



Detection of bias in measurement using analytical redundancy

Frédéric Kratz, José Ragot, Didier Maquin, Antoine Despujols

► To cite this version:

Frédéric Kratz, José Ragot, Didier Maquin, Antoine Despujols. Detection of bias in measurement using analytical redundancy. 7th International Symposium on Technical Diagnostics, Sep 1990, Helsinki, Finland. hal-00293579

HAL Id: hal-00293579

<https://hal.science/hal-00293579>

Submitted on 15 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.

DETECTION OF BIAS IN MEASUREMENT USING ANALYTICAL REDUNDANCY

F. KRATZ*, J. RAGOT*, D. MAQUIN*, A. DESPUJOLS**

* Centre de Recherche en Automatique de Nancy
Centre de Recherche sur la Valorisation des Minerais
BP 40 - Rue du Doyen Marcel Roubault
54 501 Vandœuvre les Nancy - FRANCE

** Electricité de France - DER
6, quai Watier
78 400 Chatou - FRANCE

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 use in preference 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. These validation techniques must be implemented to increase systems reliability and to facilitate detection of failures. This paper presents two methods of detection and location of measurement failures : the first is the standardized imbalance residuals approach and the second is the approach based on standardized least square residuals. These techniques are presented by using linear models that we extend to non-linear models. We assume that all variables are measured (redundant system), that the measurement errors are ruled by a zero mean normal distribution and known variances, and that the process model is correct.

A - LINEAR SYSTEM

I - STANDARDIZED IMBALANCE RESIDUALS APPROACH

I.1 - Problem formulation

In practice the raw process data may also contains other types of errors which are caused by non-random events. The presence of these errors invalidates the balance equations. It is then necessary to examine the residuals of the imbalance. The R vector of imbalance is generated according to the following equation :

$$R = M X \quad (1)$$

where M is the incidence matrix of the system network with n nodes and v streams.

The measurement vector X and the measurement error vector ϵ are connected with the true values vector X^* by the model :

$$X = X^* + \epsilon \quad (2)$$

Under the previous hypothesis relative to the measurement errors, one can demonstrate [1] that R vector is normally distributed with zero mean and a variance matrix $V_R = M V M^T$. In order to compare the components of the R vector, let us define a normalized imbalanced vector R_N with component $R_N(i)$ defined by :

$$R_N(i) = \frac{R(i)}{\sqrt{V_R(i,i)}} \quad \text{for } i = 1, \dots, n \quad (3)$$

Each component $R_N(i)$ has a normal distribution with a zero mean and unity variance. Then, a 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 0.05 .

Nodal imbalances result, if at least an adjacent stream is faulty. An error identification scheme is the nodal aggregation procedure used by R.S.H. Mah [2]. The proposed algorithm consists of three steps. In the first step our aim is to isolate the "bad" streams. We calculate each normalized imbalanced vector component and for each component, we shall associate a logical variable with each stream. Then a value of "1" will be assigned to a stream adjacent to a "good" node, and a value of "0" to a stream adjacent to a "good" node. Otherwise no value will be assigned for the other streams. The second step is the same as first for each aggregated node generated from the "bad" nodes. Finally, the product of the logical variable for each stream permits the detection of the faulty stream. Then a value of "0" represent a "good" stream and a value of "1" a "bad" stream.

I.2 - Example

To illustrate the structure of the standardized imbalance residuals approach, we now consider the simple example system shown in figure 1. For this example, the streams 3 and 7 are faulty. So in the first step, we find that all nodes are "bad". We can apply the test to the balance around to the aggregated nodes, $I + II$, $I + III$, $II + III$, $III + IV$, $I + III + IV$, $II + III + IV$, $I + II + III + IV$ (environment node).

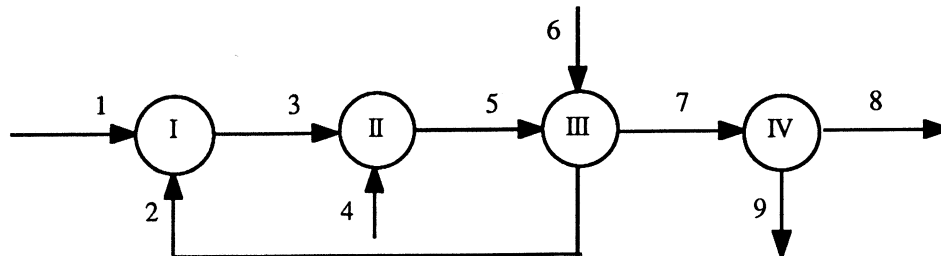


Figure 1

The results of this fault detection are shown in table 1. The last line of table 1 gives the final logical variable for each stream.

	"bad" node	1	2	3	4	5	6	7	8	9
I	yes	1	1	1
II	yes	.	.	1	1	1
III	yes	.	1	.	.	1	1	1	.	.
IV	yes	1	1	1
I + II	no	0	0	.	0	0
I + III	yes	1	.	1	.	1	1	1	.	.
II + III	yes	.	1	1	1	.	1	1	.	.
III + IV	no	.	0	.	.	0	0	.	0	0
I + III + IV	yes	1	.	1	.	1	1	.	1	1
II + III + IV	yes	.	1	1	1	.	1	.	1	1
Environment node	no	0	0	.	.	.	0	.	0	0
Logical product		0	0	1	0	0	0	1	0	0

Table 1

In the present application, we find the "bad" streams 3 and 7. In other cases, the location of faults is not so easy. Some of the streams can be identified as bad streams, but the remaining streams cannot be so identified, and thus are potentially bad streams. For our example, a fault on the stream 8 is also located to the stream 9 (symmetrical stream in respect to node IV).

II - STANDARDIZED LEAST SQUARE RESIDUALS ANALYSIS

II.1 - Presentation of method

The location of the fault necessitates the generation of additional equations to obtain information on the streams. This information can be given directly by the analysis of the least square residual which gives an idea of the quality of the estimates.

II.2 - Principle of method

In the absence of gross errors, we can use the model (2) which links the measurement vector X with the true values vector X^* and the measurement error vector ϵ . The estimation problem can be solved as :

$$\min \phi = \frac{1}{2} \| \hat{X} - X \|_{V^{-1}}^2 \text{ subject to the constraint : } M\hat{X} = 0 \quad (4)$$

Applying the Lagrange multiplier technique, the least square estimate of X^* is written as :

$$\hat{X} = (I_v - VM^T(MVM^T)^{-1}M) X \quad (5)$$

The least square residual vector E_C of dimension v is :

$$E_C = X - \hat{X} = VM^T(MVM^T)^{-1} MX \quad (6)$$

The expected value of E_C is zero and the variance-covariance matrix V_E is :

$$V_E = VM^T(MVM^T)^{-1} MV \quad (7)$$

Also, E_C follows a normal distribution with zero mean and variance V_E . To compare the different components of E_C , we define the E_N standardized least square vector which follows a normal distribution with zero mean and unit variance :

$$E_N(i) = \frac{E_C(i)}{\sqrt{V_E(i,i)}} \text{ for } i = 1, \dots, v \quad (8)$$

As for the residuals of the imbalance, each $E_N(i)$ is compared with a critical test value (defined by the overall probability α of error type I). 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 [1]. The proposed method is an iterative method. At each stage, the measurement corresponding to the largest normalized residual is deleted (by nodal aggregation) ; this measure is identified as containing a gross error. The test takes into account the effect of modification of number of stream. We choose the overall probability β instead of α [3], with $\beta = 1 - (1-\alpha)^{1/v}$ where v is the number of tested measurements.

B - NON-LINEAR SYSTEM

The two methods for detection and location of gross errors explained above are applied in the same way to non-linear models. The additional hypothesis which we formulate is that the algebraic constraint equations are continuously differentiable and monotonous in interval where the measures are defined. In the general case, the process model can be written :

$$f(X^*) = 0 \quad (9)$$

where f is an n -vector of non-linear functions and X^* is a set a true values.

I - STANDARDIZED IMBALANCE RESIDUALS ANALYSIS

I.1 - Presentation of method

As in the linear case, the measurement vector X does not satisfy the constraint equations. The constraint residuals vector, or imbalance residuals vector, R is given by :

$$R = f(X) \quad (10)$$

With the previous hypothesis, one shows that the vector R follows a normal distribution with zero mean and covariance V_R ($V_R = g(X) V g(X)^T$, where $g(X)$ is the Jacobian of f [4]).

One defines a normalized imbalanced vector R_N as in the linear case (3), if $R_N(i)$ exceeds the critical value, this denotes that node i is a bad node ; 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.

If we note S_i the set of variables index whose the i th equation explicitly depending and S_j^Λ ($= S_i \cap S_j$) the intersection of two sets of variables index of two equations, the fusion between the i th equation and j th equation is possible only if $S_{i,j} \neq \emptyset$. Then, for each equation

we seek to explain (if it is possible) the variables $X(p)$ ($p \in S_{i,j}$) as a function of other variables $X(i)$ ($i \in S_i, i \neq p$).

Generally, the fusion of two equations induces the elimination of a single variable $X(p)$, on condition that one can interpret this variable as a function of other variables $X(i)$ from at least one equation. We observe that the fusion of two equations is possible by many methods, depending on the variables that we eliminate. The resultant equation of the fusion comes by the substitution of the relation in another equation.

To illustrate this method, we now consider the following example :

$$f_I(X) = X(1) - \log(X(2)) - X(3)X(2)^2$$

$$f_{II}(X) = X(4) - X(2)^2X(5)$$

$$S_1 = \{1, 2, 3\}, S_2 = \{2, 4, 5\}, S_{1,2} = \{2\}$$

By fusion of I and II, we suppress the variable 2. We seek to explain this variable $X(2)$ as a function of other variables. From the equation I, it is not possible to obtain this relation ; but, from the equation II we get $X(2)$ that we substitute in the equation I. So, the resultant equation is :

$$f_{I+II}(X) = X(1) - \frac{1}{2} \log\left(\frac{X(4)}{X(5)}\right) - X(3)\frac{X(4)}{X(5)}$$

As in the linear case, the search for the faulty stream necessitates three stages (same in the linear case). The first is the computation of the normalized imbalanced vector and the search for components out of the confidence interval, this denotes the "bad" nodes. The next step is the same as the first one for each fusion of "bad" equations. Finally, the last stage permits the location of "bad" stream by the production of the logical variables (see A-I-1).

I.2 - Example

Consider the network, shown in figure 2, related by :

$$f_I(X) = 0.0045 X(1) X(2)^2 - X(3)$$

$$f_{II}(X) = X(3) - 280.86 \frac{X(2)}{X(4)}$$

$$f_{III}(X) = X(4) - \log(X(6)) - X(6)^2X(5)$$

$$f_{IV}(X) = X(6)^2 - X(6)X(7) - X(8)$$

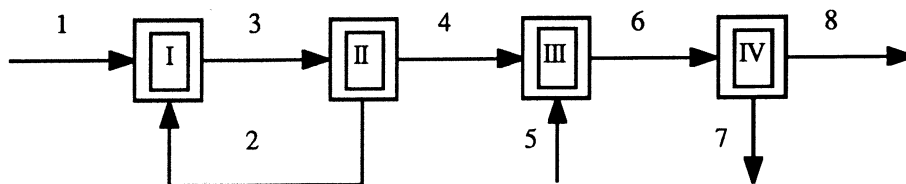


Figure 2

For this simulation example, the true values vector X^* (which satisfies the constraint equations) and the measurements vector X are :

$$X^* = [7.29 \ 58.39 \ 111.85 \ 146.62 \ 0.037 \ 62.18 \ 60.71 \ 91.40]^T$$

$$X = [7.18 \ 60.31 \ 173.13 \ 141.49 \ 0.039 \ 77.46 \ 58.04 \ 95.57]^T$$

We define the measurement covariance matrix as a diagonal matrix (the measurements are independent), where the i th term is :

$$V(i,i) = (0.05 X(i))^2 \quad (11)$$

Stage 1 : Computation of residuals vector

$$R = f(X) = [- 55.61 \ 53.41 \ - 96.86 \ 1409.48]^T$$

Stage 2 : Computation of the standardized imbalance residuals

$$R_N = [- 3.53 \ 4.41 \ - 3.57 \ 3.22]^T$$

Stage 3 : Search for "bad" nodes

For the overall probability $\alpha = 0.05$, we have a confidence interval $[- 1.96 , 1.96]$. We look for the component out of the confidence interval and we find all nodes are "bad".

Stage 4 : Aggregation procedure

We compute the following fusions :

$$\begin{aligned} & (I + II)_2, \text{ where the subscript refer the variable deleted by fusion} \\ & (I + II)_3, II + III, III + IV, (I + II)_2 + III, II + III + IV \\ & (I + II)_2 + III + IV, (I + II)_3 + III + IV \end{aligned}$$

The search for the fault is summarized in table 2. The method used is the same as the linear case.

	"bad" node	1	2	3	4	5	6	7	8
I	yes	1	1	1
II	yes	.	1	1	1
III	yes	.	.	.	1	1	1	.	.
IV	yes	1	1	1
(I + II) ₂	yes	1	.	1	1
(I + II) ₃	no	0	0	.	0
II + III	yes	.	1	1	.	1	1	.	.
III + IV	no	.	.	.	0	0	.	0	0
(I + II) ₂ + III	yes	1	.	1	.	1	1	.	.
II + III + IV	yes	.	1	1	.	1	.	1	1
(I + II) ₂ + III + IV	yes	1	.	1	.	1	.	1	1
(I + II) ₃ + III + IV	no	0	0	.	.	0	.	0	0
Logical product		0	0	1	0	0	1	0	0

Table 2

The streams 3 and 6 are identified as "bad" streams. We must notice that the location of the failures can be more difficult. Besides the problem of symmetry in the linear case, the problem of the fusion of non-linear equations is not very easy and is generally difficult to

automate. The utilization of formal language should permit, to a certain extent, the resolution of this problem.

II - STANDARDIZED LEAST SQUARE RESIDUALS ANALYSIS

II.1 - Presentation of method

Besides the problem of fusions of non-linear equations, the supplementary problem that we meet in this method is the computation of the estimate of true values vector. Newton - Raphson type iteration coupled with numerical elimination of n variables as a function of the v' - n other variables, is found to be an effective solution strategy [5] [6]. Notice that in the linear case, we have an exact solution by matrix resolution.

The different iterations are :

- compute the estimate vector \hat{X} ,
- compute the least square residuals vector $E_C = X - \hat{X}$,
- compute the covariance matrix : $V_E = V g(\hat{X})^T [g(\hat{X}) V g(\hat{X})^T]^{-1} g(\hat{X}) V$,
- compute the standardized least square residuals as in the linear case (8)
- detect and locate the failed measurement.

II.2 - Example

Consider the network in figure 2 :

Step 1 : Compute the estimate vector

$$\hat{X} = [7.11 \ 63.49 \ 128.95 \ 138.27 \ 0.04 \ 60.96 \ 59.39 \ 95.63]^T$$

Step 2 : Compute the standardized least square residuals vector

$$E_N = [0.32 \ -1.40 \ 7.10 \ 0.56 \ 2.71 \ 4.60 \ -0.55 \ -0.52]^T$$

Step 3 : Location of the failed stream

For $\alpha = 0.05$ and $v' = 8$, we have β equals 0.0064 and a confidence interval $[-2.72 , 2.72]$. One observes that the 3th measurement corresponding to the largest value of $|E_N(i)|$ out of the confidence interval, denotes that stream 3 is a "bad" stream.

Step 4 : Elimination of the faulty measure

By fusion of equations I and II : $(I + II)_3$, we delete the variable 3.

Step 5 : Compute the estimate vector

$$\hat{X} = [7.08 \ 62.07 \ 142.10 \ 0.04 \ 61.57 \ 60.02 \ 95.64]^T$$

Step 6 : Compute the standardized least square residuals vector

$$E_N = [0.47 \ -0.73 \ -0.10 \ 2.41 \ 4.43 \ -0.80 \ -0.64]^T$$

Step 7 : Location of the failed stream

For $\alpha = 0.05$ and $\nu' = 7$, we have β equals 0.0073 and a confidence interval $[-2.68, 2.68]$. One observes that the 6th measure corresponding to the largest value of $|E_N(i)|$ out of the confidence interval, denotes that stream 6 is a "bad" stream.

Step 8 : Elimination of the faulty measure

By fusion of equations III and IV, we delete the variable 6.

Step 9 : Compute the estimate vector

$$\hat{X} = [7.19 \ 61.15 \ 141.82 \ 0.04 \ 58.20 \ 95.57]^T$$

Step 10 : Compute the standardized least square residuals vector

$$E_N = [-0.07 \ -0.34 \ -0.05 \ 0.45 \ -0.07 \ -0.04]^T$$

Step 11 : Location of the failed stream

For $\alpha = 0.05$ and $\nu' = 6$, we have β equals 0.0085 and a confidence interval $[-2.63, 2.63]$. One observes that all measures are "good".

We find again, as in the standardized imbalance method, that the streams 3 and 6 are at fault.

CONCLUSION

This paper seeks to stress the problem of the detection and location of errors in a non-linear system. From the techniques applied to a linear system [7], we present, with an example, an extension of these methods applied to a non-linear system. If in a linear case, these methods are easy to apply, in a non-linear case the most important problem remains the fusion of equations.

REFERENCES

- [1] J. FAYOLLE ; Validation de données et diagnostic des procédés industriels, Thèse de Doctorat de l'Université de Nancy I (1987).
- [2] R.S.H. MAH, G.M. STANLEY and D. DOWNING ; Reconciliation and rectification of process flow and inventory data, Ind. Eng. Chem., Process Des. Dev., Vol. 15, n° 1 (1976).
- [3] R.S.H. MAH, A. C. TAMHANE ; Detection of gross errors in process data, AIChE Journal, Vol. 28, n° 5 (1982).
- [4] C.L.E. SWARTZ ; Data reconciliation for generalized flowsheet applications, American Chemical Society National Meeting, Dallas (1989).
- [5] H. SCHWETLICK, W. SCHELLONG and V. TILLER ; Gauss-Newton like methods for non-linear least squares with equality constraints - Local convergence and applications to parameter estimation in implicit models, Statistics, Vol. 16, n° 2 (1985).
- [6] J. LEGRAS ; Algorithmes et programmes d'optimisation non-linéaires avec contraintes, Editions Masson (1980).
- [7] A. AITOUCHE, F. KRATZ, D. MAQUIN, J. RAGOT ; Détection et localisation d'erreurs de mesure par la technique de l'espace de parité, Congrès Ciame - Capteurs 89, Paris (1989).