On a Strategy of Serial Identification with Collective Compensation for Multiple Gross Error Estimation in Linear Steady-State Reconciliation

Qiyou Jiang and Miguel J. Bagajewicz*

School of Chemical Engineering and Material Science, University of Oklahoma, 100 E. Boyd Street, T-335, Norman, Oklahoma 73019-0628

This article presents a serial identification with collective size estimation strategy for multiple gross errors in linear steady-state data reconciliation. With use of the theory of gross error equivalency (Bagajewicz, M.; Jiang, Q. *Comput. Chem. Eng.* **1998**, in press), an effective strategy for leak detection is also presented. The issue of exact identification of gross errors is discussed, and the conversion of gross error sizes between any two equivalent sets is formulated. Finally, the proposed strategy is tested and compared with existing approaches.

Introduction

One of the most researched issues in the field of data reconciliation in process plants is the detection of gross errors. Not only it is desirable to have the proper location of gross errors, but it is also advantageous to have an estimate of their size. The former helps instrument maintenance in biases and process monitoring in leaks. The latter constitute valuable information for production accounting.

Statistical tests to perform identification include the Global Test (Reilly and Carpani, 1963), the Measurement Test (Mah and Tamhane, 1982; Crowe et al., 1983), the Nodal Test (Reilly and Carpani, 1963; Mah et al., 1976), Generalized Likelihood Ratios (GLR) (Narasimhan and Mah, 1987), Bonferroni tests (Rollins and Davis, 1992), and Principal Component Tests (PCT) (Tong and Crowe, 1995), among others.

Multiple gross errors can be identified by using a serial elimination strategy in combination with some of these tests (Serth and Heenan, 1986; Rosenberg et al.,1987). To obtain an estimate of the size of the gross errors, a serial compensation strategy is also available using error size estimation formulas developed by Madron (1985) and Narasimhan and Mah (1987). In this strategy, gross errors are estimated and measurements are compensated one by one, rather than eliminated. Recently an alternative idea for collective compensation (or estimation) was proposed (Rollins and Davis, 1992; Keller et al., 1994; Kim et al., 1997; Sanchez, 1996; Bagajewicz and Jiang, 1997, 1998; Sanchez et al., 1998).

Many researchers have evaluated the performance of these approaches. Serial elimination is simple but has the drawback of losing redundancy and is not applicable to gross errors that are not directly associated with measurements, for instance leaks (Mah, 1990). Serial compensation is applicable to all types of gross errors and can keep redundancy during the procedure, but its results depend completely on the accuracy of estimation for the size of gross errors (Rollins and Davis, 1992). Collective compensation is considered to be more correct but computationally too intensive and impractical (Keller, 1994). Nevertheless, results from simultaneous/collective compensation methods seem to be more accurate. For example, the simultaneous estimation method developed by Sanchez (1996) and later modified by Sanchez et al. (1998) is very accurate. However, it is still not suitable for large systems, because it becomes combinatorially expensive. Thus, an accurate and computationally inexpensive collective method is needed. This is the object of this article.

A serial identification and collective estimation of gross errors in linear steady-state data reconciliation is presented in this paper. Instead of the combinatorial approach as the one undertaken by Sanchez et al. (1998), this method uses the measurement test to identify a list of suspect gross error candidates at each stage. Then the location of the most probable gross error is confirmed. As gross errors are added to a list of confirmed gross errors, they are estimated again at each stage until no gross error is detected.

This article is organized as follows: The model for collective estimation of biases is presented first. Because singularities arise from certain choices of gross errors, the theory of gross error equivalency is presented next, thus summarizing the reasons for singularities. The strategy for serial identification with collective compensation is presented next, with comparisons with other techniques. In the next section, it is explained how leaks can be detected through the use of the measurement test and equivalent biases. The issue of exact identification of gross errors, raised by Bagajewicz and Jiang (1998), is revisited and the conversion of gross error sizes between two equivalent sets is formulated. Finally, the proposed strategy is tested and compared with existing approaches.

A Model for Collective Estimation of Multiple Biases in Linear Steady-State Reconciliation

A linear steady-state problem can be formulated as

$$\left.\begin{array}{l}\min (x-x^{+})Q^{-1}(x-x^{+})\\ \text{s.t.}\\ Ax=0\end{array}\right\} \qquad (1)$$

where x is a vector of reconciled data and x^{+} is a vector

* To whom correspondence should be addressed. E-mail: bagajewicz@ou.edu.

10.1021/ie980431e CCC: $18.00
Published on Web 04/08/1999

of measurements. Q and A are the covariance matrix and incidence matrix, respectively. In the presence of unmeasured variables several methods exist to obtain a matrix of only redundant measurement: matrix projection (Crowe, 1983), QR decomposition (Swartz, 1989; Sanchez and Romagnoli, 1996), co-optation (Madron, 1992), and node aggregation (Sanchez et al., 1998).

Assume a set of instrument biases is present. We have:

$$x = \hat{x} + \delta \tag{2}$$

where δ is a vector that contains biases for bias candidates. We are interested in considering only a limited number of biased streams. Therefore, the rest of the elements of δ are zero. This is accomplished by writing δ as follows:

$$\delta = L\hat{\delta} \tag{3}$$

where

$$\hat{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{nb} \end{bmatrix} \qquad L = [e_1 e_2 \cdots e_{nb}] \tag{4}$$

and e_i is a vector with unity in a position corresponding to a bias candidate and zero elsewhere.

Therefore, if we further assume all variables are measured, the reconciliation and gross error estimation model is:

$$\left.\begin{array}{l} \min(\hat{x} + L\hat{\delta} - x^+)^T Q^{-1}(\hat{x} + L\hat{\delta} - x^+) \\ \text{s.t.} \\ A\hat{x} = 0 \end{array}\right\} \tag{5}$$

The solution is given by:

$$\hat{x}^* = \tilde{x} + h\hat{\delta}^* \tag{6}$$

$$\hat{\delta}^* = [(h + L)^T Q^{-1}(h + L)]^{-1}(h + L)^T Q^{-1}(x^+ - \tilde{x}) \tag{7}$$

where $\tilde{x}$ is the solution for the reconciliation problem assuming no biases:

$$\tilde{x} = (I - QA^T(AQA^T)^{-1}A)x^+ \tag{8}$$

and

$$h = [QA^T(AQA^T)^{-1}A - I]L \tag{9}$$

This model has been presented by Sanchez (1996) and Sanchez et al. (1998), and in its dynamic form, by Bagajewicz and Jiang (1998). In several cases $\{(h + L)TQ^{-1}(h + L)\}$ is singular. These singularities are originated in the fact that gross errors belonging to loops are included in the candidate set (Bagajewicz and Jiang, 1998). In many other cases, different sets of gross errors have the same objective function value. This is also explained by the theory of equivalent gross errors (Bagajewicz and Jiang, 1998). It is therefore desirable to eliminate all these candidate sets from consideration.

Equivalency of Gross Errors

In a recent article, Bagajewicz and Jiang, (1998) presented a series of concepts regarding the equivalency

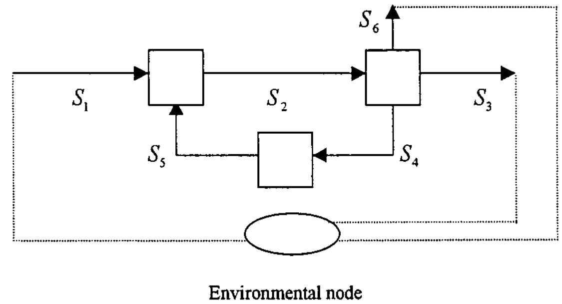

Figure 1. Illustration of gross error equivalency.

Table 1. Illustration of Equivalent Sets in $\{S_2, S_4, S_5\}$ of Figure 1

		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 2. Illustration of Degenerate Cases in $\{S_2, S_4, S_5\}$ of Figure 1

		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				

of sets of gross errors. 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 the objective function. Therefore, the equivalent sets of gross errors are theoretically undistinguishable. In other words, when a set of gross errors is identified, an equal possibility exists 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 streams/leaks form a loop in an augmented graph consisting of the original graph representing the flowsheet with the addition of environmental node. Every possible set of gross errors in a loop can always be equivalent to a set of gross errors with the number equal to the number of nodes in it minus one, which is defined as gross error cardinality in Bagajewicz and Jiang (1998).

As an illustration, consider the flowsheet of Figure 1 where all streams are measured. Loops can be identified as $\{S_3, S_6\}$, $\{S_2, S_4, S_5\}$,$\{S_1, S_2, S_3\}$, etc. In loop $\{S_2, S_4, S_5\}$, the gross error cardinality is 2, as shown in Table 1, a bias of (−2) in S_4 and a bias of (+1) in S_5 (case 1) can be represented equally by two alternative sets of two gross errors (cases 2 and 3). The similar equivalent sets in other loops can also be obtained.

Degeneracy. The number of gross errors in an equivalent set is usually equal to the gross error cardinality. However, there are examples in which a number of gross errors lower than the gross error cardinality can represent a set of gross errors, which are called degeneracy in Bagajewicz and Jiang (1998). One such example has been shown in Table 2, in which a set of two gross errors (case 1) is equivalent to one gross error (case 2). These cases are rare, because they require that the two real gross errors have equal sizes.

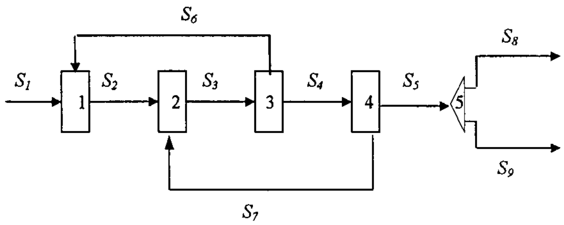

Figure 2. Illustration of gross error degeneracy.

Consider the process depicted in Figure 2. The true values for this system are $x = [1,2,3,2,1,1,1,0.4,0.6]$, and the standard deviations are 2% of each measurement.

One of the characteristics of collective compensation methods is that they will absorb part of the random errors. Thus, to make the illustrations of gross error detection clear, random errors have been eliminated and the measurements are equal to the true values plus the gross errors simulated. Although this does not resemble practical situations, as we shall see it helps to identify the phenomenon that will be described.

Consider the cases in Table 3. In case 1, the system consisting of S_1, S_4, S_5, S_6, S_9 has a cardinality = 4. The system cardinality is 4 in case 2 and 5 in case 3. The number of gross errors identified can be larger (case 1), smaller (case 2), or equal (case 3) to the number of gross errors introduced.

Quasi-Degeneracy. In practice many numbers are considered equal if they are close enough within a certain tolerance. Thus situations similar to degeneracy may happen. Table 4 provided some cases, which in this article are called quasi-degeneracy.

In light of what has been defined in the literature as identification performance, one could consider this as a failure. We will show how this is actually a result of part of the gross errors being absorbed as corrections to random errors, a well-known phenomenon. Let us consider case 1, first note that the set $\{S_2, S_3, S_6\}$ has gross error cardinality two, that is, any set of gross errors in this set is equivalent to no more than two gross errors. Therefore the real set of gross errors is equivalent to $\delta_2 = 0.030$ and $\delta_6 = -0.380$. Now note that the value for δ_2 in this new equivalent set is within the size of the standard deviation of the measurement. Therefore, as such it is absorbed as a random error.

Serial Identification with Collective Compensation Strategy

A serial identification with collective compensation (SICC) strategy for linear dynamic data reconciliation was presented recently (Bagajewicz and Jiang, 1998). The steady-state version of this strategy can be summarized as follows:

1. Run the data reconciliation and calculate the measurement tests (MT). If no MT flags, declare no gross error and stop. Otherwise go to step 2.

2. Construct a list of candidates (LC) by including all variables that failed the MT. If any two members in LC form a loop, erase one of them. Create a list of confirmed gross errors (LCGE). This list is empty at this stage.

3. Run the data reconciliation with gross error estimation model simulating a gross error in all the members of the LCGE and one member of the LC at a time.

4. Determine which member of the LC leads to the smallest value of the objective function. Add that variable to the LCGE.

5. Calculate MT for the run chosen in step 4. Erase all elements of the LC and place the latest flagged variables in LC. If there are any two members in LC forming a loop with any member(s) in LCGE, erase one of them. If LC is empty, go to step 6. Otherwise, go to step 3.

6. Determine all equivalent sets and corresponding gross error sizes. Declare all members in LCGE in suspect and stop.

Table 3. Illustration of Degenerate Cases of Figure 2

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
true value	1	2	3	2	1	1	1	0.4	0.6
case 1									
gross error introduced					−0.235				−0.235
gross error estimated	0.235			0.235		−0.235			
reconciled data	0.765	2.000	3.000	1.765	0.765	1.235	1.000	0.400	0.365
case 2									
gross error introduced	0.380	0.380			0.380				
gross error estimated						0.380	−0.380		
reconciled data	1.380	2.380	3.000	2.000	1.380	1.000	0.620	0.780	0.600
case 3									
gross error introduced	0.389		0.389		0.389				
gross error estimated		−0.389		−0.389				−0.389	
reconciled data	1.389	2.389	3.389	2.389	1.389	1.000	1.000	0.789	0.600

Table 4. Illustration of Quasi-Degenerate Cases of Figure 2

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
true value	1	2	3	2	1	1	1	0.4	0.6
case 1									
gross error introduced		0.410	0.380						
gross error estimated						−0.401			
reconciled data	1.000	2.401	3.400	1.998	1.000	1.401	0.998	0.400	0.600
case 2									
gross error introduced	0.350			0.350	0.400				
gross error estimated						0.371		−0.373	
reconciled data	1.373	2.001	2.997	2.368	1.373	0.629	0.996	0.773	0.600
case 3									
gross error introduced					0.235				0.24
gross error estimated	−0.238			−0.238		0.238			
reconciled data	1.238	2.000	3.000	2.238	1.238	0.762	1.000	0.400	0.839

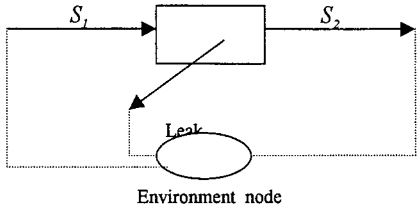

Figure 3. A simple process with a leak.

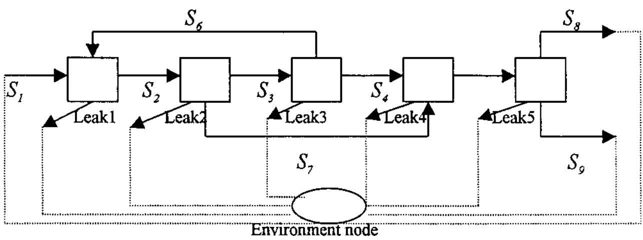

Figure 4. A process with several candidate leaks.

This strategy is based on the use of the measurement test to perform the identification of the gross error candidates. It does not contain any combinatorial search, and as such it can be applied to large systems. In addition, it has a performance that is similarly good.

Detection of Leaks

Leaks have been an elusive issue in gross error detection for years. A few strategies have been available for detecting leaks. They are graphic–theoretic analysis (Mah et al., 1976), generalized likelihood ratio method (GLR) (Narasimhan and Mah, 1987), unbiased estimation technique (UBET) (Rollins and Davis, 1992), and principal component analysis (PCA) (Tong and Crowe, 1995). Recently, the SEGE method based on the global test and collective compensation has been proposed to identify and estimate leaks and biases (Sanchez, 1996). This method was later modified by Sanchez et al. (1998) (MSEGE) making its automatic implementation possible with good performance results.

The SICC strategy can be used to identify leaks. To show this, the theory of gross error equivalency (Bagajewicz and Jiang, 1998) will be used.

If one considers a leak just as another stream, a leak forms at least one loop with some streams or other leaks in the augmented graph. Therefore it will be represented with at least 1 equivalent set of biases identified with the model.

Let us illustrate this issue with two examples. Consider the process in Figure 3. Assume that there is a leak with the size of +5 and the measurements for flowrates are 100 and 95, respectively. The SICC strategy identifies S_1 with a bias of 5. The equivalent sets are a leak with a size of 5 and a bias of −5 in S_2.

Consider now the process of Figure 4. The equivalencies are shown in Table 5.

Thus, any leak is equivalent to a set of equal size biases in a set of streams connecting to the unit where the leak occurs and the environment. This is an important result and leads to the following conclusion: *Because a leak is equivalent to a set of biased streams, any steady-state method that contains a test to detect biased instruments can be used in conjunction with the equivalency theory to asess the existence of leaks.*

Thus, one can use the SICC method to identify only biases. Leaks (if any) will be part of the equivalent set.

Table 5. Identification of Leaks

case no.	gross error introduced	gross error identified	equivalent sets
1	Leak1	S_1	{Leak1}
2	Leak2	same size biases in S_1, S_2	{Leak2}, {Leak1, S_2}
3	Leak5	S_8	{Leak5}, {S_9}
4	Leak1, Leak2	different size biases in S_1, S_2	{Leak1, S_2},{Leak1, Leak2} {S_1, Leak2}, {S_2, Leak2}
5	Leak1, S_3	S_1, S_3	{Leak1, S_3}

Identification of Equivalent Sets

Equivalent sets can be identified using the method presented by Bagajewicz and Jiang (1998). The method is based on obtaining a set that contains the set of n identified gross errors Ψ, such that the gross error cardinality that remains is equal to n, the number of elements (set cardinality) of Ψ. The set can be constructed by adding all streams that are in a loop with all subsets of Ψ. Thus, the list of equivalent basic sets $\{\varphi_i\}$ is constructed by identifying all subsets of n variables in Λ, such that $\{\varphi_i\}$ does not form a loop.

Consider the process in Figure 4. Assume that a set of gross errors $\Psi = \{S_1, S_2\}$ ($n = 2$) is identified. Because Leak1 is forming a loop with S_1 and Leak2 is in a loop with $\{S_1, S_2\}$, one can construct Λ by adding Leak1 and Leak2 to Ψ, i.e., $\Lambda = \{S_1, S_2, \text{Leak1}, \text{Leak2}\}$. Thus all equivalent sets of Ψ can be obtained as $\{S_1$, Leak2$\}$, $\{S_2$, Leak1$\}$, $\{S_2$, Leak2$\}$, and {Leak1, Leak2}. It should be noted that $\{S_1$, Leak1$\}$ is not an equivalent set because S_1 and Leak1 form a loop.

Conversion between Equivalent Sets

Once all the equivalent sets have been identified, one may want to determine the sizes of the gross errors in one equivalent set, when the sizes in another equivalent set are known. In addition, one wants to know what are the new values of the reconciled streams.

From eq 5, one can further obtain the combined model of data reconciliation and gross error estimation for both biases and leaks:

$$\left.\begin{array}{l} \min(\hat{x} + L\hat{\delta} - x^+)Q^{-1}(\hat{x} + L\hat{\delta} - x^+) \\ \text{s.t.} \\ A\hat{x} - K\mu = 0 \end{array}\right\} \tag{10}$$

where $K = [e_1 e_2 \cdots e_{nsl}]$ and μ is a vector of leaks.

Let

$$f = \hat{x} + L\hat{\delta} \tag{11}$$

Therefore one can rewrite eq 10 as follows:

$$\left.\begin{array}{l} \min(f - x^+)Q^{-1}(f - x^+) \\ \text{s.t.} \\ Af - AL\hat{\delta} - K\mu = 0 \end{array}\right\} \tag{12}$$

Assume that one already obtained a set of gross errors and its sizes $\hat{\delta}_1$ and μ_1 and wants to know the gross error sizes $\hat{\delta}_2$ and μ_2 of an equivalent set. The following holds between two equivalent sets:

$$f_1 = f_2 \tag{13}$$

Thus one gets:

$$AL_1\hat{\delta}_1 + K_1\mu_1 = AL_2\hat{\delta}_2 + K_2\mu_2 \quad (14)$$

or

$$[AL_1 \quad K_1]\begin{bmatrix}\hat{\delta}_1\\ \mu_1\end{bmatrix} = [AL_2 \quad K_2]\begin{bmatrix}\hat{\delta}_2\\ \mu_2\end{bmatrix} \quad (15)$$

Premultiplying both $[AL_1\ K_1]$ and $[AL_2\ K_2]$ by a certain particular matrix, one can transform $[AL_2\ K_2]$ into a canonical form, and obtain the new gross error sizes $\hat{\delta}_2$ and μ_2. To illustrate this procedure, consider the system in Figure 2. Suppose that one identified a set of gross error $\{S_2, S_3\}$ with sizes {1.5, 2.5} and wants to know the sizes in one of its equivalent sets, for example $\{S_3, S_6\}$. One has:

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 \end{bmatrix}$$

$$L_1 = \begin{bmatrix} 0 & 0\\ 1 & 0\\ 0 & 1\\ 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0 \end{bmatrix} \quad K_1 = 0 \quad L_2 = \begin{bmatrix} 0 & 0\\ 0 & 0\\ 1 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 1\\ 0 & 0\\ 0 & 0\\ 0 & 0 \end{bmatrix} \quad K_2 = 0$$

Thus, omitting leaks (the original and equivalent sets contain only biases) we obtain:

$$AL_1 = \begin{bmatrix} -1 & 0\\ 1 & -1\\ 0 & 1\\ 0 & 0\\ 0 & 0 \end{bmatrix} \quad AL_2 = \begin{bmatrix} 0 & 1\\ -1 & 0\\ 1 & -1\\ 0 & 0\\ 0 & 0 \end{bmatrix}$$

Premultiplying both sides by P:

$$P = \begin{bmatrix} 0 & -1 & 0 & 0 & 0\\ 0 & -1 & -1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

so that PAL_2 has a canonical form:

$$PAL_1\,\hat{\delta}_1 = \begin{bmatrix} -1 & 1\\ -1 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 0 \end{bmatrix}\hat{\delta}_1 = \begin{bmatrix} 1 & 0\\ 0 & 1\\ 0 & 0\\ 0 & 0\\ 0 & 0 \end{bmatrix}\hat{\delta}_2$$

Thus one gets the gross error sizes of $\{S_3, S_6\}$ as {1.0, −1.5}.

Comparative Performance Evaluation

Performance Measures. Different performance measures have been used in the past: overall power (OP), average number of type I errors (AVTI), and expected fraction of correct identification (OPF). They are defined as follows:

$$\mathrm{OP} = \frac{\text{No. of gross errors correctly identified}}{\text{No. of gross errors simulated}} \quad (16)$$

$$\mathrm{AVTI} = \frac{\text{No. of gross errors incorrectly identified}}{\text{No. of simulation trials}} \quad (17)$$

$$\mathrm{OPF} = \frac{\text{No. of trials with perfect identification}}{\text{No. of simulation trials}} \quad (18)$$

The first two measures were proposed by Mah and Narasimhan (1987), and the last one was presented by Rollins and Davis (1992). With the introduction of equivalency theory, OP and AVTI are obviously no longer proper performance measures for gross error detection strategies.

Recently, Sanchez et al. (1998) extended the definition of OPF to include all trials that identify the exact sets, the equivalent sets, and the degenerate sets as successful trials. This new measure is called overall performance of equivalent identification (OPFE):

$$\mathrm{OPFE} = \frac{\text{No. of trials with successful identification}}{\text{No. of simulation trials}} \quad (19)$$

In this article, OPFE and the estimated gross error size and its standard deviation are used as the overall performance and the estimation performance measures for all strategies selected.

Outline of Other Collective Compensation Strategies. Because the collective compensation strategies are superior to all other methods based on serial elimination and serial compensation (Rollins and Davis, 1992; Keller et al., 1994; Bagajewicz et al., 1998; Sanchez et al., 1998), the comparison is done only for collective compensation strategies. Besides SICC, three other collective compensation strategies have been identified.

1. UBET (Rollins and Davis, 1992). It proposes to use nodal strategies reported by Mah et al. (1976) and Heenan and Serth (1986) to isolate the suspect nodes and to construct the candidate bias/leak list from the suspect nodes. Then it proposes to apply the unbiased estimation technique to obtain the size estimation for the components in the candidate list. Finally, Bonferroni test is used to identify the gross errors. For comparison we use the modified version of UBET (Sanchez et al., 1998), which proposes to ensure that the number of elements in the candidate list equals the number of constraints, and all elements do not form any loop before performing the estimation. This modified version is called MUBET.

2. CGLR (Keller et al., 1994). This strategy is an improvement of Serial Compensation based on GLR (Narasimhan and Mah, 1987). This improvement consists of performing a collective compensation at each GLR identification step. The method consists of running the data reconciliation and calculating the GLR test statistic. If the GLR test does not flag, one declares no gross error and stops. Otherwise it uses the magnitudes estimated collectively by the GLR to compensate for the gross errors. The CGLR was also modified to avoid certain singularities. This modification consists of a loop check between a new candidate and the gross errors

Table 6. Performance Comparison for One Gross Error

		simulation results							
gross error introduced		average estimate/standard deviation				OPFE			
place	size	MUBET	MCGLR	MSEGE	SICC	MUBET	MCGLR	MSEGE	SICC
S1	1.000	0.999	0.999	1.000	1.000	0.898	0.915	0.960	0.929
		0.086	0.088	0.072	0.072				
S2	2.000	2.000	2.001	2.000	2.000	0.891	0.918	0.957	0.928
		0.154	0.147	0.141	0.142				
S3	3.000	3.000	3.001	3.001	3.001	0.914	0.923	0.964	0.932
		0.216	0.209	0.208	0.206				
S4	2.000	2.000	2.000	2.000	2.000	0.936	0.921	0.966	0.928
		0.159	0.147	0.150	0.144				
S5	1.000	0.999	0.999	0.999	0.999	0.918	0.911	0.958	0.927
		0.081	0.077	0.071	0.071				
S6	1.000	1.004	1.001	1.000	1.000	0.890	0.918	0.955	0.934
		0.197	0.170	0.125	0.125				
S7	1.000	0.996	1.000	1.001	1.001	0.915	0.919	0.956	0.935
		0.151	0.139	0.128	0.128				
S8	0.400	0.402	0.400	0.401	0.401	0.890	0.911	0.959	0.936
		0.074	0.071	0.061	0.061				
S9	0.600	0.601	0.601	0.601	0.601	0.924	0.943	0.970	0.945
		0.076	0.073	0.062	0.062				
L1	1.000	1.001	1.000	0.999	1.000	0.969	0.986	0.992	0.987
		0.105	0.102	0.080	0.088				
L2	1.000	1.002	1.000	1.000	1.001	0.909	0.925	0.962	0.981
		0.145	0.108	0.072	0.145				
L3	1.000	1.008	0.999	1.001	1.001	0.913	0.924	0.964	0.989
		0.199	0.097	0.075	0.130				
L4	1.000	1.001	1.000	1.000	0.993	0.944	0.922	0.977	0.982
		0.139	0.083	0.075	0.143				
L5	0.400	0.400	0.401	0.401	0.399	0.901	0.924	0.963	0.939
		0.075	0.071	0.061	0.061				

Table 7. Performance Comparison for Two Gross Errors

		simulation results							
gross error introduced		average estimate/standard deviation				OPFE			
place	size	MUBET	MCGLR	MSEGE	SICC	MUBET	MCGLR	MSEGE	SICC
S1	1.400	1.521	1.468	1.421	1.400	0.989	0.999	1.000	0.955
		0.165	0.179	0.118	0.073				
S2	1.600	1.598	1.533	1.466	1.600				
		0.168	0.176	0.137	0.142				
S4	2.800	2.804	2.800	2.800	2.800	0.919	0.933	1.000	0.954
		0.159	0.149	0.144	0.145				
S5	0.800	0.808	0.799	0.799	0.799				
		0.097	0.078	0.072	0.072				
S5	1.400	1.367	1.400	1.399	1.399	0.867	0.919	0.978	0.942
		0.124	0.078	0.072	0.072				
S6	0.800	0.853	0.806	0.802	0.802				
		0.195	0.171	0.126	0.126				
S6	1.400	1.414	1.410	1.399	1.399	0.861	0.911	0.961	0.927
		0.199	0.189	0.128	0.128				
S7	0.800	0.806	0.800	0.804	0.804				
		0.175	0.141	0.129	0.130				
S7	1.400	1.395	1.399	1.397	1.399	0.904	0.926	0.934	0.962
		0.153	0.143	0.130	0.131				
S8	0.320	0.327	0.323	0.327	0.322				
		0.074	0.072	0.058	0.062				
L1	1.400	1.398	1.397	1.394	1.396	0.987	0.965	0.959	0.968
		0.156	0.153	0.150	0.152				
L2	0.800	0.805	0.807	0.806	0.807				
		0.169	0.156	0.139	0.150				
L2	1.400	1.371	1.359	1.344	1.366	0.913	0.876	0.831	0.899
		0.218	0.198	0.174	0.216				
L3	0.800	0.833	0.840	0.858	0.840				
		0.194	0.184	0.167	0.150				
L3	1.400	1.395	1.397	1.395	1.426	0.969	0.941	0.956	0.999
		0.226	0.165	0.147	0.223				
L4	0.800	0.800	0.804	0.806	0.798				
		0.150	0.147	0.142	0.146				
L4	1.400	1.398	1.396	1.392	1.372	0.952	0.887	0.914	0.933
		0.153	0.093	0.078	0.133				
L5	0.320	0.324	0.326	0.332	0.330				
		0.072	0.070	0.067	0.068				

Table 8. Performance Comparison for Three Gross Errors

gross error introduced		simulation results							
		average estimate/standard deviation				OPFE			
place	size	MUBET	MCGLR	MSEGE	SICC	MUBET	MCGLR	MSEGE	SICC
		a. 7, 5, 3 Standard Deviations							
S4	2.800	2.792	2.797	2.774	2.798				
		0.168	0.148	0.258	0.146	0.705	0.891	0.886	0.907
S5	1.000	1.002	0.999	0.994	0.998				
		0.115	0.076	0.085	0.072				
S6	0.600	0.687	0.624	0.647	0.610				
		0.148	0.170	0.313	0.122				
S5	1.400	1.273	1.397	1.210	1.397				
		0.210	0.088	0.364	0.079	0.638	0.864	0.866	0.880
S6	1.000	1.089	1.023	0.996	1.000				
		0.204	0.190	0.130	0.127				
S7	0.600	0.737	0.605	0.807	0.612				
		0.157	0.140	0.335	0.125				
S6	1.400	1.394	1.403	1.392	1.395				
		0.156	0.200	0.130	0.126	0.795	0.827	0.770	0.832
S7	1.000	1.003	0.990	0.995	0.993				
		0.227	0.153	0.134	0.138				
S8	0.240	0.266	0.257	0.260	0.258				
		0.060	0.075	0.051	0.067				
S7	1.400	1.386	1.346	1.391	1.355				
		0.170	0.173	0.140	0.155	0.927	0.590	0.843	0.605
S8	0.400	0.402	0.403	0.404	0.430				
		0.076	0.081	0.072	0.086				
L1	0.600	0.612	0.603	0.612	0.601				
		0.137	0.130	0.083	0.126				
S8	0.560	0.563	0.584	0.565	0.577				
		0.075	0.091	0.076	0.073	0.910	0.859	0.816	0.848
L1	1.000	0.989	1.031	0.985	1.023				
		0.145	0.127	0.134	0.123				
L2	0.600	0.618	0.563	0.635	0.561				
		0.150	0.157	0.144	0.130				
L2	1.400	1.388	1.381	1.376	1.371				
		0.229	0.220	0.204	0.22	0.942	0.912	0.882	0.876
L3	1.000	1.004	1.009	1.001	1.024				
		0.219	0.214	0.210	0. 200				
L4	0.600	0.609	0.610	0.614	0.608				
		0.137	0.134	0.130	0.129				
L3	1.400	1.395	1.399	1.395	1.404				
		0.233	0.175	0.161	0.195	0.808	0.739	0.691	0.680
L4	1.000	0.984	0.980	0.978	0.974				
		0.151	0.148	0.147	0.148				
L5	0.240	0.264	0.272	0.276	0.280				
		0.060	0.056	0.055	0.053				
		b. 9, 7, 5 Standard Deviations							
S4	3.600	3.602	3.599	3.565	3.600				
		0.174	0.148	0.315	0.146	0.964	0.957	0.967	0.982
S5	1.400	1.402	1.399	1.393	1.399				
		0.122	0.076	0.093	0.072				
S6	1.000	1.001	1.000	1.045	1.001				
		0.201	0.189	0.374	0.127				
S5	1.800	1.774	1.799	1.786	1.799				
		0.243	0.083	0.123	0.075	0.979	0.954	0.966	0.978
S6	1.400	1.408	1.403	1.401	1.402				
		0.230	0.204	0.138	0.126				
S7	1.000	1.027	0.999	1.016	1.000				
		0.211	0.144	0.167	0.129				
S6	1.800	1.798	1.803	1.799	1.801				
		0.154	0.205	0.131	0.127	0.988	0.962	0.957	0.992
S7	1.400	1.400	1.394	1.400	1.396				
		0.236	0.152	0.134	0.137				
S8	0.400	0.401	0.403	0.401	0.402				
		0.077	0.083	0.063	0.072				
S7	1.800	1.786	1.810	1.796	1.803				
		0.170	0.176	0.141	0.154	0.971	0.879	0.957	0.883
S8	0.560	0.561	0.559	0.561	0.562				
		0.077	0.078	0.073	0.076				
L1	1.000	1.003	0.992	1.004	1.000				
		0.149	0.116	0.092	0.112				
S8	0.720	0.721	0.737	0.722	0.738				
		0.077	0.089	0.075	0.078	0.981	0.989	0.987	0.988
L1	1.400	1.399	1.496	1.414	1.493				
		0.153	0.201	0.168	0.199				
L2	1.000	1.000	0.872	0.990	0.868				
		0.173	0.235	0.177	0.217				

Table 8 (Continued)

		simulation results							
gross error introduced		average estimate/standard deviation				OPFE			
place	size	MUBET	MCGLR	MSEGE	SICC	MUBET	MCGLR	MSEGE	SICC
b. 9, 7, 5 Standard Deviations									
L2	1.800	1.800	1.800	1.800	1.800				
		0.236	0.232	0.216	0.236	0.999	0.999	0.999	1.000
L3	1.400	1.400	1.400	1.400	1.400				
		0.235	0.235	0.235	0.235				
L4	1.000	1.000	1.000	1.000	1.000				
		0.151	0.150	0.148	0.147				
L3	1.800	1.799	1.800	1.799	1.800				
		0.235	0.180	0.164	0.184	0.997	0.996	0.994	0.995
L4	1.400	1.400	1.400	1.400	1.400				
		0.154	0.154	0.153	0.154				
L5	0.400	0.400	0.400	0.400	0.400				
		0.076	0.076	0.076	0.075				

Table 9. Performance Comparison for Four Gross Errors

		simulation results							
gross error introduced		average estimate/standard deviation				OPFE			
place	size	MUBET	MCGLR	MSEGE	SICC	MUBET	MCGLR	MSEGE	SICC
S2	3.600	3.758	3.759	3.651	3.673				
		0.258	0.283	0.171	0.179	0.990	1.000	1.000	1.000
S3	4.200	4.238	4.198	4.081	4.016				
		0.269	0.237	0.309	0.331				
S4	2.000	2.308	2.007	1.989	1.979				
		0.203	0.165	0.156	0.153				
S5	0.600	0.638	0.607	0.603	0.599				
		0.109	0.104	0.081	0.074				
S3	5.400	5.408	5.409	5.409	5.411				
		0.234	0.233	0.233	0.232	0.958	0.956	0.960	0.948
S4	2.800	2.799	2.799	2.799	2.799				
		0.151	0.150	0.149	0.147				
S5	1.000	0.998	0.998	0.998	0.998				
		0.081	0.079	0.077	0.074				
S6	0.600	0.613	0.613	0.613	0.616				
		0.137	0.136	0.135	0.131				
S7	1.800	1.796	1.800	1.800	1.801				
		0.153	0.151	0.140	0.140	0.955	0.966	0.971	0.981
S8	0.560	0.563	0.560	0.565	0.565				
		0.076	0.077	0.076	0.076				
L1	1.000	0.994	1.019	1.028	1.029				
		0.149	0.104	0.095	0.092				
L2	0.600	0.598	0.581	0.573	0.569				
		0.156	0.108	0.111	0.085				
S8	0.720	0.730	0.723	0.728	0.723				
		0.076	0.076	0.077	0.074	0.679	0.821	0.662	0.839
L1	1.400	1.387	1.420	1.386	1.401				
		0.151	0.176	0.151	0.157				
L2	1.000	0.925	0.888	0.910	0.891				
		0.182	0.167	0.173	0.148				
L3	0.600	0.697	0.695	0.719	0.713				
		0.144	0.159	0.163	0.142				
L1	1.800	1.799	1.799	1.799	1.797				
		0.154	0.154	0.153	0.154	0.946	0.924	0.927	0.876
L2	1.400	1.390	1.385	1.386	1.371				
		0.230	0.227	0.228	0.221				
L3	1.000	1.003	1.008	1.007	1.024				
		0.219	0.214	0.215	0.200				
L4	0.600	0.609	0.610	0.610	0.608				
		0.137	0.135	0.134	0.129				
L2	1.800	1.800	1.800	1.800	1.801				
		0.237	0.237	0.221	0.239	0.819	0.758	0.772	0.680
L3	1.400	1.399	1.400	1.399	1.399				
		0.236	0.236	0.235	0.236				
L4	1.000	0.984	0.979	0.981	0.974				
		0.151	0.149	0.150	0.149				
L5	0.240	0.264	0.271	0.269	0.280				
		0.060	0.057	0.058	0.053				

previously identified. If there is any loop, the new candidate is eliminated from the candidate list. This modified version is called MCGLR.

3. MSEGE (Sanchez et al., 1998). This strategy was originally proposed by Sanchez and Romagnoli (1994) and later modified by Sanchez et al. (1998). It consists of two stages. The first stage is the construction of the gross error candidate list. It starts with one constraint and performs the data reconciliation and calculates the objective function value (ofv). If the global test fails, then it disregards the constraint and puts the node with the largest GT in the candidate node list. If this node is already in the list, it puts the node with the second largest GT in the candidate node list and so on. Then, a new constraint is added and the entire procedure is repeated. It then proposes to include all measurements involved in the list of suspect nodes and the nodes itself in a candidate bias/leak list. The second stage is the identification and estimation of gross errors. It assumes that there are n gross errors in the candidate bias/leak list ($n = 1$ at the beginning). It determines which combination of n gross errors gives the lowest objective function value (the set forming a loop is omitted). It then increases n ($n = n + 1$) and repeats this search until the global test is satisfied.

Results and Discussion

The performance of SICC was tested and compared with MUBET, MCGLR, and MSEGE on a basis of 10 000 simulation trials in which the random errors are changed and the magnitudes of gross errors are fixed.

We use the process in Figure 2 as an example to conduct this comparison. The true flowrate values are $x^+ = [10,20,30,20,10,10,10,4,6]$, and the standard deviation for each flowrate was taken as 2% of its true value. Measurement values for each simulation trial were taken as the average of 10 random generated values. To compare results on the same basis, the level of significance of each method was chosen such that it gives an AVTI equal to 0.1 under the null hypothesis. This common basis of comparison for gross error detection schemes was introduced by Rosenberg et al. (1987). The same confidence levels for comparison have also been used by other authors. If a comparison is performed on such basis the conclusions of this article do not vary.

Table 6 shows the comparison when only one gross error is introduced. The size of the bias is five times of the corresponding flowrate standard deviation, and the size of the leak is five times of the minimum standard deviation of flowrates connected to the corresponding process.

In general, SICC and MSEGE have slightly higher OPFE and lower standard deviations for gross error estimates, whereas all the four methods reach relatively high power and similar performance in the average size estimation.

Tables 7−9 show the comparisons for the increasing number of gross errors introduced. The results for two gross errors with sizes of 7 and 4 standard deviations are in Table 7. Three gross errors with sizes of 7, 5, 3 and 9, 7, 5 standard deviations are simulated in Tables 8a and 8b, and four gross errors with sizes of 9, 7, 5, 3 standard deviations are simulated in Table 9.

None of the four methods has consistently higher power than the others. Table 8a and 8b indicate the effect of gross error sizes on the power; as expected, the power for each method is improved when gross error sizes are increased.

Conclusions

The SICC method as applied to steady-state bias and leak detection has been introduced. Any leak can be detected through its equivalent biases; thus the SICC can be applied to all gross error identification and compensation. The conversion of gross error sizes from one equivalent set to another is also formulated. Overall, the results show that the SICC has comparable, sometimes better performance than three other collective compensation strategies.

Acknowledgment

Partial financial support from KBC Advanced Technologies for Q. Jiang is acknowledged.

Nomenclature

A = constraint matrix
$AVTI$ = average number of type I errors
e_i = ith unit vector
f = vector defined by eq 11
h = matrix defined by eq 9
I = identity matrix
K = matrix defined by eq 10
L = matrix defined by eq 4
OP = overall power
OPF = expected fraction of perfect identification
OPFE = expected fraction of successful identification
Q = covariance matrix
x = reconciled data without assuming biases
$\hat{x}$ = reconciled data when assuming biases
x^+ = measurements

Greek Letters

δ = vector of instrument biases
μ = vector of leaks

Subscripts or Superscripts

T = transpose of matrix
-1 = inverse of matrix

Literature Cited

(1) Bagajewicz, M.; Jiang, Q. An Integral Approach to Dynamic Data Reconciliation. *AIChE J.* **1997**, *43*, 2546.

(2) Bagajewicz, M.; Jiang, Q. Gross Error Modeling and Detection in Plant Linear Dynamic Reconciliation. *Comput. Chem. Eng.* **1998**, in press.

(3) Bagajewicz, M.; Jiang, Q.; Sanchez, M. Removing Singularities and Assessing Uncertainties in Two Efficient Gross Error Collective Compensation Methods. *Chem. Eng. Commun.* **1998**, submitted for publication.

(4) Crowe, C. M.; Hrymak A.; Campos, Y. A. G. Reconciliation of Process Flow Rates by Matrix Projection. I. The Linear Case. *AIChE J.* **1983**, *29*, 881.

(5) Keller, J. Y.; Darouach M.; Krzakala, G. Fault detection of Multiple Biases or process leaks in Linear Steady-State Systems. *Comput. Chem. Eng.* **1994**, *18*, 1001.

(6) Kim I.; Kang, M. S.; Park S.; Edgar T. Robust Data Reconciliation and Gross Error Detection: The Modified MIMT using NLP. *Comput. Chem. Eng.* **1997**, *21*, 775.

(7) Madron, F. A New Approach to the Identification of Gross Errors in Chemical Engineering Measurements. *Chem. Eng. Sci.* **1985**, *40*, 1855.

(8) Madron, F. *Process Plant Performance. Measurement and Data Processing for Optimization and Retrofits;* Ellis Horwood Series in Chemical Engineering: Chichester, England, 1992.

(9) Mah, R. S. H. *Chemical Process Structures and Information Flows;* Butterworth: Stoneham, MA, 1990.

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

(11) Mah, R. S. H.; Tamhane, A. C. Detection of Gross Errors in Process Data. *AIChE J.* **1982**, *28*, 828.

(12) Narasimhan, S.; Mah, R. S. H. Generalized Likelihood Ratio Method for Gross Error Detection. *AIChE J.* **1987**, *33*, 1514.

(13) Reilly, P. M.; Carpani, R. E. Application of Statistical Theory of Adjustment to Material Balances. Presented at 13th Canadian Chemical Engineering Conference, Montreal, Quebec, 1963.

(14) Rollins, D. K.; Davis, J. F. Unbiased Estimation of Gross Errors in Process Measurements. *AIChE J.* **1992**, *38*, 563.

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

(16) Sanchez, M. Ph.D. Dissertation, Universidad Nacional del Sur, Argentina, 1996.

(17) Sanchez, M.; Romagnoli, J.; Jiang Q.; Bagajewicz, M. Simultaneous Estimation of Biases and Leaks in Process Plants. *Comput. Chem. Eng.* **1998**, submitted for publication.

(18) Sanchez, M.; Romagnoli, J. Use of Orthogonal Transformations in Classification/Data Reconciliation. *Comput. Chem. Eng.* **1996**, *20*, 483.

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

(20) Swartz, C. L. E. Data Reconciliation for Generalized Flowsheet Application. Presented at the American Chemical Society National Meeting, Dallas, TX, 1989.

(21) Tong, H.; Crowe, C. M. Detection of Gross Errors in Data Reconciliation by Principal Component Analysis. *AIChE J.* **1995**, *41*, 1712.

Received for review July 6, 1998
Accepted February 9, 1999

IE980431E