Chem. Eng. Comm., 2000, Vol. 177, pp. 139–155
Reprints available directly from the publisher
Photocopying permitted by license only

© 2000 OPA (Overseas Publishers Association) N.V.
Published by license under
the Gordon and Breach Science
Publishers imprint.
Printed in Malaysia.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

A MIXED INTEGER LINEAR PROGRAMMING-BASED TECHNIQUE FOR THE ESTIMATION OF MULTIPLE GROSS ERRORS IN PROCESS MEASUREMENTS

MIGUEL J. BAGAJEWICZ* and QIYOU JIANG

School of Chemical Engineering and Materials Science, University of Oklahoma, 100 E. Boyd St., T-335, Norman OK 73019-0628

(Received 20 November 1998; In final form 11 May 1999)

This paper presents a new strategy for detecting, identifying, and estimating gross errors (measurement biases and leaks) in linear steady state processes. The MILP-based gross error detection and identification model is constructed aiming at identifying the minimum number of gross errors and their sizes. One significant advantage of the method is that the detection, identification, and estimation of gross errors can be performed simultaneously without using any test statistics.

Keywords: Linear programming; error estimation; process measurements

INTRODUCTION

Since process measurements contain random errors, data reconciliation has been used to adjust the measurements so that they comply with conservation laws. Measurements also contain gross errors (instrument biases and leaks), which behave differently from the random noise. For years, researchers have been developing good techniques to detect and eliminate these gross errors.

Two central issues are of concern in the case of gross errors: proper location of gross errors and estimation of their sizes. Three kinds of strategies

*Corresponding author.

are available. They are: Serial Elimination, Serial Compensation and Collective Compensation. Serial Elimination (Ripps, 1965; Serth and Heenan, 1986; Rosenberg *et al.*, 1987) and Serial Compensation (Narasimhan and Mah, 1987) have been proven to be relatively efficient when one gross error is present, but do not perform well in the presence of multiple gross errors. While serial elimination does not address leaks (Mah, 1990), serial compensation is applicable to all types of gross errors and can maintain redundancy during the procedure, but its results are completely dependent on the accuracy of estimation for the size of gross errors (Rollins and Davis, 1992). To improve these methods, simultaneous or collective compensation proposes the estimation of all gross errors simultaneously (Rollins and Davis, 1992; Keller *et al.*, 1994; Sánchez *et al.*, 1999; Bagajewicz and Jiang, 1999).

In this paper, a new strategy for the collective identification and estimation of gross errors using mixed integer linear programming technique is presented. First, two useful concepts of the equivalency theory and the spanning tree are reviewed. Then the model is constructed and its performance are evaluated and discussed.

REVIEW OF EQUIVALENCY THEORY

In a recent paper (Bagajewicz and Jiang, 1998) a series of concepts regarding the equivalency of sets of gross errors were presented. Two sets of gross errors are equivalent when they have the same effect in data reconciliation, that is, when simulating either one using a compensation model, 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, 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 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. This number is the Gross Error Cardinality of the set (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 I, a bias of (-2) in S_4 and a bias of $(+1)$ in S_5 (Case 1) can be

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

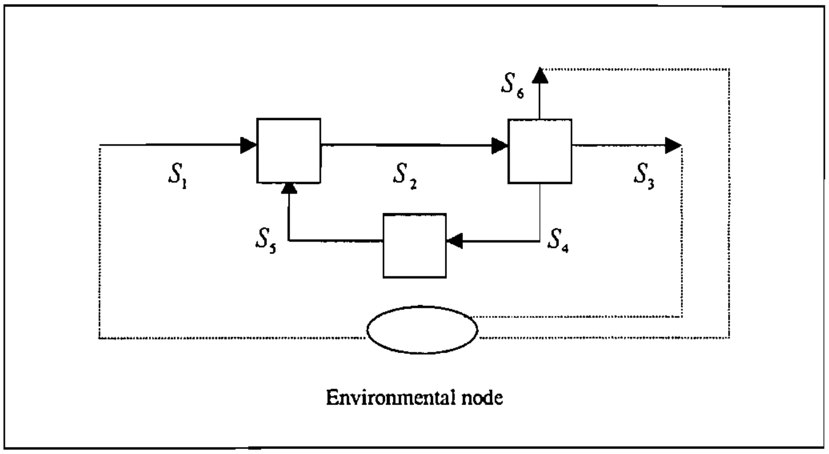

FIGURE 1 Illustration of Gross Error Equivalency.

TABLE I 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	

equally represented by two alternative sets of two gross errors (Cases 2 and 3). 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 where a number of gross errors different with 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 II, 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.

Consider now 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.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

TABLE II 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				

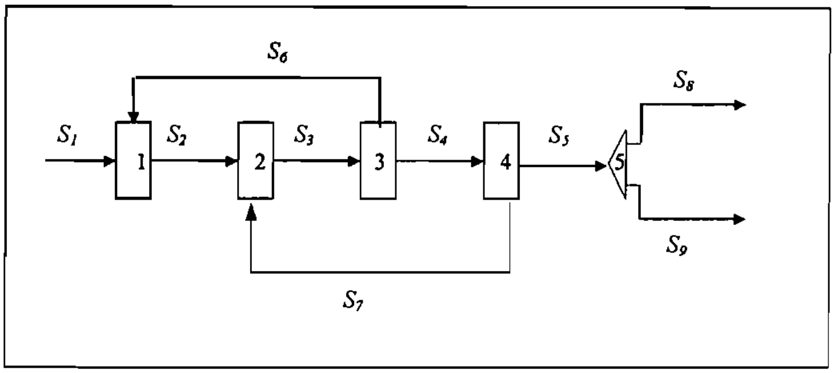

FIGURE 2 Illustration of gross error degeneracy.

One of the characteristics of collective compensation methods is that they absorb part of the random errors. Thus, to make the illustrations of gross error detection clear, random errors have been ignored 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 III. In Case 1, the system consisting of S_1, S_4, S_5, S_6, S_9 has a cardinality $\Gamma = 4$. 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. A more detailed discussion of this can be found in Bagajewicz and Jiang (1998).

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 IV provided some cases, which in this paper are called Quasi-Degeneracy.

TABLE III 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 IV 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

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

In light of what has been defined in the literature as identification performance, one could consider this as a failure. We will show that this is actually a result of part of the gross errors being absorbed as corrections to random errors. 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. This phenomenon is called "masking" by some practitioners (DATACON, Simsci), it is discussed by Madron (1992) and is related to the concept of error detectability (Bagajewicz, 1997).

THE RELATION BETWEEN TWO EQUIVALENT SETS

A detailed explanation for the relation between two equivalent sets is presented in another paper (Jiang and Bagajewicz, 1999) and is briefly summarized as follows.

A linear steady state data reconciliation problem can be formulated as

$$\begin{aligned} &\text{Min}\,(x - x^+)^T Q^{-1}(x - x^+) \\ &Ax = 0 \end{aligned} \tag{1}$$

where x is a vector of reconciled data and x^+ is a vector of measurements. Q and A are the covariance matrix and incidence matrix, respectively. For simplicity, in this formulation it is assumed that all variables are measured. This is not a limitation since in the presence of unmeasured variables several methods exist to obtain a matrix of only redundant measurement: Matrix Projection (Crowe *et al.*, 1983), QR decomposition (Swartz, 1989; Sanchez and Romagnoli, 1996), Co-optation (Madron, 1992) and node aggregation (Sanchez *et al.*, 1999).

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

$$\begin{aligned} &\text{Min}(\hat{x} + L\hat{\delta} - x^+)^T Q^{-1}(\hat{x} + L\hat{\delta} - x^+) \\ &\text{s.t.} \\ &\qquad A\hat{x} - K\hat{\mu} = 0 \end{aligned} \tag{2}$$

where $\hat{\delta}$ is a vector that contains biases for bias candidates, $L = [e_1\, e_2 \ldots e_{nb}]$ and e_i is a vector with unity in a position corresponding to a bias candidate and zero elsewhere, $K = [e_1\, e_2 \ldots e_{nl}]$ and $\hat{\mu}$ is a vector of leaks.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

Let

$$f = \hat{x} + L\hat{\delta} \tag{3}$$

Therefore one can rewrite Eq. (1) as follows:

$$\begin{aligned} &\mathrm{Min}(f - x^{+})^{T} Q^{-1}(f - x^{+}) \\ &\text{s.t.} \\ &\quad Af - AL\hat{\delta} - K\hat{\mu} = 0 \end{aligned} \tag{4}$$

The following holds between two equivalent sets:

$$f_1 = f_2 \tag{5}$$

Thus one gets:

$$AL_1\hat{\delta}_1 + K_1\hat{\mu}_1 = AL_2\hat{\delta}_2 + K_2\hat{\mu}_2 \tag{6}$$

Let

$$\delta = L\hat{\delta} \tag{7}$$

$$\mu = K\hat{\mu} \tag{8}$$

Then one also has:

$$A\delta_1 + \mu_1 = A\delta_2 + \mu_2 \tag{9}$$

Equation (9) constitutes the relationship between one set of gross errors (δ_1, μ_1) and another (δ_2, μ_2).

THE CONCEPT OF SPANNING TREE

In graph theory, a spanning tree is defined as a connected sub-graph of a connected graph which contains all the vertices of the graph but does not

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

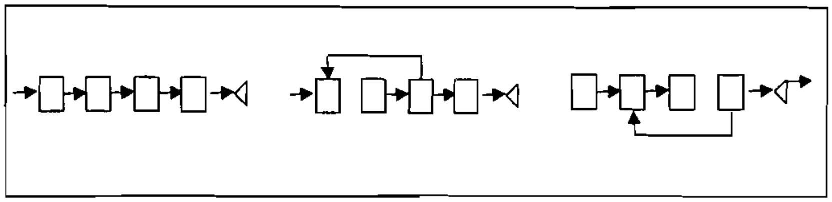

FIGURE 3 Three spanning trees of Figure 2.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

contain any circuits (Seshu and Reed, 1961). Three spanning trees of the graph from Figure 2 are shown in Figure 3.

If one resorts to the connection between graph theory and equivalency theory, a spanning tree is a basic set of maximum cardinality, that is, of cardinality equal to the number of units in a system. Thus, one can use the spanning tree as a set of gross error candidates in a compensation model and capture all gross errors. Those elements in the spanning tree that show a value lower than a threshold can be disregarded.

THE MILP BASED STRATEGY

We start with an important assumption as the theoretical basis for this technique:

ASSUMPTION *If gross errors exist in a system, n gross errors have always larger probability to appear than $n+1$ gross error.*

The assumption needs a small explanation. If a certain number of gross errors exist, gross error detection techniques can detect a (sometimes large) number of gross errors that are equivalent to the set introduced. This is due to the phenomenon of degeneracy for pure biases and to the fact that certain tests do not detect leaks, and therefore they detect their equivalent biases. Therefore, the assumption relies on the fact that these degenerate cases are less likely to happen because they require having similar bias/leak size. The assumption is also leaning towards a certain pattern of gross errors and is probabilistic in nature. Hence, it might not always detect all occurrences of errors.

Let $\hat{\delta}$ be the set of gross errors identified using the spanning tree as candidate in a compensation model. Assume now that the least number of gross errors is to be identified (this will eliminate degeneracy). Then, we use the following model, which minimizes the number of gross errors:

$$\left.\begin{aligned}&\text{Min} \sum_{i=1}^{Ns} Y_i + \sum_{j=1}^{Nu} Z_j\\&\text{s.t.}\\&|\delta_i| - UY_i \leq 0\\&|\delta_i| - \varepsilon_i Y_i \geq 0\\&|\mu_j| - UZ_j \leq 0\\&|\mu_j| - \varepsilon_j Z_j \geq 0\\&A\hat{\delta} - A\delta - \mu = 0\end{aligned}\right\} \tag{10}$$

The binary variable Y_i represents the presence of a gross error of size $\delta_i \geq \varepsilon_i$ in stream i (U is a large number). When $Y_i = 0$, δ_i is forced to be zero, whereas when $Y_i = 1$, δ_i can take any value such larger than ε_i or lower than $-\varepsilon_i$. The same type of constraints is used for leaks. The last constraint represents the equivalency between the values of the gross errors in the spanning tree $\hat{\delta}$ and the set of new set, hopefully of smaller cardinality (δ, μ).

Remark 1 The proposed MILP-based strategy does not use any statistical test.

Remark 2 The Unbiased estimation technique, as proposed by Rollins and Davis (1992) and modified to avoid singularities by Bagajewicz *et al.* (1999) is based on selecting a spanning tree to perform the estimation.

SOLUTION PROCEDURE

To solve this problem the following is used:

$$|\delta| = \text{Max}(\delta, -\delta) \tag{11}$$

We now replace the *Max operator* by a set of mixed integer linear constraints proposed by Bagajewicz and Manousiouthakis (1992):

$$C = \text{Max}(A, B) \Leftrightarrow \begin{cases} A + \lambda\Gamma \geq C \\ C \geq A \\ B + (1-\lambda)\Gamma \geq C \\ C \geq B \end{cases} \tag{12}$$

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

Then, denoting $d_i = |\delta_i|$ and $m_j = |\mu_j|$ the above problem becomes

$$\left.\begin{array}{l}
\text{Min} \displaystyle\sum_{i=1}^{Ns} Y_i + \sum_{j=1}^{Nu} Z_j \\
\text{s.t.} \\
\quad d_i - UY_i \leq 0 \\
\quad d_i - \varepsilon_i Y_i \geq 0 \\
\\
\delta_i + \lambda_i \, \Gamma \geq d_i \\
d_i \geq \delta_i \\
-\delta_i + (1 - \lambda_i)\Gamma \geq d_i \\
d_i \geq -\delta_i \\
\\
\quad m_j - UZ_j \leq 0 \\
\quad m_j - \varepsilon_j Z_j \geq 0 \\
\\
\mu_j + \kappa_j \, \Gamma \geq m_j \\
m_j \geq \mu_j \\
-\mu_j + (1 - \kappa_j)\Gamma \geq m_j \\
m_j \geq -\mu_j \\
\\
\quad A\hat{\delta} - A\delta - \mu = 0
\end{array}\right\} \tag{13}$$

which is a MILP problem.

ALGORITHM

The gross error identification steps can be described as:

A. Let U equal to 1000 times of the largest measurement in the system and ε_k equal to two times of the corresponding standard deviation. For leak candidates, ε_k is equal to two times of the minimum standard deviation in the system.
B. Obtain the redundant system from the original system.
C. Get a spanning tree for the redundant system.
D. Use the spanning tree as a set of gross error candidates in a compensation model and estimate the sizes of gross errors.
E. Those elements in the spanning tree that have a lower gross error size than a threshold are disregarded. This threshold is selected as two times of the minimum standard deviation of the streams in the spanning tree.
F. Run the MILP model to identify a suspect set of gross errors with the minimum number.
G. Find out all equivalent sets to the identified set.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

In this procedure, the threshold ε_k for leaks is selected as two times of the minimum standard deviation in the system, that is the minimum value among all the standard deviations of the measurements in the system. Although this seems to be an arbitrary choice, there is rationale behind it. In a bias detection scheme a leak manifests as a sequence of biases (Jiang and Bagajewicz, 1999) from the unit in question to one outlet stream to the environment. If such a leak is to be considered, all the equivalent biases should be detectable, that is, the threshold should be the smallest of all possible biases that can be part of an equivalent set to this leak. More sophisticated strategies can be used. For example, one can only consider the minimum standard deviation of the streams that are likely to be part of an equivalent set. Finally, as in UBET, there is no hypothesis testing involved in these choices. They are simply a result of the application of equivalency theory and the flowsheet-inherent uncertainty of the location of gross error, which is method independent.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

RESULTS

The proposed strategy was tested in a variety of situations. In order to avoid the problems described above, the measurements in a system are limited in a certain range and the sizes for the introduced gross errors are intentionally designed. The first example is for the process in Figure 4. The true values for this system are $x = [12, 18, 10, 6, 6, 2]$ and the standard deviations are 2% of each measurement. Let us first focus on the performance when all measurements have no random errors.

Table V shows the results of the application of the technique. The spanning tree (S_1, S_2, S_5) is identified and used.

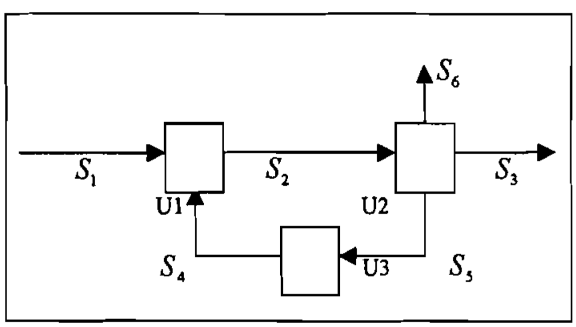

FIGURE 4

TABLE V Gross error identification with the minimum cardinality model

Case no.	*Gross error introduced*	*Initial set with spanning tree*	*Reduced initial set*	*Gross errors identified*
1	S_2 (2.0)	S_1 (0.0) S_2 (2.0) S_5 (0.0)	S_2 (2.0)	S_2 (2.0)
2	S_4 (1.0)	S_1 (0.0) S_2 (− 1.0) S_5 (− 1.0)	S_2 (− 1.0) S_5 (− 1.0)	S_4 (1.0)
3	U_2 (1.0)	S_1 (1.0) S_2 (1.0) S_5 (0.0)	S_1 (1.0) S_2 (1.0)	U_2 (1.0)
4	S_1 (1.0) S_2 (2.0)	S_1 (1.0) S_2 (2.0) S_5 (0.0)	S_1 (1.0) S_2 (2.0)	S_1 (1.0) S_2 (2.0)
5	S_5 (1.0) U_2 (2.0)	S_1 (2.0) S_2 (2.0) S_5 (1.0)	S_1 (2.0) S_2 (2.0) S_5 (1.0)	U_2 (1.0) U_3 (1.0)
6	S_1 (1.0) S_2 (3.0) S_4 (1.0)	S_1 (1.0) S_2 (2.0) S_5 (− 1.0)	S_1 (1.0) S_2 (2.0) S_5 (− 1.0)	S_1 (1.0) S_2 (2.0) S_5 (− 1.0)
7	S_1 (1.0) S_2 (2.0) S_4 (1.0)	S_1 (1.0) S_2 (1.0) S_5 (− 1.0)	S_1 (1.0) S_2 (1.0) S_5 (− 1.0)	U_2 (2.0) U_3 (− 1.0)

Note: S_n means a bias in stream n; U_n means a leak in unit n.

In Runs 1, 2, 3 and 4, exact errors are identified. In Runs 4, 5 and 6, equivalent sets are obtained. Run 7 is a degenerate case.

Table VI shows the results when random errors to all measurements. The measurements are $x^+ = [11.96, 18.05, 10.04, 5.98, 6.04, 2.01]$.

In comparing Tables V and VI, one may notice that in 4 out of 7 cases they have exactly the same results. For Case 5, the two results are equivalent. For Cases 3 and 7, the results are different since random errors affect the occurrence of degeneracy.

A second example was prepared for the case of Figure 3. Suppose 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. Table VII shows the results of these experiments.

In this experiment, there are two failures. They are the fourth case and the third case of four gross errors. All others are successful, either exactly or equivalently. "Exact" means the same gross errors as introduced have been identified and "equivalent" means the gross errors equivalent to the introduced have been identified.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

TABLE VI Gross error identification with the minimum cardinality model

Case no.	*Gross error introduced*	*Initial set with spanning tree*	*Reduced initial set*	*Gross errors identified*
1	S_2 (2.0)	S_1 (−0.09) S_2 (2.02) S_5 (0.06)	S_2 (2.02)	S_2 (2.02)
2	S_4 (1.0)	S_1 (−0.09) S_2 (−0.98) S_5 (−0.94)	S_2 (−0.98) S_5 (−0.94)	S_4 (0.98)
3	U_2 (1.0)	S_1 (0.91) S_2 (1.02) S_5 (0.06)	S_1 (0.91) S_2 (1.02)	S_1 (0.91) S_2 (1.02)
4	S_1 (1.0) S_2 (2.0)	S_1 (0.91) S_2 (2.02) S_5 (0.06)	S_1 (0.91) S_2 (2.02)	S_1 (0.91) S_2 (2.02)
5	S_5 (1.0) U_2 (2.0)	S_1 (1.91) S_2 (2.02) S_5 (1.06)	S_1 (1.91) S_2 (2.02) S_5 (1.06)	S_5 (−0.96) U_3 (2.02)
6	S_1 (1.0) S_2 (3.0) S_4 (1.0)	S_1 (0.91) S_2 (2.02) S_5 (−0.94)	S_1 (0.91) S_2 (2.02) S_5 (−0.94)	S_1 (0.91) S_2 (2.02) S_5 (−0.94)
7	S_1 (1.0) S_2 (2.0) S_4 (1.0)	S_1 (0.91) S_2 (1.02) S_5 (−0.94)	S_1 (0.91) S_2 (1.02) S_5 (−0.94)	S_1 (0.91) S_2 (1.02) S_5 (−0.94)

GROSS ERROR MASKING

The new MILP-based strategy, like other strategies available so far, has no guarantee to be successful in all cases. Following, we analyze some cases in which it may fail.

Threshold Value Criterion

When a spanning tree is used for in a gross error compensation model, it may capture not only gross errors but also some random errors. Therefore, one needs a certain criterion to determine the elements that should be considered as gross errors in this spanning tree. This criterion is crucial since the probability of type I/type II error is dependent on it to some extents. One may use some tests, like Bonferroni tests used by Rollins and Davis (1992), as the criterion for the threshold. For simplicity, in this strategy the criterion is selected as two times of the smallest standard deviation involved.

Consider the process in Figure 3. Assume the measurements are 1.01, 2.04, 2.97, 2.015, 0.998, 1.01, 0.99, 0.406, 0.606 and the standard deviations are 2% of each measurement. If one selects {$S1$, $S2$, $S3$, $S4$, $S5$} as the

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

TABLE VII Gross error identification with the minimum cardinality model

	Gross error introduced		Gross error captured with spanning tree	Minimum number of gross error identified	
Number	Location	Size	(S1, S2, S3, S4, S5)	Location	Size
1	S1	0.08	0.08, 0., 0., 0., 0.	S1	0.08
	S3	0.24	0., 0., 0.24, 0., 0.	S3	0.24
	S5	0.08	0., 0., 0., 0., 0.08	S5	0.08
	S7	0.08	0., 0., −0.08, −0.08, 0.	S7	0.08
	L3	0.08	0.08, 0.08, 0.08, 0., 0.	L3	0.08
2	S1, S4	0.1, 0.12	0.1, 0., 0., 0.12, 0.	S1, S4	0.1, 0.12
	S3, S8	0.3, 0.09	−0.09, −0.09, 0.21, −0.09, −0.09	S3, L5	0.3, −0.09
	S6, S7	0.1, 0.09	0., −0.1, −0.19, −0.09, 0.	S6, S7	0.1, 0.09
	S2, S5	0.2, 0.09	0., 0.2, 0.,0., 0.09	S2, S5	0.2, 0.09
	S4, L2	0.2, 0.09	0.09, 0.09, 0., 0.2, 0.	S4, L2	0.2, 0.09
3	S1, S3, S5	0.1, 0.24, 0.09	0.1, 0., 0.24, 0., 0.09	S1, S3, S5	0.1, 0.24, −0.09
	S2, S4, S8	0.2, 0.16, 0.09	−0.09, 0.11, −0.09, 0.07, −0.09	S2, S4, L5	0.2, 0.16, −0.09
	S6, S7, S9	0.2, 0.16, 0.14	−0.14, −0.34, −0.50, −0.30, −0.14	S6, S7, L5	0.20, 0.16, −0.14
	S1, L2, L4	0.2, 0.16, 0.14	0.50, 0.30, 0.14, 0.14, 0.	L1, L2, L4	0.20, 0.16, 0.14
	S7, S8, L3	0.2, 0.16, 0.14	−0.02, −0.02, −0.22, −0.36, −0.16	S3, S4, S5	−0.22, −0.36, −0.16
4	S1, S2, S4, S5	0.2, 0.16, 0.14, 0.18	0.20, 0.16, 0., 0.14, 0.18	S1, S2, S4, S5	0.20, 0.16, 0.14, 0.18
	S4, S6, S7, S8	0.2, 0.16, 0.14, 0.18	−0.18, −0.34, −0.48, −0.12, −0.18	S4, S6, S7, L5	0.20, 0.16, 0.14, −0.18
	S8, S5, S3, S1	0.2, 0.16, 0.14, 0.18	−0.02, −0.2, −0.06, −0.2, −0.04	S2, S3, S4, S5	−0.20, −0.06, −0.2, −0.04
	S8, S6, L4, L2	0.2, 0.16, 0.14, 0.18	0.12, −0.04, −0.22, −0.06, −0.20	S6, L2, L4, L5	0.16, 0.18, 0.14, −0.20
	S1, L2, L3, L4	0.2, 0.16, 0.14, 0.18	0.68, 0.48, 0.32, 0.18, 0.	S1, S2, S3, S4	0.68, 0.48, 0.32, 0.18

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

spanning tree and introduces a gross error 0.3 in stream 6, one obtains the following gross errors: {0.00, − 0.28, − 0.34, 0.01, − 0.01}. If one applies the criterion for the threshold, one obtains the initial suspect set as {$S2$, $S3$} with sizes {− 0.28, − 0.34}. Obviously, all gross errors lower than the threshold will be missed.

The criterion also has effect on the occurrence of quasi-degenerate cases. Assume the gross errors introduced in the above example are {$S5$, $S8$} with {0.4, 0.3}. The spanning tree identifies: {$S1$, $S2$, $S3$, $S4$, $S5$} with {− 0.35, − 0.33, 0.01, − 0.34, −0.36}. After applying the criterion, one gets {$S1$, $S2$, $S4$, $S5$} with {− 0.35, − 0.33, −0.34, − 0.36}. Quasi-degeneracy happened when it's not supposed to happen.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

Gross Errors in Streams with Smaller Measurements

As stated above, a gross error should be considered only when its size is above a certain threshold. In reality, however, there are special situations where one may miss a gross error even when its size is really larger than the threshold. This problem usually happens because of the wide range in measurement sizes.

We illustrate this problem with the process in Figure 4. Suppose the flowrate measurements for each stream are $x^+ = [1000, 1010, 990, 10, 10, 10]$ and the standard deviation for each variable is 2% of its measurement.

Assume the measurements follow a multivariate normal distribution. If one wants to identify all biases with sizes larger than two times of their corresponding standard deviations, the confidence level for each stream at that time is 95.45%. Now introduce a bias of 0.5 to S_4, which is 2.5 times of its corresponding standard deviation. Table VIII shows that all available statistic tests fail to detect it.

We now show how the new strategy can have the same type of gross error masking problems.

Assume the spanning tree identified for the example of Figure 4 above is {$S1$, $S2$, $S5$}. The gross error sizes when using this spanning tree in a compensation model can be obtained as {0.0, − 0.5, − 0.5}. If the size criterion is selected as two times of the smallest standard deviation involved,

TABLE VIII An example for failed detection

	GT	*MT*	*NT*	*GLR*	*PCA*
Largest Test Statistic	3.13	1.77	1.77	3.13	1.77
Critical Value	3.84	2.0	2.0	3.84	2.0

i.e., 0.4 in this case. Then the new initial set is $\{S2, S5\}$ and the sizes are $\{-0.5, 0.5\}$. The MILP-based strategy can successfully identify the real gross error $S4$ with size of 0.5. Now suppose that one introduces two gross errors in $\{S4, S5\}$ with sizes $\{0.5, 0.5\}$ and uses the same spanning tree. One obtains the initial solution as $\{S1, S2, S5\}$ with $\{0.0, -0.5, 0.0\}$. Applying the same size criterion to this case, one will find that no one in suspect for this case since the gross error in $S2$ is less than two times of its standard deviation.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010

CONCLUSIONS

A new strategy based on MILP for detecting, identifying, and estimating gross errors (measurement biases and leaks) in linear steady state processes has been described in this paper. This strategy has the advantage that the detection, identification, and estimation of gross errors are performed simultaneously without using any statistic tests. The paper reinforces the notion that multiple sets of equivalent gross errors can exist, that this is independent of the detection method used and that without additional information (like field information for example) it is not possible to determine which set correspond the exact location of these gross errors. The implications for inventory losses and production accounting are therefore significant.

Acknowledgment

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

NOTATION

- A constraint matrix
- E a matrix representing the leaks in the constraints
- Q covariance matrix of random measurement errors
- U large number
- x measured variables
- x^+ measured values
- Y binary variable
- Z binary variable

Greek Symbols

ε small number
δ vector of sizes of biases
μ vector of sizes of leaks

References

Almasy, G. A. and Sztano, T. (1975) "Checking and Correction of Measurements on the Basis of Linear System Model", *Problems of Control and Information Theory*, **4**, 57.

Bagajewicz, M., "Optimal Sensor Location in Process Plants", *AIChE Journal*, **43**(9), 2300, September 1997.

Bagajewicz, M. and Manousiouthakis, V. (1992) "On the Mass/Heat Exchanger Network Representations of Distillation Networks", *AIChE J.*, **38**(11), 1769–1800.

Bagajewicz, M. and Jiang, Q. (1998) "Gross Error Modelling and Detection in Plant Linear Dynamic Reconciliation", *Computers and Chemical Engineering*, **22**(12), 1789–1810.

Bagajewicz, M., Jiang, Q. and Sánchez, M. (1999) "*Removing Singularities and Assessing Uncertainties in Two Efficient Gross Error Collective Compensation Methods*", To appear in Chem. Eng. Communications.

Crowe, C. M., Garcia Campos, Y. A. and Hrymak, A. (1983) "Reconciliation of Process Flow Rates by Matrix Projection Part I: Linear Case", *AIChE J.*, **29**, 881–888.

Jiang, Q. and Bagajewicz, M. (1999) "On a Strategy of Serial Identification with Collective Compensation for Multiple Gross Error Estimation in Linear Data Reconciliation", *Ind. & Eng. Chem. Research*, **38**(5), 2119–2128.

Keller, J. Y., Darouach, M. and Krzakala, G. (1994) "Fault detection of Multiple Biases or process leaks in Linear Steady State Systems", *Comp. & Chem. Eng.*, **18**, 1001.

Madron, F. V., Veverka, V. and Vanecek, V. (1977) "Statistical Analysis of Material Balance of a Chemical Reactor", *AIChE J.*, **23**, 482.

Madron, F. (1992) "Process Plant Performance. Measurement and Data Processing for Optimization and Retrofits", Ellis Horwood Ltd., Chichester, England.

Mah, R. S. H. (1990) "Chemical Process Structures and Information Flows", Butterworths.

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

Reilly, P. M. and Carpani, R. E. (1963) "Application of Statistical Theory of Adjustment to Material Balances", *13th Can. Chem. Eng. Conf.*, Montreal, Que.

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

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

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.

Sánchez, M. and Romagnoli, J. (1996) Use of Orthogonal Transformations in Classification/Data Reconciliation, *Comp. & Chem. Engng.*, **20**, 483–493.

Sánchez, M., Romagnoli, J., Jiang, Q. and Bagajewicz, M. (1999) "Simultaneous Estimation of Biases and Leaks in Process Plants", *Comp. & Chem. Engng.*, **23**(7), 841–858.

Seshu, S. and Reed, M. B. (1961) "Linear Graphs and Electrical Networks", Addison-Wesley Publishing Co., Reading, Mass.

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

Swartz, C. L. E. (1989) Data Reconciliation for Generalized Flowsheet Applications. American Chemical Society of National Meeting. Dallas, TX.

Downloaded By: [University of Okla Libraries] At: 20:40 27 August 2010