = b_1 \end{cases} \quad (16)$$

with $A_1 \in \mathbb{R}^{n_1 \times v}$ and $b_1 \in \mathbb{R}^{n_1}$.

The estimate \hat{x}_1 is given by (15) :

$$\hat{x}_1 = \Sigma_1 H^T V^{-1} Z + R_1 A_1^T (A_1 R_1 A_1^T)^{-1} b_1 \quad (17)$$

where $\Sigma_1 = P_{r1} R_1$

$$R_1 = (H^T V^{-1} H + A_1^T A_1)^{-1}$$

$$P_{r1} = I - R_1 A_1^T (A_1 R_1 A_1^T)^{-1} A_1$$

with additional linear constraints, the system S_1 becomes :

$$(S_2) \quad \begin{cases} Z = H x + \epsilon \\ A_1 x = b_1 \\ A_2 x = b_2 \end{cases} \quad (18)$$

with $A_2 \in \mathbb{R}^{n_2 \times v}$ and $b_2 \in \mathbb{R}^{n_2}$

Then the state estimation problem can be formulated as :

$$\min \Phi = \| Z - H \hat{x}_2 \|_{V^{-1}}^2 \quad (19)$$

subject to the constraints :

$$\begin{cases} A_1 \hat{x}_2 = b_1 \\ A_2 \hat{x}_2 = b_2 \end{cases} \quad (20)$$

Forming the lagrangian:

$$L = \Phi + \lambda^T (A_1 \hat{x}_2 - b_1) + \mu^T (A_2 \hat{x}_2 - b_2) \quad (21)$$

where the λ 's and μ 's are lagrangian multipliers associated with (20) ; the first order necessary conditions yield the $(n_1 + n_2 + v)$ simultaneous equations:

$$(H^T V^{-1} H) \hat{x}_2 + A_1^T \lambda + A_2^T \mu = H^T V^{-1} Z \quad (22)$$

$$A_1 \hat{x}_2 = b_1 \quad (23)$$

$$A_2 \hat{x}_2 = b_2 \quad (24)$$

using expression (17) we obtain :

$$\hat{x}_2 = \Sigma_1 H^T V^{-1} Z + R_1 A_1^T (A_1 R_1 A_1^T)^{-1} b_1 - \Sigma_1 A_2^T \mu \quad (25)$$

From (25) and (24) we have:

$$\hat{x}_2 = (I - \Sigma_1 A_2^T (A_2 \Sigma_1 A_2^T)^{-1} A_2) \hat{x}_1 + \Sigma_1 A_2^T (A_2 \Sigma_1 A_2^T)^{-1} b_2 \quad (26)$$

Statistical properties of the estimate \hat{x}_2 :

We have:

$$E(\hat{x}_2) = (I - \Sigma_1 A_2^T (A_2 \Sigma_1 A_2^T)^{-1} A_2) E(\hat{x}_1) + \Sigma_1 A_2^T (A_2 \Sigma_1 A_2^T)^{-1} b_2 \quad (27)$$

$$\text{Var}(\hat{x}_2) = \Sigma_2 = \Sigma_1 - \Sigma_1 A_2^T (A_2 \Sigma_1 A_2^T)^{-1} A_2 \Sigma_1 \quad (28)$$

Generalization

More generally, if we call \hat{x}_k the estimation and Σ_k its variance in presence of the constraint $A_k \hat{x} = b_k$, the new estimation \hat{x}_{k+1} and its variance can be established in terms of the additional constraint $A_{k+1} \hat{x} = b_{k+1}$

where :

$$\hat{x}_{k+1} = (I - \Sigma_k A_{k+1}^T (A_{k+1} \Sigma_k A_{k+1}^T)^{-1} A_{k+1}) \hat{x}_k + \Sigma_k A_{k+1}^T (A_{k+1} \Sigma_k A_{k+1}^T)^{-1} b_{k+1} \quad (29)$$

and

$$\Sigma_{k+1} = \Sigma_k - \Sigma_k A_{k+1}^T (A_{k+1} \Sigma_k A_{k+1}^T)^{-1} A_{k+1} \Sigma_k \quad (30)$$

This recurrence is especially useful if the matrix A_{k+1} reduces to a row vector. In this case, calculations of Σ_{k+1} and consequently \hat{x}_{k+1} only requires the inversion of a scalar.

APPLICATION TO MASS BALANCES

The described procedure was applied to validate measurements using mass balances equations. In this case the system is represented by directed graph. Its arcs represent the process streams and its nodes represent the units (tanks, reactors flotation cells, etc.)

The system can be represented by :

- Overall mass balances equations

$$A x = 0 \quad (31)$$

where A is the incidence matrix

x is the vector of total flow rate

- Measurement equations

$$Z = H x + \varepsilon = [I_m | O] x + \varepsilon \quad (32)$$

where I_m denotes the identity $m \cdot m$ matrix.

According to the above partition of the matrix H, the matrix A and the vector x are partitioned as well,

$$A = (A_m | A_{\bar{m}})$$

and

$$x = \begin{pmatrix} x_m \\ x_{\bar{m}} \end{pmatrix}$$

With these partitions the balance equation (32) becomes :

$$A_{\bar{m}} x_{\bar{m}} + A_m x_m = 0 \quad (33)$$

The observability of the system (31), (32) reduces to that of system (33). The procedure (3) used previously can be applied to (33). However the structure of the incidence matrix is sparse and only contains +1's and -1's ; an echelon form transformation is more suitable [4]. Thus if rank $A_{\bar{m}} = r$, a permutation of the rows and columns of $A_{\bar{m}}$ allows the regular part $A_{\bar{m}12}$ to be isolated.

After these transformations, the matrix A can be expressed as [5]:

$$A = \begin{pmatrix} A_{\bar{m}12}^{-1} & A_{\bar{m}12}^{-1} A_{\bar{m}11} & I \\ A_{m2} - A_{\bar{m}22} A_{\bar{m}12}^{-1} A_{m1} & O & O \end{pmatrix} \quad (34)$$

The analysis of this matrix point out the observability conditions and the redundancy of all the variables :

- Block 4 (second row, first column) concerns the equations of redundancy as it only includes measured variables. The corresponding incidence matrix is :

$$A_r = A_{m2} - A_{\bar{m}22} A_{\bar{m}12}^{-1} A_{m1} \quad (35)$$

- If the block 2 (first row, second column) contains rows of zeros then the rows of the same rank in block 3 (identity matrix) give the indices of the unmeasured variables which can be determined.
- The measured variables, which can not be estimated correspond to the columns of block 4 composed of all zero elements.
- The decision variables of the global observability are the no zero elements of block 2.

The procedure using these results to validate data by mass balances can be schematized by figure 2.

This diagram shows the decomposition of the initial problem (31), (32) into two smaller sub-problems which can be solved sequentially.

First sub-problem

$$\begin{aligned} \min \Phi &= \|Z - \hat{x}_m\|_{V^{-1}}^2 \\ \text{subject to } A_r \hat{x}_m &= 0 \end{aligned}$$

is solved by using the previous recurrent estimation technique (29), (30). Taking into account that $H = I$ and $b = 0$, we obtain :

$$\hat{x}_m = \hat{V}_m V^{-1} Z \quad (37)$$

with

$$\hat{V}_m = \text{Var}(\hat{x}_m) = V - V A_r^T (A_r V A_r^T)^{-1} A_r V \quad (38)$$

in practice A_r is partitioned in the form :

$$A_r = \begin{pmatrix} A_i \\ A_{i+1} \end{pmatrix}$$

the state estimate for the system described by A_i is :

$$\hat{x}_i = \hat{V}_i V^{-1} Z \quad (39)$$

Adding the constraint corresponding to the matrix A_{i+1} , we obtain the new estimate :

$$\hat{x}_{i+1} = \hat{V}_{i+1} V^{-1} Z \quad (40)$$

with

$$\hat{V}_{i+1} = \hat{V}_i - \hat{V}_i A_{i+1}^T (A_{i+1} \hat{V}_i A_{i+1}^T)^{-1} A_{i+1} \hat{V}_i \quad (41)$$

When the partitioning is performed according to the nodes of the system, each block A_i becomes a row vector. Matrix \hat{V}_m can be calculated in n iterations, each of which only requires the inversion of a scalar.

We apply (41) for $i = 0, n-1$ with $\hat{V}_0 = V$ and $\hat{V}_m = \hat{V}_n$

Second sub-problem

Once the level 1 of validation has been achieved, the estimation of the observable part of the unmeasured variables can be made by direct deduction from the estimate variables of the first level and the relationships corresponding to the rows of the elements in the matrix $A_{m12}^{-1} A_{m11}$.

CONCLUSION

The method of data validation presented here takes into account the topology of a process (its structure) and the data which are available (the measurements, their location and their precision). It involves decomposing the problem of validation into two smaller sub-problems. In the case of large scale systems a recurrence formula provides an estimation of the redundant variables.

There are a great number of practical applications of this method. Particular mention may be made of the diagnosis of industrial systems in which the process configuration and the location of the measurement points change with time. The extension to the case of bi-linear systems will be treated in a later study.

REFERENCES

- 1 - G.M. Stanley, R.S.H. Mah, "Observability and redundancy in process data estimation". *Chem. Eng. Sciences*, Vol. 36, pp. 259-272, 1981
- 2 - J.A. Romagnoli, "On data reconciliation : Constraints processing and treatment of bias". *Chem. Eng. Sciences*, Vol. 38, pp. 1107-1117, 1983.
- 3 - M. Darouach, J. Ragot, D. Maquin, J. Fayolle, "Validation des mesures par équilibrage hiérarchisé de bilans matière. Concepts et méthodologie". *International Journal of Mineral Processing*, Vol. 17, pp. 273-285, 1986.
- 4 - M. Darouach, "Observabilité et validation des données de systèmes de grande dimension. Application à l'équilibrage des bilans de mesures". Thèse d'état, Université de Nancy I, Nancy, juin 1986.
- 5 - G.H. Golub, G.A. Meurant, "Résolution numérique des grands systèmes linéaires". Collection de la division des études et recherches E.D.F., Eyrolles 1983.
- 6 - D. Maquin, M. Darouach, J. Fayolle, J. Ragot, "Le positionnement des capteurs dans les systèmes industriels de grande dimension". Congrès I.M.A.C.S. 86, Villeneuve d'Ascq, 3 - 6 June 1986. Applied Modelling and Simulation of Technological Systems, Elsevier 1987.

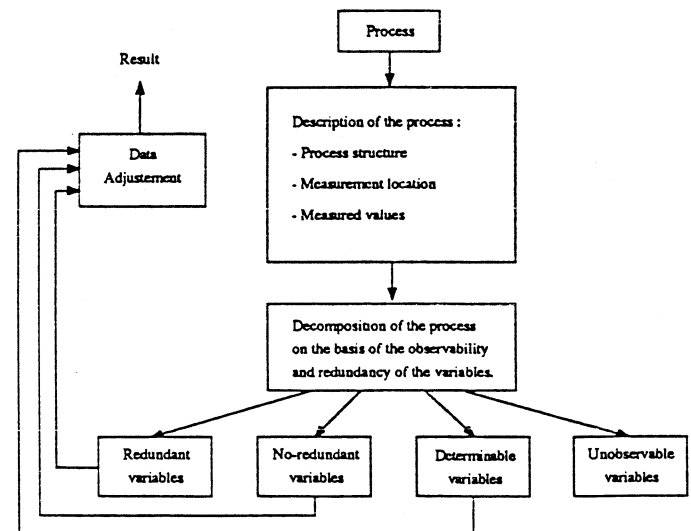


Figure 1 : General strategy for data adjustment

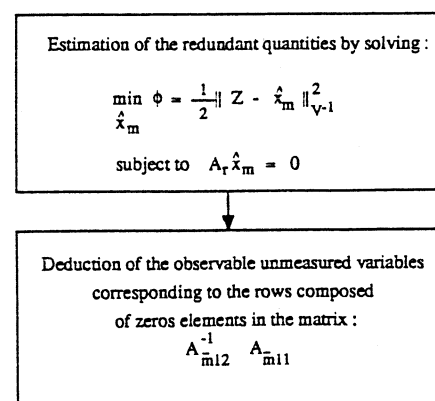


Figure 2 : Schematic diagram of the procedure for data validation by linear mass balances.

Numerical example : Mass balance of flotation circuit.

The circuit considered in this example is described by the flowsheet of figure 3, it consists of 22 nodes and 32 streams. The measurements location of the flow rates are indicated by x. The results of the data validation are summarized in tables 1 and 2.

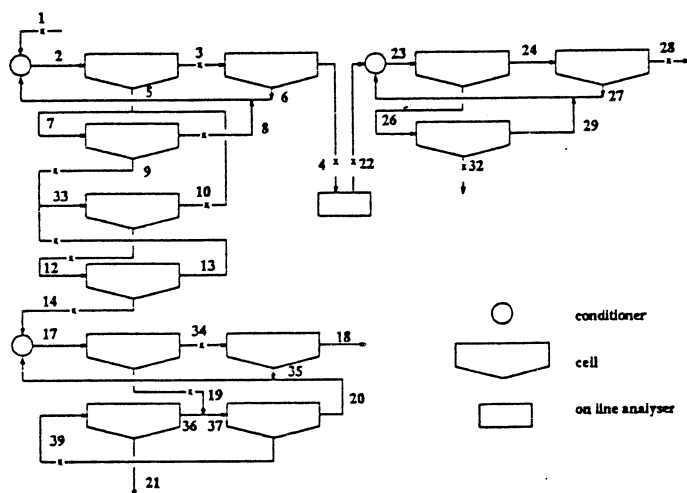


Figure 3 : Flotation circuit

Stream	Measurement	Estimation	Standard deviation	Estimated standard deviation
1	4,683	4,595	0,234	0,096
6		2,242		0,322
8	0,531	0,531	0,027	0,027
2		7,368		0,310
3	6,166	6,166	0,308	0,308
5		1,202		0,036
4	4,000	3,924	0,200	0,094
10	0,300	0,295	0,015	0,014
7		1,497		0,037
9	0,917	0,966	0,046	0,026
13	0,264	0,271	0,013	0,013
33		1,237		0,028
12	1,050	0,942	0,053	0,026
14	0,652	0,671	0,033	0,024
35				
20				
17		1,115		0,042
19	0,756	0,756	0,038	0,038
34	0,359	0,359	0,018	0,018
18				
39	0,550	0,550	0,028	0,028
21				
36				
37				
22	3,700	3,924	0,185	0,094
27				
29				
23				
24				
26				
28	3,000	2,932	0,150	0,097
32	1,000	0,992	0,050	0,048

Table 2 : Flotation mass balance results.

Stream	Measured		Unmeasured	
	Redundant	No-redundant	Determinable	Undeterminable
1	x	-	-	-
6	-	-	x	-
8	-	x	-	-
2	-	-	x	-
3	-	x	-	-
5	-	-	x	-
4	x	-	-	-
10	x	-	-	-
7	-	-	x	-
9	x	-	-	-
13	x	-	-	-
33	-	-	x	-
12	x	-	-	-
14	x	-	-	-
35	-	-	-	x
20	-	-	-	x
17	-	-	x	-
19	-	x	-	-
34	-	x	-	-
18	-	-	-	x
39	-	x	-	-
21	-	-	-	x
36	-	-	-	x
37	-	-	-	x
22	x	-	-	-
27	-	-	-	x
29	-	-	-	x
23	-	-	-	x
24	-	-	-	x
26	-	-	-	x
28	x	-	-	-
32	x	-	-	-

Table 1 : Result of variables classification.