Generalized Likelihood Ratio Method for Gross Error Identification

A new method for detecting, identifying, and estimating gross errors in steady state processes is described in this paper. The generalized likelihood ratio method is based on the likelihood ratio statistical test and provides a general framework for identifying any type of gross error that can be modeled. The procedure is illustrated with gross errors caused by measurement biases and leaks. One significant advantage of the method is that the identification of gross errors is not confounded by departure from steady state conditions, which may now be accounted for by "leaks." Also proposed is a new strategy for identifying multiple gross errors using serial compensation of gross errors, which may be applied to all types of gross errors including leaks and biases and which requires less computing time than the existing strategies.

S. Narasimhan, R. S. H. Mah
Department of Chemical Engineering
Northwestern University
Evanston, IL 60201

Introduction

Gross error detection in steady state chemical processes has received considerable attention in the past 15 years and several statistical tests have been developed (see Tamhane and Mah, 1985, for a recent review). All methods for gross error identification make use of one or more statistical tests in combination with an identification strategy. The most recent methods are (DMT), and (EMT) (Rosenberg, 1985), (IMT) and (SC) (Serth and Heenan, 1986). All of these methods are designed to detect and identify only biases in measuring instruments. However, it is also important to detect and differentiate other types of gross errors (Mah, 1982). Without such capabilities even departure from steady state conditions could cause mistaken identification of gross errors. The main drawback of the above methods is that they do not utilize information regarding the effect of each type of gross error on the process, and hence they cannot differentiate between various types of gross errors.

In this paper we first describe the application of a new statistical approach based on the likelihood ratio test for identifying measurement biases and process leaks in steady state chemical processes. This approach was developed by Willsky and Jones (1974) to identify abrupt failures in dynamic systems and is referred to as the generalized likelihood ratio (GLR) method. In order to use this approach for our purpose, we develop a mathematical model that describes the effect of a leak and/or a bias on the process. Although we restrict our considerations to identifying measurement biases and process leaks in steady state processes, this approach provides the framework for identifying any type of gross error that can be mathematically modeled. In addition to the generality of this method, we also obtain as a by-product an estimate of the magnitude of the gross error, which is useful in judging the impact of the gross error on the process.

In order to detect and identify multiple gross errors, a serial strategy is often used (Romagnoli and Stephanopoulos, 1981; Rosenberg, 1985; Serth and Heenan, 1986). Typically, in the serial elimination strategy measurements suspected of containing gross errors are eliminated, the test statistic is recomputed, and if its value is now below the critical value, then the suspected measurements are declared in gross error. Such a procedure exploits the association of a gross error with a measurement. It is not applicable to gross errors that are not directly associated with measurements, for instance, a leak, since it assumes that eliminated measurements are corrupted by gross errors. We propose a new strategy based on serial compensation of gross errors to be used with the GLR technique. In this strategy we identify one gross error at a time and compensate for it using its estimated magnitude before identifying more gross errors. In contrast to the previous strategies, the serial compensation strategy can be used to identify multiple gross errors of any type.

Recent simulation studies (Iordache et al., 1985; Rosenberg et al., 1987) indicate that methods based on the measurement test (Mah and Tamhane, 1982) give the best performance for identifying measurement biases. We first show theoretically that for the special case when measurement biases only are considered, the GLR test reduces to the measurement test (see

Correspondence concerning this paper should be addressed to R. S. H. Mah.

appendix B), and therefore both approaches will give the same performance under identical assumptions. We also show through simulation studies of a steam metering process (Serth and Heenan, 1986) that the serial compensation strategy is computationally more efficient than serial elimination for identifying multiple gross errors.

In conclusion, the GLR method is preferred over existing approaches for gross error identification in steady state chemical processes, since it alone can differentiate between various types of gross errors. This capability is important, since we cannot always be sure that the process operates under strictly steady state conditions. For identifying multiple gross errors, the serial compensation strategy in combination with the GLR method is recommended.

Process Model

The steady state model of a chemical process in the absence of gross errors can be described by

$$\underline{z} = \underline{x} + \underline{v} \tag{1}$$

$$\underline{\underline{A}}\,\underline{x} = \underline{0} \tag{2}$$

Equation 1 is the measurement model where $\underline{z}$: $n \times 1$ is a measurement vector, $\underline{x}$: $n \times 1$ is a vector of true values of state variables and $\underline{v}$: $n \times 1$ is a vector of measurement errors. The measurement errors are assumed to be normally distributed with mean $\underline{0}$ and known covariance matrix $\underline{\underline{Q}}$. Equation 2 describes the linear or linearized mass and energy conservation constraints where $\underline{\underline{A}}$: $m \times n$ is a known constraint matrix. More general steady state process models (described by Eqs. 1 and 2 in Iordache et al., 1985) can be treated similarly after using appropriate transformations.

Measurement bias model

The model for a bias of unknown magnitude b in measuring instrument i is given by

$$\underline{z} = \underline{x} + \underline{v} + b\underline{e}_i \tag{3}$$

where $\underline{e}_i$ is a vector with unity in position i and zero elsewhere.

Process leak model

The presence of a leak affects the balance constraints. A mass flow leak in process unit (node) j of unknown magnitude b can be modeled by

$$\underline{\underline{A}}\,\underline{x} - b\underline{m}_j = \underline{0} \tag{4}$$

The elements of vector $\underline{m}_j$ are chosen as follows:

(a) Corresponding to a total mass flow constraint associated with node j, the element in $\underline{m}_j$ is chosen to have a value of unity.

(b) Corresponding to an energy flow constraint associated with node j, the value of the element in $\underline{m}_j$ is chosen as an average of the enthalpy of the streams incident on node j. In other words, we have assumed that the energy flow loss due to a mass flow leak in a node is an average of the enthalpies of the streams incident on that node multiplied by the magnitude of the leak. The preceding two statements are also applied to component flow constraints associated with node j, by replacing energy with component and enthalpy with concentration.

(c) The elements in $\underline{m}_j$ corresponding to constraints not associated with node j are chosen to be zero.

Pure energy or component flow losses in node j can also be modeled by Eq. 4 by choosing the corresponding element in $\underline{m}_j$ to be unity and all other elements to be zero.

The choice of the elements of $\underline{m}_j$ requires careful consideration. Strictly speaking, if a leak occurs in a stream (between the stream measurement and the process unit), then the composition and the enthalpy of that stream should be used. If a node is associated with several streams, each of a different composition, then potentially the composition and the enthalpy may be different for leaks associated with the different streams and the node. One might be tempted to introduce more gross error models, one for each situation. But this proliferation will clearly impair the diagnostic ability of the method. A preferred alternative is that we exercise good engineering judgment and introduce these models parsimoniously. A leak model should be introduced only if a leak is judged to be a distinct possibility. The selection rules given above represent an approximation in the absence of any other information. If precise values of enthalpies and concentrations are not available, nominal (design) values of these variables may be used in constructing the model for the leak. In this way we make effective use of available information and reduce the number of parameters to be estimated by the GLR method. Finally, for a physical leak we may choose to allow b in Eq. 4 to have a nonnegative value only, and interpret any negative value as a reflection of unsteady state process behavior.

The GLR Method

Procedure for single gross error

Given measurements, $\underline{z}$, we would like to determine the presence of each gross error, if any is present, and identify its source or cause. We shall first consider the case when at most one gross error is present.

Let the balance residuals be given by

$$\underline{r} = \underline{\underline{A}}\,\underline{z} \tag{5}$$

Since $\underline{r}$ is a linear transformation of $\underline{z}$, it has a multivariate normal distribution.

If no gross errors are present, we can show that

$$E[\underline{r}] = \underline{0} \tag{6}$$

$$Cov[\underline{r}] = \underline{\underline{V}} = \underline{\underline{A}}\,\underline{\underline{Q}}\,\underline{\underline{A}}' \tag{7}$$

If a gross error due to bias of magnitude b is present in measurement i, then by using Eqs. 2, 3, and 5 we can show that

$$E[\underline{r}] = b\underline{\underline{A}}\,\underline{e}_i \tag{8}$$

If a gross error due to a process leak of magnitude b is present in node j, then by using Eqs. 1, 4, and 5 we can show that

$$E[\underline{r}] = b\underline{m}_j \tag{9}$$

Therefore, when a gross error due to a bias or a process leak is

present, we can write

$$E[\underline{r}] = b\underline{f}_i \tag{10}$$

where

$$\underline{f}_i = \begin{cases} \underline{Ae}_i & \text{for a bias in measurement } i \\ \underline{m}_j & \text{for a process leak in node } j \end{cases} \tag{11}$$

Therefore, if we define $\underline{\mu}$ as the unknown expected value of $\underline{r}$, we can formulate the hypotheses for gross error detection as

$$\begin{aligned} H_0&: \underline{\mu} = \underline{0} \\ H_1&: \underline{\mu} = b\underline{f}_i \end{aligned} \tag{12}$$

where H_0 is the null hypothesis that no gross errors are present and H_1 is the alternative hypothesis that either a process leak or a measurement bias is present. The alternative hypothesis has two unknown parameters, b and $\underline{f}_i$. The parameter b can be any real number and the parameter $\underline{f}_i$ can be any vector from the set F, which is given by

$$F = \{\underline{Ae}_i, \underline{m}_j; i = 1 \ldots n, j = 1 \ldots m\} \tag{13}$$

Henceforth, we will refer to the vectors $\underline{f}_i$ as gross error vectors.

In order to test the hypotheses given by Eqs. 12 and estimate the unknown parameters if H_1 is accepted, we make use of the likelihood ratio test (Bickel and Doksum, 1977). The likelihood ratio test statistic in our case is given by

$$\lambda = \sup \frac{Pr\{\underline{r}|H_1\}}{Pr\{\underline{r}|H_0\}} \tag{14}$$

where the supremum in Eq. 14 is computed over all possible values of the parameters present in the hypotheses.

Using the normal probability density function for $\underline{r}$, we can write Eq. 14 as

$$\lambda = \sup_{b,\underline{f}_i} \frac{\exp\{-0.5(\underline{r} - b\underline{f}_i)'\underline{V}^{-1}(\underline{r} - b\underline{f}_i)\}}{\exp\{-0.5\underline{r}'\underline{V}^{-1}\underline{r}\}} \tag{15}$$

Since the expression on the righthand side of Eq. 15 is always positive, we can simplify the calculation by choosing as the test statistic

$$T = 2 \ln \lambda = \sup_{b,\underline{f}_i} \underline{r}'\underline{V}^{-1}\underline{r} - (\underline{r} - b\underline{f}_i)'\underline{V}^{-1}(\underline{r} - b\underline{f}_i) \tag{16}$$

The computation of T proceeds as follows. For any vector $\underline{f}_i$, we compute the estimate $\hat{b}$ of b, which gives the supremum in Eq. 16. Thus we obtain the maximum likelihood estimate

$$\hat{b} = (\underline{f}_i'\underline{V}^{-1}\underline{f}_i)^{-1}(\underline{f}_i'\underline{V}^{-1}\underline{r}) \tag{17}$$

Substituting this value of $\hat{b}$ in Eq. 16 and denoting the corresponding value of T by T_i we get

$$T_i = d_i^2/C_i \tag{18}$$

where

$$d_i = \underline{f}_i'\underline{V}^{-1}\underline{r} \tag{19}$$

$$C_i = \underline{f}_i'\underline{V}^{-1}\underline{f}_i \tag{20}$$

This calculation is performed for every vector $\underline{f}_i$ in set F and the test statistic T is therefore obtained as

$$T = \sup_i T_i \tag{21}$$

Let $\underline{f}^*$ be the vector that leads to the supremum in Eq. 21. The test statistic T is compared with a prespecified threshold ϵ, and a gross error is detected if T exceeds ϵ. The gross error that corresponds to the vector $\underline{f}^*$ is identified as the gross error and its magnitude is estimated from Eq. 17 using $\underline{f}^*$ for $\underline{f}_i$.

The method described above is called the generalized likelihood ratio (GLR) method (Willsky and Jones, 1974). It should be noted that in the above formulation we have not imposed the additional constraint that the magnitude b should be nonnegative for process leaks. Instead, we explicitly take this constraint into account as follows. If the estimated magnitude for a process leak is negative, we disregard that process leak as a possible source of gross error. In other words, when we take the supremum over all gross errors in Eq. 21, we disregard all process leaks for which the estimated magnitudes are negative.

In general, upper bounds on the magnitudes of leaks and upper and lower bounds on the magnitudes of biases may also be available. These can be taken into account in a similar way.

Selection of significance level

It is easily seen that under H_0

$$d_i \sim N(0, C_i) \tag{22}$$

Therefore, T_i has a central chi-square distribution with one degree of freedom under H_0. Since the T_i's are not independent, the distribution of T cannot be obtained. However, we can use arguments similar to those used by Mah and Tamhane (1982) and choose as the critical value $\chi^2_{1,1-\beta}$, the upper $1 - \beta$ quantile of the chi-square distribution. For a given level of significance α, the value of β is computed as

$$\beta = 1 - (1 - \alpha)^{1/p} \tag{23}$$

where p is the number of gross errors hypothesized (cardinality of set F). This would ensure that the probability of type I error is less than or equal to α.

Strategy for multiple gross errors

The treatment above pertains to the detection and identification of a single gross error. For multiple gross errors we propose a new strategy based on serial compensation of gross errors. In this strategy, we identify one gross error at a time by applying the GLR test. If a gross error is identified, we use its estimated magnitude to compensate for the gross error. We repeat this process until no further gross errors are detected. It should be noted that this strategy can be applied to all types of gross errors. A mathematical description of the serial compensation method is given in appendix A. The algorithm for implementing

the serial compensation strategy with the GLR test is outlined in Figure 1.

Strictly speaking, the serial procedure described above should be applied only if the detection of a gross error is not affected by the presence of any other gross error. If this condition is not true, then a more correct procedure is to apply the test to all postulated combinations of gross errors. That is, one at a time, two at a time, and so on. The drawback of the combinatorial approach is the escalated dimensionality of the problem and the concomitant increase in computational requirements. For this reason we propose to evaluate the performance of the serial compensation strategy by simulation experiments.

Performance Evaluation Using Simulation

Simulation procedure

Given a process network and associated constraints, and the covariance matrix of measurement errors $\underline{Q}$, the measurements are simulated using random numbers from the standard normal distribution. The method for simulating measurements, when no gross errors are present, or when a bias of given magnitude is present in some measurement, is described clearly by Iordache et al. (1985).

In the case when no process leaks are present, we use the simulated measurements $\underline{z}$ to compute the balance residuals from Eq. 5. If a leak of magnitude b in node i is simulated, then we compute the balance residuals as

$$\underline{r} = \underline{A}\underline{z} + b\underline{m}_i \tag{24}$$

The balance residuals are used by the GLR method to detect and identify gross errors.

The simulation can be carried out for a gross error of given type, location, and magnitude to study the performance of the method for identifying a specific gross error, or the number, type, and magnitudes of the gross errors can be chosen randomly, as in Serth and Heenan (1986). Once the appropriate conditions are specified, a simulation run consisting of 10,000 simulation trials (Iordache et al., 1985) is made. In each simulation trial, a different set of measurements is randomly generated. For ease of cross reference, the simulation runs are numbered such that the first digit is the same as the number of the figure that shows the corresponding process network. The performance of the GLR method is evaluated in the next two sections.

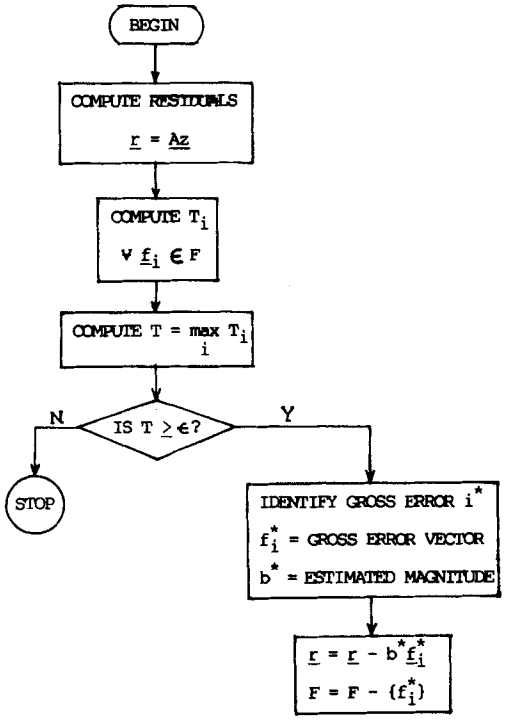

Figure 1. GLR test with serial compensation strategy.

Performance measures

Rosenberg (1985) proposed several measures that may be used to evaluate the performance of gross error identification schemes when a single gross error is present but multiple gross errors may be predicted or declared. These measures may be suitably modified to allow for the occurrence of multiple gross errors. In our case the following two performance measures are used:

1. The overall power of the method to identify gross errors correctly is given by

$$\text{Overall power} = \frac{\text{Number of gross errors correctly identified}}{\text{Number of gross errors simulated}} \tag{25}$$

The overall power is computed only for simulation runs in which gross errors are simulated (Serth and Heenan, 1986).

2. The average number of type I errors (AVTI), which defines the number of misidentifications made by a method, is given by

$$\text{AVTI} = \frac{\text{Number of gross errors wrongly identified}}{\text{Number of simulation trials made}} \tag{26}$$

The measure AVTI is computed separately for each simulation run, whether or not gross errors are simulated. This measure gives the average number of gross errors mispredicted per application of the method. AVTI under H_0 is set to the same value when different schemes are compared.

Results and Discussion

The first set of simulation runs was performed on the process network shown in Figure 2. In these simulation runs, a single specified gross error is simulated and the GLR method applied as described for identifying at most one gross error. These simulation runs give an indication of the ability of the GLR method to detect measurement biases and process leaks. The results of the simulation that are relevant in deriving useful conclusions are presented in Table 1.

In Table 1, each run number corresponds to a simulation run. The gross error simulated and its true magnitude in each case are given in the second and third columns, respectively. The computed overall power and AVTI under H_1 for each case are listed in the fourth and fifth columns, respectively. The estimated magnitude of the gross error, averaged over all trials in which the gross error is correctly identified, as shown in the last column.

For these simulation runs, all the variables shown in Figure 2

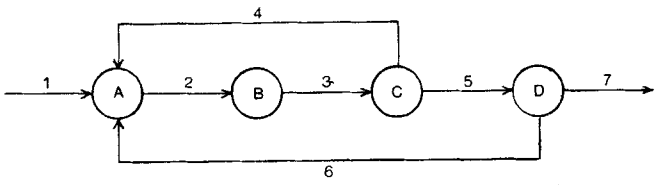

Figure 2. Recycle process network.

are assumed to be measured. The constraint matrix used is shown in Table 2A. For runs 2.1, 2.2, 2.3, and 2.4 only the first four rows of the matrix, corresponding to total mass balances in the four nodes, are used. For run 2.5 the last two rows of the matrix, corresponding to component and energy balances in node A, are also used. The gross error vectors for measurement biases and process leaks used in the simulation runs are shown in Table 2B. It should be noted that the dimensionality of the gross error vectors equals the number of rows in the constraint matrix used in each run. In all the simulation runs, $\underline{Q}$ is taken to be an identity matrix, and a level of significance α equal to 0.14 is used. This level of significance gives an AVTI approximately equal to 0.1 under the null hypothesis (when no gross errors are present).

In run 2.1 we assume that one gross error due to a bias in one of the seven measurements can occur. In runs 2.2 and 2.3, process leaks in nodes B and C are also included as gross error possibilities. In other words, the set F contains seven vectors for run 2.1 and nine vectors for runs 2.2 and 2.3. Comparing the results of runs 2.1 and 2.2 for the simulation of a bias in measurement 1, we observe that the inclusion of two additional gross error possibilities leads to a decrease in the power (8%) and a corresponding increase in AVTI. Thus, if we want to identify many different gross errors, we should use additional constraints or measurements in order to obtain good performance. Run 2.3 shows that the GLR method has the ability to identify and distinguish a leak in node B from measurement biases.

In runs 2.4 and 2.5, in addition to biases in the seven measurements and leaks in nodes B and C, a leak in node A is included as a gross error possibility. The results of run 2.4 show that the AVTI is very high for identifying a leak in node A. This is due to the fact that the gross error vectors for a bias in measurement 1 and for the leak in node A are identical, as observed from column three of Table 2B. The GLR method cannot distinguish between gross errors with identical gross error vectors. However, by adding other available constraints for node A, we can make the gross error vector for a bias in measurement 1 different from that of a process leak in node A, as observed from column four of Table 2B. Comparing the results of runs 2.4 and 2.5, we observe that the addition of the energy and component balances for node A has considerably increased the power for identifying the leak in node A and has also lowered the AVTI. Thus we can make better use of the GLR method by including energy, component, and other constraints in addition to total mass balances. The accuracy in the estimated magnitude of the gross error increases with the power of the GLR method. Runs 2.1, 2.2, 2.3, and 2.5 show that the error in the estimate of the gross error is less than 6%, if the power of the GLR test is greater than 0.8.

Table 1. Simulation Results of Runs Associated with Figure 2

Run No.	Gross Error Simulated	Magnitude of Gross Error	Overall Power	AVTI	Est. Magnitude of Gross Error
2.1	Bias in sensor 1	5.0	0.90	0.06	5.19
2.2	Bias in sensor 1	5.0	0.82	0.14	5.27
2.3	Leak in node B	4.25	0.88	0.11	4.37
2.4	Leak in node A	3.0	0.36	0.57	4.15
2.5	Leak in node A	3.0	0.93	0.07	3.02

Table 2A. Constraint Matrix for Process Network in Figure 2

Type of Constraint	Associated Node	Streams 1	2	3	4	5	6	7
Mass balance	A	1	−1	0	1	0	1	0
Mass balance	B	0	1	−1	0	0	0	0
Mass balance	C	0	0	1	−1	−1	0	0
Mass balance	D	0	0	0	0	1	−1	−1
Component balance	A	0.7	−1.0	0	0.1	0	0.2	0
Energy balance	A	125	−300	0	100	0	75	0

Table 2B. Gross Error Vectors Used in Runs Associated with Figure 2

Gross Error Type	Associated Node or Measurement	Gross Error Vectors f_i' Runs 2.1–2.4	Run 2.5
Sensor biases	1	[1 0 0 0]	[1 0 0 0 0.7 125]
	2	[−1 1 0 0]	[−1 1 0 0 −1 −300]
	3	[0 −1 1 0]	[0 −1 1 0 0 0]
	4	[1 0 −1 0]	[1 0 −1 0 0.1 100]
	5	[0 0 −1 1]	[0 0 −1 1 0 0]
	6	[1 0 0 −1]	[0 0 0 −1 0.2 75]
	7	[0 0 0 −1]	[0 0 0 −1 0 0]
Process leaks	A	[1 0 0 0]	[1 0 0 0 1 300]
	B	[0 1 0 0]	[0 1 0 0 0 0]
	C	[0 0 1 0]	[0 0 1 0 0 0]

Comparison of serial compensation and serial elimination

The second set of simulation runs was performed on the steam metering process network (Serth and Heenan, 1986) shown in Figure 3. The purpose of this simulation is to compare the performance of serial compensation with the serial elimination strategy for identifying multiple gross errors. The GLR test, which uses serial compensation of the gross error, is compared with the iterative measurement test (IMT) proposed by Serth and Heenan (1986). The IMT is chosen because it is based on the measurement test (MT) and uses the serial elimination strategy for identifying measurement biases. For this comparison, only measurement biases are simulated, since the IMT can only identify them. Note that if only measurement biases are considered, then the GLR test reduces to MT. The proof is given in appendix B. Hence, if the same strategy is used, then both methods will give the same performance. Therefore, any difference in the performance of the methods may be entirely attributed to the different strategies that are used.

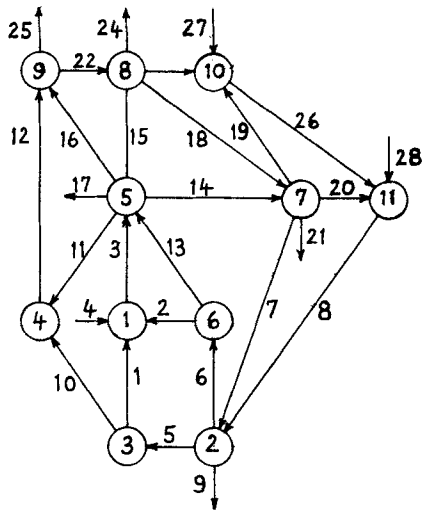

Figure 3. Steam-metering process network.

In the simulations, the number of gross errors to be simulated is fixed but the locations of the gross errors are randomly chosen. The magnitudes of the gross errors are also randomly generated between 5 and 15 times the standard deviations of the measurement errors. The constraint matrix and the matrix Q are the same as those used by Serth and Heenan. The case of no gross errors is first simulated and the two methods applied. The level of significance for each method is chosen such that it gives an AVTI under the null hypothesis approximately equal to 0.1. Thus the two methods are compared on the same basis. The results for the cases when 3, 5, or 7 gross errors are present are given in Table 3.

In Table 3, runs 3.1a, 3.2a, and 3.3a correspond to the GLR method and runs 3.1b, 3.2b, and 3.3b are those in which the IMT method is used. The same random number generator seeds are used in runs with identical numerical indices so that the same data are generated for both methods. The overall power and AVTI for each case are shown in columns three and four, respectively. Column five in this table gives the average percentage error reduction after data reconciliation, where the average is taken over all simulation trials. The percentage error reduction is given by $100*(E_1 - E_2)/E_1$, where E_1 is a measure of errors in the raw data,

$$E_1 = \sum_{i=1}^{n} |z_i - x_i| \tag{27}$$

Table 3. Simulation Results of Runs Associated with Figure 3

Run No.	No. Gross Errors Simulated	Overall Power	AVTI	% Error Reduction	Computation Time, s
3.1a	3	0.33	0.08	65.0	435
3.1b	3	0.33	0.09	64.6	1,574
3.2a	5	0.44	0.87	36.6	474
3.2b	5	0.44	0.72	39.9	2,368
3.3a	7	0.37	1.47	18.2	523
3.3b	7	0.37	1.19	20.1	2,748

Level of significance of GLR test = 0.15
Level of significance of IMT = 0.14

and E_2 is a measure of errors remaining after reconciliation

$$E_2 = \sum_{i=1}^{n} |\hat{z}_i - x_i| \tag{28}$$

and $\hat{z}_i$ are the elements of the estimates obtained by data reconciliation after the measurements identified as containing gross errors are eliminated or compensated, as the case may be. The percentage error reduction is a useful performance measure for applications involving data reconciliation. The last column in Table 3 shows the total computation time in seconds, taken for each run of 10,000 simulation trials. The computations are performed on a CDC Cyber 180/845.

The results for all three runs show that in terms of the measures used, both methods give almost the same performance. The AVTI increases for both methods as more gross errors are present. Runs 3.3a and 3.3b show that the IMT method gives a slightly lower AVTI than the GLR method when more gross errors are present. The most significant difference in the methods is in the computing time. The GLR method takes about one-fifth to one-half the time required by the IMT. This is due to the fact that serial elimination requires recomputation of the test statistics at each step of the serial procedure, which in turn requires the computation of inverses of different matrices, whereas in the case of serial compensation, the test statistics are obtained easily by updating the statistics from the previous step. Our implemention of IMT allows for nondiagonal Q, but is restricted to mass flow networks. If Q is diagonal, then IMT may be implemented more efficiently by the procedure described by Romagnoli and Stephanopoulos (1981). On the other hand, for general process networks with energy and component balances, the IMT will require the construction of projection matrices at each step to eliminate the measurements suspected of containing gross errors. A lengthier computation will be required. Based on these considerations and the simulation results, we conclude that the serial compensation strategy is much faster and performs almost as well. However, its most important advantage is its ability to handle all types of gross errors, which makes it a clearly superior choice.

A last simulation run is made on the steam metering network shown in Fig. 3 to demonstrate the practical utility of the GLR method for identifying biases and leaks. In addition to biases in the 28 measurements, leaks in nodes 3, 4, and 6 are considered as gross error possibilities. Leaks in other nodes are not considered since they cannot be distinguished from biases in streams connecting these nodes to the environment. The simulation is similar to that used for comparing serial elimination with serial compensation. The magnitude of the leaks in this simulation is randomly generated between 5 and 15. The results of the simulation are shown in Table 4. Columns two, three, four, and five in this table have the same interpretation as the corresponding columns of Table 3. The last column of Table 4 gives the per-

Table 4. Simulation of Steam-Metering Process Including Leaks

Run No.	No. Gross Errors Simulated	Overall Power	AVTI	% Error Reduction	% Leaks Identified
3.4a	3	0.36	0.13	58.0	67.0

centage of simulated leaks that are correctly identified. Comparing the results with run 1a of Table 3, in which leaks were excluded, we observe that the power and AVTI have increased (3 to 5%). More importantly, the last column of Table 4 shows that 67% of the simulated leaks are identified. Thus the GLR method is a useful practical tool for identifying leaks and biases.

Closing Remarks

Hitherto, a general framework for identifying gross errors other than measurement biases in steady state chemical processes was not available. The generalized likelihood ratio (GLR) method described in this study provides such a framework. In this paper the procedure is illustrated with gross errors arising from measurement biases and process leaks. But the procedure is equally applicable to other gross error models.

If the variables are bounded, the serial compensation strategy may be extended by checking for bound violations when data reconciliation is performed. If a bound on a variable is violated, then one possible strategy is to identify the gross error with the maximum test statistic (even if it is not rejected by the GLR test) and continue with the serial compensation procedure. This strategy is analogous to the dynamic measurement test (DMT) proposed by Rosenberg et al. (1987).

Acknowledgment

This work was supported by the National Science Foundation, Grant No. CPE 8115161, and by the Du Pont Educational Foundation.

Notation

$\underline{A}$ = process constraints matrix
$\underline{a}^*$ = vector, Eq. B2
a_i^* = ith element of $\underline{a}^*$
b = magnitude of gross error
$\hat{b}$ = estimated magnitude of gross error
b_k^* = estimated magnitude of gross error identified in application k of GLR test
$\underline{b}_k^*$ = vector of estimated magnitudes of gross errors, Eq. A9
C_i = quantity, Eq. 20
d_i = quantity, Eq. 19
E_1 = absolute sum of true errors in measurements, Eq. 27
E_2 = absolute sum of measurement adjustments, Eq. 28
$\underline{E}_k^*$ = matrix, Eq. A1
$\underline{e}_i$ = ith unit vector
F = set of gross error vectors
F_k = set of gross error vectors not identified in first k applications of GLR test
$\underline{G}_k^*$ = matrix, Eq. A5
$\underline{f}_i$ = gross error vector for gross error i
$\underline{f}^*$ = gross error vector that gives maximum GLR test statistic
$\underline{f}_k^*$ = gross error vector corresponding to gross error identified in application k of GLR test
$\underline{H}$ = matrix, Eq. B3
H_{ii} = ith diagonal element of $\underline{H}$
$\underline{M}_k^*$ = matrix, Eq. A3
$\underline{m}_i$ = gross error vector for leak in node i
m = number of constraints
n = number of measurements
p = number of gross errors hypothesized
$\underline{Q}$ = covariance matrix of measurement errors
$\underline{r}$ = vector of constraint residuals
$\underline{r}_k$ = vector of compensated constraint residuals after k applications of GLR test
T_i = GLR test statistic for gross error i
T = maximum GLR test statistic
$\underline{V}$ = covariance matrix of constraint residuals
$\underline{v}$ = vector of measurement errors
v_{ii} = ith diagonal element of $\underline{V}$
$\underline{x}$ = vector of true values of process variables
$\underline{z}$ = vector of measurements
$\underline{z}_k$ = vector of compensated measurements after k applications of GLR test
z_i = ith measurement
$\hat{z}_i$ = least-squares estimate of ith process variable
$\underline{0}$ = vector of zeros

Greek letters

α = level of significance
β = adjusted level of significance, Eq. 23
μ = expected value of balance residuals
λ = likelihood ratio, Eq. 14

Other symbols

a' = transpose of a
a^{-1} = inverse of a
$Cov[a]$ = covariance matrix of a
$E[a]$ = expected value of a
$Pr[a]$ = probability of a
$Pr[a|b]$ = conditional probability of a given b
$\chi^2_{1,1-\beta}$ = upper $1-\beta$ quantile of chi-square distribution with one degree of freedom

Appendix A: Serial Compensation Strategy

In the serial compensation strategy, the GLR test is applied one or more times depending on the number of gross errors identified. After each gross error is identified, the constraints or measurements are compensated. Let us denote the compensated measurements and residuals after k applications of the GLR test as $\underline{z}_k$ and $\underline{r}_k$, respectively. Let $\underline{f}_k^*$ and b_k^*, be the gross error vector and estimated magnitude, respectively, of the gross error that is identified in application k of the GLR test. Let the matrices $\underline{E}_k^*$, $\underline{M}_k^*$, and $\underline{G}_k^*$ be defined as follows:

$$\underline{E}_k^* = [\underline{e}_1^*, \underline{e}_2^* \cdots \underline{e}_k^*] \tag{A1}$$

where

$$\underline{e}_i^* = \begin{cases} \underline{0} & \text{if a bias is not identified in application } i \\ \underline{e}_j & \text{if a bias in sensor } j \text{ is identified in application } i \end{cases} \tag{A2}$$

$$\underline{M}_k^* = [\underline{m}_1^*, \underline{m}_2^* \cdots \underline{m}_k^*] \tag{A3}$$

where

$$\underline{m}_i^* = \begin{cases} \underline{0} & \text{if a leak is not identified in application } i \\ \underline{m}_j & \text{if a leak in node } j \text{ is identified in application } i \end{cases} \tag{A4}$$

$$\underline{G}_k^* = [\underline{f}_1^*, \underline{f}_2^* \cdots \underline{f}_k^*] \tag{A5}$$

It can be seen from the above definitions and the definition of the gross error vectors $\underline{f}_i$, that

$$\underline{G}_k^* = \underline{A}\underline{E}_k^* + \underline{M}_k^* \tag{A6}$$

The compensated measurements and constraints at the end of application k are then given by

$$\underline{z}_k = \underline{z} - \underline{E}_k^* \underline{b}_k^* \tag{A7}$$

$$\underline{A}\underline{x} - \underline{M}_k^* \underline{b}_k^* = 0 \tag{A8}$$

where

$$\underline{b}_k^* = [b_1^*, b_2^* \ldots b_k^*]' \tag{A9}$$

If we define the compensated residuals $\underline{r}_k$ as

$$\underline{r}_k = \underline{A}\underline{z}_k - \underline{M}_k^* \underline{b}_k^* \tag{A10}$$

then by using Eqs. A6, A7, and 5 we can see that

$$\underline{r}_k = \underline{r} - \underline{G}_k^* \underline{b}_k^* \tag{A11}$$

The hypotheses for application $k + 1$ of the GLR test can be formulated as

$$\begin{aligned} H_{0,k}&: E[\underline{r}_k] = \underline{0} \\ H_{1,k}&: E[\underline{r}_k] = b\underline{f}_i; \quad \underline{f}_i \in F_k \end{aligned} \tag{A12}$$

where F_k is the set of gross error vectors corresponding to the gross errors that are not identified in the first k applications of the GLR test. If we assume that the gross errors identified in the first k applications are actually present in the data and the actual magnitudes of these gross errors are equal to the estimated magnitudes, then under $H_{0,k}$ the true constraint model is given by Eq. A8 and the true measurement model is given by

$$\underline{z} = \underline{x} + \underline{E}_k^* \underline{b}_k^* + \underline{v} \tag{A13}$$

Then by using Eqs. 5, A6, A8, A11, and A13, we can easily show that

$$\underline{r}_k \sim N(\underline{0}, \underline{V}) \text{ under } H_{0,k} \tag{A14}$$

The test statistic for each gross error vector in F_k can therefore be calculated through Eqs. 18, 19, and 20 by using $\underline{r}_k$ for $\underline{r}$. It should be noted that the test statistics for application $k + 1$ are conditional test statistics, conditioned on the assumption that the gross errors and their estimated magnitudes identified in the first k applications are actually present in the data.

If a gross error is not detected, say in application n of the GLR test, then the serial compensation strategy is stopped. The compensated residuals $\underline{r}_{n-1}$ are used in data reconciliation to obtain the estimates of all variables.

Appendix B: Equivalence of GLR Test and Measurement Test

The gross error vector $\underline{f}_i$, for a bias in sensor i is equal to $\underline{A}\underline{e}_i$. Substituting for $\underline{f}_i$ in Eq. 17 and using Eqs. 18 and 19, we obtain the GLR test statistic for a bias in sensor i as

$$T_i = \frac{(\underline{e}_i' \underline{A}' \underline{V}^{-1} \underline{r})^2}{(\underline{e}_i' \underline{A}' \underline{V}^{-1} \underline{A} \underline{e}_i)} \tag{B1}$$

where $\underline{V}$ is given by Eq. 7.

Let us define

$$\underline{a}^* = \underline{A}'(\underline{A}\underline{Q}\underline{A}')^{-1}\underline{r} \tag{B2}$$

and

$$\underline{H} = \underline{A}'(\underline{A}\underline{Q}\underline{A}')^{-1}\underline{A} \tag{B3}$$

Then T_i is given by

$$T_i = (a_i^*)^2 / H_{ii} \tag{B4}$$

where a_i^* is the ith element of $\underline{a}^*$ and H_{ii} is the ith diagonal element of $\underline{H}$. But the righthand side of Eq. B4 is equal to the square of the measurement test (MT) statistic for sensor bias i. The MT statistic is given by Eq. 4.6 in Tamhane and Mah (1985), where the notation Ω is used for $\underline{V}$, $\underline{C}$ for $\underline{A}$, $\underline{y}$ for $\underline{z}$, and $\underline{d}$ for $\underline{0}$. Thus the GLR test statistic is the square of the MT statistic. If the same level of significance is used for both methods, then the GLR test criterion is also equal to the square of the MT criterion. Therefore, under the same assumptions the GLR test and the MT are equivalent for identifying sensor biases.

Literature cited

Bickel, P. J., and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco, 209 (1977).

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

Mah, R. S. H., "Design and Analysis of Performance Monitoring Systems," *Chemical Process Control II*, D. E. Seborg, T. F. Edgar, eds., Engineering Foundation, New York, 525 (1982).

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

Romagnoli, J. A., and G. Stephanopoulos, "Rectification of Process Measurement Data in the Presence of Gross Errors," *Chem. Eng. Sci.*, **36**(11), 1849 (1981).

Rosenberg, J., "An Evaluation of Strategies for Detecting and Identifying a Single Gross Error in Process Data," Northwestern Univ., Evanston, IL (1985).

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

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

Tamhane, A. C., and R. S. H. Mah, "Data Reconciliation and Gross Error Detection in Chemical Process Networks," *Technometrics*, **27**(4), 409 (1985).

Willsky, A. S., and H. L. Jones, "A Generalized Likelihood Ratio Approach to State Estimation in Linear Systems Subject to Abrupt Changes," *Proc. IEEE Conf. Decision and Control*, 846 (1974).

Manuscript received Dec. 1, 1986, and revision received Mar. 19, 1987.