Chem. Eng. Comm., 2000, Vol. 178, pp. 1–20
Reprints available directly from the publisher
Photocopying permitted by license only

REMOVING SINGULARITIES AND ASSESSING UNCERTAINTIES IN TWO EFFICIENT GROSS ERROR COLLECTIVE COMPENSATION METHODS

MIGUEL BAGAJEWICZ[a,*], QIYOU JIANG[a] and MABEL SÁNCHEZ[b]

[a] *School of Chemical Engineering and Materials Science, University of Oklahoma, Norman, OK, USA;* [b] *Planta Piloto de Ingenieria Química (UNS-CONICET), Camino La Carrindanga. Km 7 (8000), Bahía Blanca-Argentina*

(Received 23 June 1998; In final form 30 March 1999)

Two collective estimation strategies, the Unbiased Estimation Technique (Rollins and Davis, 1992) and the recursive Generalized Likelihood Ratio (Keller *et al.*, 1994), have been shown to be very efficient in detecting and estimating multiple gross errors in linear process systems. However, these strategies run into singularities and uncertainties that prevent them from being used in automatic schemes. This paper uses a recently presented theory on gross error equivalency to explain when and how these singularities and uncertainties take place. The procedures presented by these two methods are modified to prevent the singularities from appearing and allowing their automatic implementation.

Keywords: Gross error detection; data reconciliation; bias estimation; leak estimation

INTRODUCTION

Measurements in process plants contain errors, random and systematic. To make these measurements abide by basic conservation laws, these errors are eliminated by the application of data reconciliation techniques. Systematic errors as well as process leaks, which have been together called gross errors, have been the object of several publications in the past three decades. It is desired to obtain the exact location of gross errors and it is advantageous to have an estimate of their size. The former helps instrument maintenance in

*Corresponding author.

the case of biases and process monitoring in the case of leaks. The latter constitutes valuable information for production accounting.

All compensation techniques that deal with gross errors rely on two steps, identification and size estimation. Three kinds of strategies have been developed in identifying multiple gross errors: Serial Elimination (Ripps, 1965; Serth and Heenan, 1986; Rosenberg *et al.*, 1987) which identifies one gross error at a time using a test statistic and eliminates the corresponding measurement until no gross error is detected; Serial Compensation (Narasimhan and Mah, 1987) which identifies the gross error and its size, compensates the measurement and continues until no error is found; Simultaneous or Collective Compensation (Keller *et al.*, 1994; Kim *et al.*, 1997; Sánchez and Romagnoli, 1994), which proposes the estimation of all gross errors simultaneously. In addition, Bagajewicz and Jiang (1998) proposed a collective compensation strategy for dynamic systems that can be used for steady state cases. Finally the Unbiased Estimation Technique (UBET) proposed by Rollins and Davis (1992) makes the identification first and then a simultaneous estimation.

This paper concentrates on two published collective compensation methods, the Unbiased Estimation Technique (UBET; Rollins and Davis, 1992) and a modification of the Generalized Likelihood Ratio (GLR, Narasimhan and Mah, 1987) strategy (Keller *et al.*, 1994), which we call Collective GLR (CGLR), because size estimation is done for all gross errors at the same time. UBET relies first on an identification step that requires determining a certain number of gross error candidate locations. In a second step, it performs an estimation of the gross error size and it applies Bonferroni tests to determine if the estimates are statistically significant. The CGLR technique is based on the Generalized Likelihood Ratio test that locates one gross error in each step and compensates/estimates all gross errors previously identified.

Both these techniques have shown problems for special cases of locations of gross errors. These problems consist mainly of singularities of matrices that need inversion. These problems will be explained in detail in view of the Gross Error Equivalency Theory recently developed by Bagajewicz and Jiang (1998). This theory is also used to modify certain steps of both methods so that automatic implementation of both techniques is possible. After these modifications are presented, the two methods are numerically compared.

UBET and CGLR are first briefly reviewed and special sections describe the problems encountered. The theory of gross error equivalency is summarized next and in the following section, the modifications needed to avoid the problems are introduced. Finally, the methods are compared.

REVIEW OF UBET

In this section, we review the unbiased estimation technique presented by Rollins and Davis (1992). In this paper, the basic elements of the technique are presented and suggestions are made on how to implement it. These suggestions have not been put into the form of a step by step recipe by these authors. Therefore, the description that follows as well as the step-by-step summary of the method is our own construction on how best to implement the method. The UBET is developed from the balance residuals:

$$r = Ay \tag{1}$$

and its expected value

$$\mu_r = A\delta + M\gamma \tag{2}$$

where

$$M = [m_1, \dots, m_q] \tag{3}$$

and p, q are the number of measured variables and constraint equations respectively; δ, γ the unknown $(p \times 1)$ measurement biases and $(q \times 1)$ leaks, A is a $(q \times p)$ constraint matrix with rank $(A) = q$, y is a $(p \times 1)$ vector of measurements and m_j is a $(q \times 1)$ vector with zeros in every position but a 1 in the jth.

By partitioning A, M, δ, γ and assuming there are always q gross errors, one can get:

$$\mu_r = \begin{bmatrix} A_{11} & 0 \\ A_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \delta_1 \\ \gamma_2 \end{bmatrix} = C_1\theta_1 \tag{4}$$

Finally, by introducing

$$l_i^T = \mathrm{e}_i^T C_1^{-1} \tag{5}$$

one obtains

$$l_i^T \mu_r = \mathrm{e}_i^T \theta_1 = \theta_i \tag{6}$$

Thus, $l_i^T r(i = 1, \dots, q)$ are unbiased estimators of the components of δ and γ contained in θ_1.

The procedure for applying UBET can be then summarized as follows:

(1) Use a gross error identification strategy, like the nodal strategies reported by Mah *et al.* (1976) and Heenan and Serth (1986), to isolate the suspect nodes and also construct the candidate bias/leak list from the suspect nodes.
(2) Obtain θ_1 with elements no more than the number of constraint equations (q).
(3) Construct C_1 with rank equal to q.
(4) Obtain the size estimation for the elements in θ_1.
(5) Use the Bonferroni Test to identify the gross errors.

Although the original paper suggests a fixed number of q gross errors, as the basis for the method, Rollins (1998) have suggested that with appropriate modifications, the method can be applied to a lower number of gross errors candidates. These modifications are not explored in this paper. In addition, he recommends the use of linear combination of tests (LCT), as suggested by Rollins *et al.* (1996).

SOME DIFFICULTIES IN APPLYING UBET

In UBET, it is assumed that the rank of C_1 is equal to the number of nodes (q) and there are at most q gross errors. *i.e.*, no matter how many gross errors are identified in step (1), q and only q gross errors will be used in θ_1 in step (2). Thus, the following problems arise:

Too Many Gross Error Candidates

In many cases, the gross error candidates found in step (1) are more than q. Which q gross errors should one choose? We illustrate this situation next:

Consider the process flowsheet in Figure 1, where the true values of the mass flowrates are $x = [100.\ 40.\ 80.\ 30.\ 50.\ 40.\ 60.\ 40.\ 20.\ 60.]$ and the total flowrate standard deviations are taken as 2.5% of the true flow rates. All streams are measured and their values are: $y = [115.63\ 31.40\ 77.11\ 30.42\ 49.50\ 40.69\ 61.22\ 46.71\ 20.64\ 61.00]$. Now, assume three measurement biases are present in streams S1, S2 and S8 of size 12.5, −7 and 6 respectively.

The Modified Method of Pseudonodes (MMP; Serth and Heenan, 1986) identifies only biases. Thus, to identify leaks, an extension of MMP was necessary. This extension consists of adding the corresponding leak into the

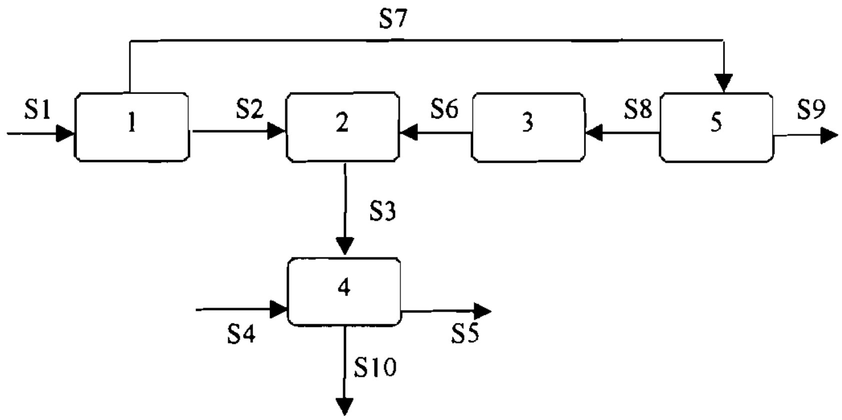

FIGURE 1 Process flowsheet 1.

candidate list when a node is identified as a suspect during the procedure. Rollins (1998) suggested that an alternative would be to use the LCT method (Rollins *et al.*, 1996).

Once the extended MMP was applied, the following list of candidates was obtained: {S1, S2, S8, Leak1, Leak 2, Leak3, Leak5}. It can be seen that the number of candidates is larger than the number of nodes in this system, so the UBET strategy can only be applied by an expert user who can decide which leaks and biases have to be included in C_1.

Not Enough Gross Error Candidates

In some other cases, the number of suspects identified by a strategy could be less than the number of nodes. If that happens, could one arbitrarily add any other candidates to θ_1?

In the example of Figure 1, consider the following measurement vector y = [102.97 38.79 79.96 29.88 47.99 46.25 58.42 41.41 19.59 60.79], which corresponds to a bias of size 6 in stream S6. The extended MMP strategy identifies the following candidate list: {S6, Leak2, Leak3}, which contains two gross errors less than required. Thus, some other gross error has to be included in the candidate list to fulfill the requirement of $q = 5$ gross errors.

Singularity of C_1

In the example of Figure 1, consider the following measurement vector y = [100.54 44.57 65.65 29.95 48.73 40.61 60.76 41.69 20.29 59.03], which corresponds to two biases of size 5.5 and −14.0 in streams 2 and 3. The extended MMP strategy identifies the following candidate list {S2, S3, Leak1, Leak2, Leak4}. The rank of C_1 based on this set is less than 5. In other words, C_1 is singular.

In addition, when the number of candidates is less than q one cannot add arbitrarily any gross errors to complete the required number without the risk of running into singularities. Likewise, when the number of candidates is larger than q, one may run into the same problem: one cannot remove the excess number arbitrarily. These suggestions have not been put forth by Rollins and Davis (1992) or papers thereafter, but rather, they are a result of the present discussion.

REVIEW OF CGLR

This strategy is a modification of the Serial Compensation Strategy based on GLR (Narasimhan and Mah, 1987), proposed by Keller *et al.* (1994). We call this new strategy CGLR.

It is based on a recursive algorithm. In each step, one new gross error is identified and CGLR collectively estimates and compensates for all gross errors identified in the previous steps rather than just estimates and compensates for the new gross error.

Suppose that k gross errors $\hat{B}_k = [b_1, b_2, \ldots, b_k]^T$ (obviously $k = 0$ at the initial time) are already detected. Assume also that there exist some other gross errors not yet identified and one of these gross errors is b_j with fault direction $f_j = A_j$ or $f_j = m_j$. The constraint residual of the system can be expressed as:

$$r = A\varepsilon + F_k B_k + f_j b_j \tag{7}$$

where $F_k = [f_1, \ldots, f_k]^T$ is the $(q \times k)$ fault directions matrix describing the gross errors already detected and ε the vector of measurement random errors. In the least square sense, the magnitude estimators of gross errors are solutions of:

$$\mathrm{Min}\left\| r - F_k\hat{B}_k - f_j\hat{b}_j \right\|^2_{V^{-1}} \tag{8}$$

where $\hat{B}_k$ is a $(k \times 1)$ vector of gross errors estimation, $\hat{b}_j$ the estimation of the hypothetical gross error and $V = \mathrm{var}(r) = \mathrm{AQA}^T$. One finally obtains:

$$\hat{b}_j = \frac{f_j^T S_k r}{f_j^T S_k f_j} \tag{9}$$

$$\hat{B}_k = E_k^{-1} F_k^T V^{-1} (r - f_j \hat{b}_j) \tag{10}$$

where

$$E_k = F_k^T V^{-1} F_k \tag{11}$$

$$S_k = V^{-1} - V^{-1} F_k E_k^{-1} F_k^T V^{-1} \tag{12}$$

The GLR test statistic is:

$$T_k = \sup T_k^j \tag{13}$$

where

$$T_k^j = \frac{(f_j^T S_k r)^2}{f_j^T S_k f_j} \tag{14}$$

According to Keller *et al.* (1994), the recursive procedure for applying CGLR can be summarized as follows:

(1) At time k ($k = 0$ at the initial time), calculate T_k^j for every vector f_j of $Z(s-k)$. Here s represents the total number of streams and nodes and $Z(s-k)$ the set of $(s-k)$ hypothetical fault directions.
(2) Obtain T_k and compare it with a pre-specified critical value. If T_k exceeds this threshold, the gross error j is present. Also obtain $\hat{b}_j$ and $\hat{B}_k$. If T_k does not exceed this threshold, go to step (4).
(3) Let $k = k+1$ and $F_{k+1} = [F_k | f_j]$, go to step (1).
(4) Declare k gross errors and stop.

SOME DIFFICULTIES IN APPLYING CGLR

At step k, the calculation of $(f_j^T S_k f_j)$ has to be performed for every vector f_j of $Z(s-k)$ in order to evaluate T_k^j and b_j. If $(f_j^T S_k f_j)$ is zero, it is not possible to calculate the test statistic for stage k (T_k), and to make a decision of which is the gross errors in suspect.

Consider the process flowsheet in Figure 2, where the true values of the mass flowrates are $x = [20.\ 40.\ 40.\ 20.\ 20.]$ and the total flowrate standard deviations are taken as 2.5% of the true flow rates. All streams are measured and their values are: $y = [31.9\ 40.9\ 34.5\ 20.7\ 19.4]$. Now, assume two gross errors are present in streams S1 and S3 of size 11.1 and -5.0 respectively.

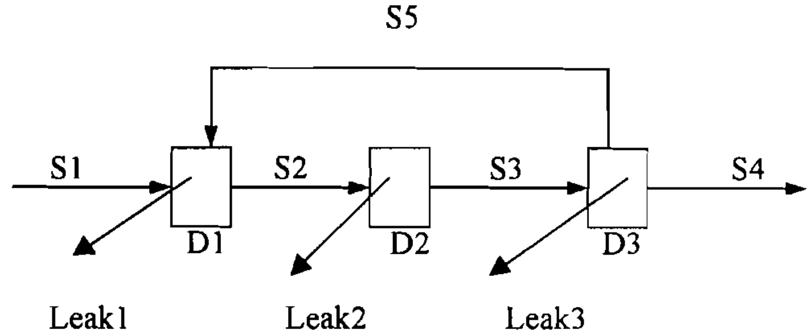

FIGURE 2 Process flowsheet 2.

The application of the CGLR strategy results in the following: At step (1), $k = 0, s = 8$, and one can calculate T_0^j for all 8 hypothetical fault directions f_j ($j = 1, \ldots, 8$). At step (2), T_0 is identified equal to T_0^1 and exceeds the critical value, therefore, the bias in S1 is detected and its size $\hat{b}_1$ is also estimated. Since T_0 exceeds the critical value, one has to go back to step (1). Now we have $k = 1$ and $s - k = 7$. Unfortunately, $f_6^T S_1 f_6$, corresponding to Leak1, is found to be zero, which makes the calculation of T_1 and the estimation of $\hat{b}_6$ impossible.

PREVIOUS REPORTS OF FAILURES

The difficulty associated with the selection of a right amount of candidates list has not been reported in the literature for the UBET strategy. If a non-automatic procedure of selection was originally applied, the expert user might choose a convenient amount of gross errors from MMP results, so this problem did not appear. In the same way, some combinations with a right amount of gross errors that create singularities problems were avoided.

Nevertheless, some difficulties can not be eliminated by an expert user. These problems were reported by Rollins and Davis (1992). When analyzing the failures of the technique in nine cases of Table V in their paper, they conclude: "*This happened because these δ's are distributed such that their unique estimation is not possible. That is, A_{11} is not of rank* 3." Later on, they add: "*It is important to note, however, that in all these cases BNS reaches the conclusion that at least two or three δ's are nonzero, which is correct.*" The authors use an ad-hoc technique, not explained in their paper, but communicated separately (Rollins, 1998) to determine that "two of three errors" are present.

In turn the problems experienced by CGLR were not reported in the original paper. All these difficulties, can be summarized as:

- Not obtaining the right amount of gross error candidates.
- Singularities.
- Uncertainties as of which are the gross error present.

ASSESSMENT OF THE FREQUENCY OF FAILURES

To determine the frequency at which these methods run into the aforementioned problems a simulation study was conducted. The method proposed by Iordache *et al.* (1985) was followed. Each result is based on 10000 simulation trials where the random errors are changed and the magnitudes of gross errors are fixed.

Three performance measures are used: overall power (OP), average number of type I errors (AVTI) and expected fraction of perfect identification (OPF). They are defined as follows:

$$\mathrm{OP} = \frac{\text{No. of gross errors correctly identified}}{\text{No. of gross errors simulated}} \tag{15}$$

$$\mathrm{AVTI} = \frac{\text{No. of gross errors incorrectly identified}}{\text{No. of simulation trials}} \tag{16}$$

$$\mathrm{OPF} = \frac{\text{No. of trials with perfect identification}}{\text{No. of simulation trials}} \tag{17}$$

The first two measures were proposed by Mah and Narasimhan (1987) and the last one by Rollins and Davis (1992).

The process flowsheet in Figure 3 is used to analyze the performance of the strategies. In this example the true flow rate values are $x = [5.\ 15.\ 15.\ 5.\ 10.\ 5.\ 5.]$. The flow rate standard deviations were taken as 2.5% of the true flow rates. All possible combinations of two measurement biases were simulated first. Fixed gross error magnitudes of 7 and 4 standard deviations were considered for the corresponding flow rates. Measurement values for each simulation trial were taken as the average of ten random generated values. In order to compare results on the same basis, the level of significance of each method was chosen in this paper such that it gives an AVTI equal to 0.1 under the null hypothesis. The results of these runs are summarized in Tables I and II.

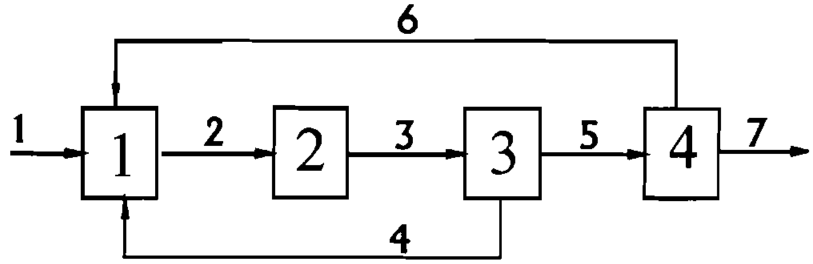

FIGURE 3 Process flowsheet 3.

TABLE I Performance of UBET. Biases

Biased stream	*AVTI*	*OP*	*OPF*	*Failure* (1.-*OPF*)			
				Singular	*Too few*	*Too many*	*Other*
1-2	0.000	0.920	0.920	0.0000	0.0237	0.0565	0.0000
1-3	0.000	0.000	0.000	0.0035	0.9500	0.0464	0.0000
1-4	0.010	0.837	0.813	0.0076	0.0808	0.0510	0.0477
1-5	0.000	0.000	0.000	0.0001	0.9924	0.0075	0.0000
2-5	0.000	0.000	0.000	0.0000	0.9649	0.0351	0.0000
2-6	0.000	0.000	0.000	0.0001	0.9752	0.0247	0.0000
2-7	0.000	0.000	0.000	0.0011	0.9725	0.0263	0.0001
3-5	0.004	0.047	0.042	0.0000	0.9192	0.0341	0.0044
3-6	0.004	0.003	0.000	0.0076	0.9629	0.0244	0.0051
3-7	0.020	0.031	0.012	0.0034	0.9653	0.0000	0.0196
4-7	0.000	0.000	0.000	0.0000	0.9894	0.0105	0.0001
5-7	0.021	0.023	0.002	0.0000	0.9572	0.0196	0.0208
1-6	0.000	0.000	0.000	0.9234	0.0311	0.0452	0.0003
1-7	0.000	0.000	0.000	0.6320	0.3586	0.0091	0.0003
2-3	0.000	0.000	0.000	0.0008	0.9121	0.0799	0.0000
2-4	0.003	0.027	0.000	0.0051	0.8790	0.0627	0.0532
3-4	0.001	0.000	0.000	0.0047	0.9323	0.0624	0.0005
4-5	0.000	0.000	0.000	0.0000	0.9162	0.0838	0.0000
4-6	0.000	0.000	0.000	0.0000	0.9569	0.0430	0.0001
5-6	0.003	0.002	0.000	0.0000	0.9419	0.0551	0.0030
6-7	0.000	0.000	0.000	0.8374	0.1268	0.0358	0.0000

From these tables, one could conclude that the power of these methods is low. In addition, the methods cannot be implemented automatically, as an expert opinion is needed to sort out the problems that arise. The identification of too few or too many gross error candidates can be blamed on the identification method of choice. Nothing suggests that other identification algorithms will perform satisfactorily in this regard either. In addition, it is interesting to note how certain combinations of gross errors lead to a high number of failures due to singularities, whereas others do not. The

TABLE II Performance of CGLR. Biases

Biased stream	AVTI	OP	OPF	Failure (1.-OPF)	
				Singular	Other
1–2	0.071	0.912	0.897	0.0321	0.0710
1–3	0.070	0.914	0.900	0.0298	0.0703
1–4	0.032	0.944	0.927	0.0392	0.0340
1–5	0.973	0.027	0.023	0.0044	0.9730
2–5	0.036	0.976	0.939	0.0244	0.0363
2–6	0.068	0.967	0.916	0.0156	0.0686
2–7	0.000	0.999	0.999	0.0008	0.0002
3–5	0.038	0.976	0.938	0.0241	0.0378
3–6	0.022	0.949	0.938	0.0405	0.0214
3–7	0.000	0.999	0.999	0.0008	0.0002
4–7	0.001	0.642	0.284	0.0010	0.7150
5–7	0.007	0.996	0.993	0.0015	0.0053
1–6	0.004	0.002	0.000	0.9967	0.0033
1–7	0.004	0.002	0.000	0.9967	0.0033
2–3	0.000	0.000	0.000	1.0000	0.0000
2–4	0.025	0.041	0.000	0.9178	0.0822
3–4	0.023	0.040	0.000	0.9195	0.0805
4–5	0.000	0.000	0.000	1.0000	0.0000
4–6	0.001	0.001	0.000	0.9985	0.0015
5–6	0.038	0.020	0.000	0.9604	0.0396
6–7	0.000	0.073	0.000	0.8538	0.1462

emergence of singularities is more frequent in CGLR than in UBET. The difference is not by any means a result of a poorer performance of the method. These singularities are not found in UBET simply because too many or too few gross error candidates are identified first.

If one seeks an automatic implementation of these techniques one needs an automatic way of sorting these difficulties efficiently. All these difficulties can be sorted using the equivalency theory (Bagajewicz and Jiang, 1998), which is summarized briefly next.

EQUIVALENCY THEORY

This theory was recently presented by Bagajewicz and Jiang (1998) and is briefly summarized next. Two sets of gross errors are equivalent when they have the same effect in data reconciliation, that is, when simulating either one, leads to the same value of objective function. Therefore, the equivalent sets of gross errors are theoretically undistinguishable. In other words, when a set of gross errors is identified, there exists an equal possibility that the true locations of gross errors are in one of its equivalent sets. From the view of graph theory, equivalent sets exist when candidate stream/leaks form a loop in an augmented graph consisting of the original graph representing the flowsheet with the addition of environmental node.

For example, consider the process of Figure 4 and assume that all streams are measured. As shown in Table III, a bias of (−2) in S_4 and a bias of (+1) in S_5 (Case 1) can be represented by two alternative sets of two gross errors (Cases 2 and 3).

By applying this theory, one can see that any proposed set of gross error candidates cannot form a loop. Otherwise the size of these gross errors is indeterminate, a condition that leads to singularities. This explains why combinations of introduced gross errors in Figure 3 like 1−6, 1−7, 6−7 lead so easily to singularities: The addition of just one stream to these sets can lead to a loop through the environmental node. Take for example the combination of simulated gross errors 1−6: just the addition of stream 7 will form a set with a loop through the environmental node.

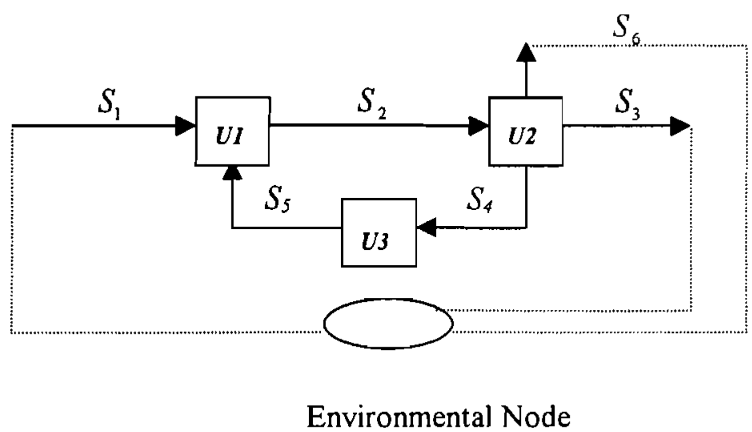

FIGURE 4 Illustration of gross error equivalency.

TABLE III Illustration of equivalent sets in $\{S_2, S_4, S_5\}$ of Figure 4

		S_1	S_2	S_3	S_4	S_5	S_6
Measurement		12	18	10	4	7	2
Case 1	Reconciled data	12	18	10	6	6	2
(Bias in S_4, S_5)	Estimated biases				−2	1	
Case 2	Reconciled data	12	19	10	7	7	2
(Bias in S_2, S_4)	Estimated biases		−1		−3		
Case 3	Reconciled data	12	16	10	4	4	2
(Bias in S_2, S_5)	Estimated biases		2			3	

TABLE IV Illustration of degenerate cases in $\{S_2, S_4, S_5\}$ of Figure 4

		S_1	S_2	S_3	S_4	S_5	S_6
Measurement		12	18	10	7	7	2
Case 1	Reconciled data	12	18	10	6	6	2
(Bias in S_4, S_5)	Estimated biases				1	1	
Case 2	Reconciled data	12	19	10	7	7	2
(Bias in S_2)	Estimated biases		−1				

Degenerate Cases

The equivalencies above are built in the assumption that the number of gross errors identified is equal to the real number of gross errors. However, there are examples where the actual number of gross errors can be larger than the number of gross errors identified. One such example is shown in Table IV, where a set of two gross errors (Case 1) is equivalent to one gross error (Case 2). These cases are rare, as they require that the two real gross errors have equal sizes. This poses an additional uncertainty. If a certain number of gross errors are detected, there is a possibility that the actual number of gross errors is larger.

Degenerate cases are, however, more frequent in the case of the presence of leaks. Indeed, take again the example of Figure 4. A leak of a certain size in unit $U1$ can be represented by a set of equal in size (but opposite sign) biases in streams S_2 and S_3. If one chooses to perform the candidate identification step using a bias detection test (like the measurement test for example), then one is likely to identify these degenerate cases.

MODIFICATIONS OF UBET AND CGLR TO ADDRESS SINGULARITIES

In view of the above theory, we can now explain why in applying UBET C_1 could be singular and how one should construct θ_1, and why in applying CGLR in some cases $f_j^T S_k f_j$ are zero. All these singularities are caused by the presence of loops.

The procedure of applying UBET is modified as follows:

(1) In step (I), after the candidate bias/leak list is constructed, the Equivalency Theory is applied to delete the possible existing loops in this list. All candidates are checked one by one. If a candidate forms a loop with any candidate(s) checked, it will be deleted from the list.
(2) After this modified step (1), in view of the Equivalency Theory, the elements of θ_1 in step (2) will certainly be no more than the number of

constraint equations (q). Therefore, θ_1 can be directly transferred from step (1).

(3) In step (3), if the number of elements in θ_1 is less than q, one has to add other streams/leaks as candidates to θ_1 to make the number equal to q. Any of these streams/leaks must be checked and be sure that it is not forming a loop with any element(s) in θ_1 before it is added.

The modification made to the CGLR procedure is:

- In step (1), at time k (when $k > 1$), one calculate T_k^j only for those vectors f_j that do not form a loop with any element(s) in F_k.

We call these new methods Modified UBET (MUBET) and Modified CGLR (MCGLR)

PERFORMANCE OF MODIFIED UBET AND CGLR IN ADDRESSING SINGULARITIES

After modifying UBET and CGLR strategies to avoid singularities, a simulation procedure was applied to study their performance in gross error detection and estimation.

Table V shows good results expressed in terms of performance measures for the two modified methods when leaks are not present. In addition, it can be seen that there is one case (biases in streams 1 and 5) where UBET is capable of performing an acceptable identification, whereas CGLR performs poorly. At this time, we have no explanation for this anomaly.

TABLE V Good performance results for MUBET and MCGLR when leaks are not present

Biased	*MUBET*			*MCGLR*		
stream	*AVTI*	*OP*	*OPF*	*AVTI*	*OP*	*OPF*
1–2	0.066	0.999	0.934	0.103	0.944	0.897
1–3	0.132	0.977	0.915	0.100	0.944	0.900
1–4	0.133	0.932	0.817	0.077	0.977	0.927
1–5	0.134	0.963	0.911	0.977	0.031	0.023
2–5	0.122	0.982	0.914	0.061	1.000	0.939
2–6	0.110	0.980	0.893	0.085	0.982	0.916
2–7	0.303	0.930	0.849	0.002	1.000	0.999
3–5	0.146	0.965	0.924	0.063	1.000	0.938
3–6	0.126	0.972	0.899	0.064	0.989	0.938
3–7	0.163	0.976	0.902	0.002	1.000	0.999
4–7	1.623	0.600	0.520	0.003	0.642	0.284
5–7	0.230	0.946	0.881	0.010	0.997	0.993

However, it serves as an indication that the power of a method seems to be location dependent.

As it is shown in Table VI, MUBET gives zero OPF for the nine combinations of gross errors. In addition, CGLR gives similar results for seven of these cases. Although UBET and CGLR have been modified to avoid singularities, the OPF measure indicates the inability of the methods to distinguish the exact locations of the gross errors, hence its inability to estimate its size.

In Table VII, bias estimation results are presented based on perfect identification trials. It is important to note that when the OPF is low, the results are based on a lower number of successful runs. Similar identification and estimation examples are presented in Tables VIII and IX when leaks are present.

TABLE VI Unsatisfactory results for MUBET and MCGLR when leaks are not present

Biased	*MUBET*			*MCGLR*		
stream	*AVTI*	*OP*	*OPF*	*AVTI*	*OP*	*OPF*
1−6	0.084	0.966	1, 6, 7*	0.068	0.997	0.935
1−7	1.061	0.500	1, 6, 7*	1.062	0.500	0.000
2−3	1.077	0.500	2, 3, 4*	1.058	0.503	0.005
2−4	0.091	0.944	2, 3, 4*	0.992	0.500	0.000
3−4	1.017	0.504	2, 3, 4*	0.992	0.500	0.000
4−5	1.083	0.500	4, 5, 6*	0.052	1.000	0.949
4−6	0.351	0.875	4, 5, 6*	0.899	0.576	0.144
5−6	1.025	0.517	4, 5, 6*	1.033	0.500	0.000
6−7	1.050	0.499	1, 6, 7*	0.910	0.500	0.000

* At least two of the three locations contain gross errors. (Rollins and Davis, 1992).

TABLE VII Biases estimation results for MUBET and MCGLR when leaks are not present

Biased		*MUBET*		*MCGLR*	
stream	*Sizes*	*Estimates*	*Standard deviation*	*Estimates*	*Standard deviation*
1−2	0.875, 1.500	0.875, 1.499	0.056, 0.144	0.876, 1.497	0.058, 0.134
1−3	0.875, 1.500	0.876, 1.499	0.056, 0.143	0.876, 1.497	0.056, 0.135
1−4	0.875, 0.500	0.874, 0.515	0.056, 0.118	0.875, 0.505	0.056, 0.123
2−5	2.625, 1.000	2.626, 1.000	0.135, 0.095	2.626, 1.000	0.133, 0.094
2−6	2.625, 0.500	2.626, 0.501	0.144, 0.092	2.626, 0.502	0.133, 0.093
2−7	2.625, 0.500	2.623, 0.501	0.153, 0.110	2.626, 0.500	0.129, 0.053
3−5	2.625, 1.000	2.624, 1.000	0.142, 0.105	2.624, 1.000	0.134, 0.092
3−6	2.625, 0.500	2.625, 0.501	0.146, 0.092	2.624, 0.500	0.145, 0.091
3−7	2.625, 0.500	2.626, 0.500	0.145, 0.056	2.625, 0.500	0.130, 0.053
4−7	0.875, 0.500	0.925, 0.478	0.121, 0.081	0.969, 0.451	0.077, 0.040
5−7	1.750, 0.500	1.754, 0.501	0.134, 0.056	1.750, 0.500	0.091, 0.054

TABLE VIII Performance results for MUBET and MCGLR when leaks are present

Gross errors	*MUBET*			*MCGLR*		
	AVTI	*OP*	*OPF*	*AVTI*	*OP*	*OPF*
L2, B4	0.168	0.961	0.900	0.052	0.994	0.953
L2, B5	0.213	0.952	0.891	0.689	0.837	0.644
L2, B6	0.154	0.969	0.917	0.052	0.996	0.949
L2, B7	0.162	0.951	0.928	0.051	0.990	0.953
L3, B2	0.327	0.921	0.831	0.095	0.981	0.944
L3, B6	0.202	0.963	0.892	0.237	0.937	0.902

Note: Ln means a leak in node *n* and Bn means a bias in stream *n*.

TABLE IX Gross error estimation for MUBET and MCGLR when leaks are present

Gross errors	*Sizes*	*MUBET*		*MCGLR*	
		Estimates	*Standard deviation*	*Estimates*	*Standard deviation*
L2, B4	1.8, 0.625	1.801, 0.628	0.147, 0.127	1.801, 0.626	0.141, 0.123
L2, B5	1.8, 1.25	1.807, 1.246	0.151, 0.093	1.809, 1.209	0.104, 0.076
L2, B6	1.8, 0.625	1.801, 0.625	0.129, 0.091	1.800, 0.625	0.109, 0.093
L2, B7	1.8, 0.625	1.801, 0.625	0.136, 0.093	1.803, 0.625	0.149, 0.093
L3, B2	.25, 1.875	1.240, 1.902	0.119, 0.141	1.253, 1.872	0.106, 0.155
L3, B6	.25, 0.625	1.251, 0.625	0.130, 0.091	1.257, 0.623	0.088, 0.092

ADDRESSING UNCERTAINTIES

The uncertainties in gross error identification, experienced by UBET as well as CGLR, can be also completely understood in view of the aforementioned theory of equivalency of gross errors. Returning to the uncertain cases of Table VI, each set of three streams ($\{1,6,7\},\{2,3,4\}$ and $\{4,5,6\}$) forms a loop, the first through the environment node, and the other two within the graph. For each set, all combinations of two gross errors form equivalent sets, so they are theoretically undistinguishable.

As a result of the equivalency of gross errors, the issue of perfect identification of gross errors should be revised. This issue has been raised by Bagajewicz and Jiang (1998), when presenting the equivalency theory, where the concepts of gross error exact detectability and gross error correct detectability were introduced. Exact detectability is possible in principle, only when there are no equivalent sets. Thus, the perfect identification performance measure (OPF), as defined by Rollins and Davis should be extended to include equivalent sets. That is, gross error identification should be considered successful if exactly the errors introduced are identified, or an equivalent set is identified. Degenerate cases should also be considered as success in such counting.

Using the concepts from Equivalency Theory the definition of OPF is extended to include all trials that identify the perfect set, equivalent sets and degenerate sets as successful trials. This new measure is called overall power of equivalent identification (OPFE)

$$\text{OPFE} = \frac{\text{No. of trials with successful identification}}{\text{No. of simulation trials}} \tag{18}$$

Using this new measure, the modified UBET and CGLR strategies achieve a significant increase in the identification performance for the cases of Tables V, VI and VIII. These results are included in Tables X and XI. Furthermore, the size estimation of all simulated gross errors can now be obtained, as it is shown in Tables XII and XIII. These results are based on successful identification trials.

TABLE X OPFE for MUBET and MCGLR strategies. Biases

Biased stream	*MUBET*	*MCGLR*
1–2	0.935	0.897
1–3	0.963	0.900
1–4	0.817	0.927
1–5	0.956	0.023
2–5	0.950	0.940
2–6	0.909	0.916
2–7	0.942	0.999
3–5	0.960	0.939
3–6	0.924	0.938
3–7	0.922	0.999
4–7	0.971	0.284
5–7	0.914	0.994
1–6	0.928	0.935
1–7	0.911	0.944
2–3	1.000	1.000
2–4	0.839	0.911
3–4	0.943	0.911
4–5	0.996	0.949
4–6	0.928	0.968
5–6	0.946	0.949
6–7	0.958	0.802

TABLE XI OPFE for MUBET and MCGLR strategies. Leaks and biases

Biased stream	*MUBET*	*MCGLR*
L2, B4	0.953	0.954
L2, B5	0.986	0.971
L2, B6	0.961	0.949
L2, B7	0.989	0.954
L3, B2	0.981	0.964
L3, B6	0.944	0.951

TABLE XII Biases estimation results for MUBET and MCGLR when leaks are not present

Biased stream	Sizes	MUBET		MCGLR	
		Estimates	Standard deviation	Estimates	Standard deviation
1–2	0.875, 1.500	0.875, 1.499	0.056, 0.144	0.876, 1.497	0.058, 0.134
1–3	0.875, 1.500	0.877, 1.501	0.093, 0.152	0.876, 1.497	0.056, 0.135
1–4	0.875, 0.500	0.874, 0.515	0.056, 0.118	0.875, 0.505	0.056, 0.123
2–5	2.625, 1.000	2.626, 1.001	0.135, 0.109	2.626, 1.000	0.133, 0.094
2–6	2.625, 0.500	2.625, 0.507	0.135, 0.103	2.626, 0.502	0.133, 0.093
2–7	2.625, 0.500	2.622, 0.496	0.152, 0.110	2.626, 0.500	0.129, 0.053
3–5	2.625, 1.000	2.625, 1.000	0.150, 0.118	2.624, 1.000	0.134, 0.093
3–6	2.625, 0.500	2.629, 0.505	0.150, 0.099	2.624, 0.500	0.145, 0.091
3–7	2.625, 0.500	2.630, 0.501	0.148, 0.057	2.625, 0.499	0.130, 0.053
4–7	0.875, 0.500	0.854, 0.502	0.141, 0.091	0.969, 0.451	0.077, 0.040
5–7	1.750, 0.500	1.751, 0.498	0.138, 0.070	1.750, 0.500	0.091, 0.054
1–5	0.875, 1.000	0.877, 0.999	0.078, 0.105	0.886, 0.819	0.057, 0.060
1–6	0.875, 0.500	0.873, 0.510	0.057, 0.115	0.875, 0.500	0.057, 0.092
1–7	0.875, 0.500	0.896, 0.524	0.109, 0.108	0.875, 0.501	0.094, 0.093
2–3	2.625, 1.500	2.627, 1.501	0.138, 0.139	2.626, 1.500	0.134, 0.135
2–4	2.625, 0.500	2.644, 0.522	0.160, 0.122	2.643, 0.520	0.160, 0.120
3–4	2.625, 0.500	2.636, 0.513	0.162, 0.125	2.640, 0.519	0.160, 0.118
4–5	0.875, 1.000	0.875, 1.000	0.134, 0.121	0.864, 1.001	0.089, 0.119
4–6	0.875, 0.500	0.882, 0.498	0.143, 0.095	0.863, 0.512	0.153, 0.117
5–6	1.750, 0.500	1.759, 0.511	0.140, 0.125	1.754, 0.505	0.138, 0.124
6–7	0.875, 0.500	0.868, 0.500	0.122, 0.057	0.893, 0.487	0.084, 0.051

TABLE XIII Gross error estimation for MUBET and MCGLR when leaks are present

Gross errors	Sizes	MUBET		MCGLR	
		Estimates	Standard deviation	Estimates	Standard deviation
L2, B4	1.8, 0.625	1.798, 0.627	0.166, 0.132	1.801, 0.627	0.144, 0.125
L2, B5	1.8, 1.25	1.799, 1.251	0.168, 0.114	1.620, 1.378	0.290, 0.262
L2, B6	1.8, 0.625	1.799, 0.627	0.143, 0.103	1.799, 0.625	0.112, 0.097
L2, B7	1.8, 0.625	1.805, 0.625	0.147, 0.096	1.802, 0.626	0.152, 0.096
L3, B2	1.25, 1.875	1.251, 1.881	0.139, 0.152	1.265, 1.854	0.139, 0.202
L3, B6	1.25, 1.875	1.251, 0.626	0.136, 0.096	1.234, 0.645	0.134, 0.130

CONCLUSIONS

In this paper, the identification difficulties that arise in the practical application of two effective strategies for gross error identification and estimation are presented. The Equivalency Theory of Gross Errors is used to explain why these techniques sometimes fail. The theory states that the problem is related to the nature of the system and that they do not depend on any presumed weaknesses of the strategies. Based on the Equivalency Theory of Gross Errors, UBET and CGLR strategies are modified. These modifications enhance their performance in identification and estimation of gross errors and allow their use in an automatic way.

NOTATION

A	$(q \times p)$ balance matrix
A_j	jth column of matrix A
AVTI	Average number of type I errors
$\hat{B}_k$	vector of gross error estimates at step k
b_j	size of gross error not yet identified
C_1	matrix defined by Eq. (4)
e_i	vector with 1 in the ith place and zero elsewhere
E	matrix defined by Eq. (11)
F_k	$(q \times k)$ fault direction matrix
f_j	jth fault direction
k	estimation step for CGLR
l	vector used for making linear combinations of measurements
m_j	vector with zero in every position but in jth
M	diagonal matrix of constants for calculating leaks
OP	overall power
OPF	expected fraction of perfect identification
OPFE	overall power of equivalent identification
p	number of measured variables
q	number of constraints
Q	covariance matrix of the measurement errors
r	equations' residuals
S	matrix defined by Eq. (12)
s	number of constraints and measurements
T_k	GLR test statistic
V	covariance matrix of r
x	vector of true values of variables
y	measurements

Greek Symbols

δ	$(p \times 1)$ measurement biases
γ	$(q \times 1)$ leaks
ε	vector of random measurement errors
μ_r	expected value of r
θ_1	vector with elements of δ and γ

References

Bagajewicz, M. and Jiang, Q. (1998) Gross Error Modeling and Detection in Plant Linear Dynamic Reconciliation. *Comp. and Chem. Engng.*, **22**(12), 1789–1810.

Iordache, C., Mah, R. and Tamhane, A. (1985) Performance Studies of the Measurement Test for Detection of Gross Errors in Process Data. *AIChE J.*, **31**, 1187–1201.

Keller, J., Dorouach, M. and Karzala, G. (1994) Fault Detection of Multiple Biases or Process Leaks in Linear Steady State Systems. *Comp. and Chem. Engng.*, **18**(10), 1001–1004.

Kim, I., Kang, M., Park, S. and Edgar, T. (1997) Robust Data Reconciliation and Gross Error Detection: The Modified MIMT using NLP. *Comp. and Chem. Engng.*, **21**(7), 775–782.

Mah, R. S. H., Stanley, G. and Downing, D. (1976) Reconciliation and Rectification of Process Flow and Inventory Data. *Ind. Eng. Chem. Process Des. Dev.*, **15**, 175–183.

Narasimhan, S. and Mah, R. (1987) Generalized Likelihood Ratio Method for Gross Error Identification. *AIChE J.*, **33**, 1514–1521.

Ripps, D. L. (1965) Adjustment of Experimental Data. *Chem. Eng. Prog. Symp. Ser.*, **61**(55), 8–13.

Rollins, D. and Davis, J. (1992) Unbiased Estimation of Gross Errors in Process Measurements. *AIChE J.*, **38**, 563–572.

Rollins, D. (1998) Private Communication.

Rollins, D. K., Cheng, Y. and Devanathan, S. (1996) Intelligent Selection of Hypothesis Tests to Enhance Gross Error Identification. *Comp. and Chem. Engng.*, **20**(5), 517.

Rosenberg, J., Mah, R. S. H. and Iordache, C. (1987) Evaluation of Schemes for Detecting and Identifying Gross Errors in Process Data. *Ind. Eng. Chem. Res.*, **26**, 555–564.

Sánchez, M. and Romagnoli, J. (1994) Monitoreo de Procesos Continuos: Análisis Comparativo de Técnicas de Identificación y Cálculo de Bias en los Sensores. *AADECA 94–XIV Simposio Nacional de Control Automático*, Argentina.

Serth, R. and Heenan, W. (1986) Gross Error Detection and Data Reconciliation in Steam Metering Systems. *AIChE J.*, **32**, 733–742.