AIChE WILEY InterScience DISCOVER SOMETHING GREAT

Downside Financial Loss of Sensor Networks in the Presence of Gross Errors

Duy Quang NguyenThanh and Kitipat Siemanond
The Petroleum and Petrochemical College, Chulalongkorn University, Bangkok 10330, Thailand

M. Bagajewicz
University of Oklahoma, Norman, OK 73019

DOI 10.1002/aic.10992
Published online September 14, 2006 in Wiley InterScience (www.interscience.wiley.com).

Theoretical expressions for assessing the expected financial loss associated with the accuracy of the instrumentation and the associated probability in the presence of two and more gross errors have been developed in previous work. However, these expressions were given in a form of integrals without closed-form solutions. This article presents a generalized method to calculate the expected financial loss and the associated probability when two or more gross errors are present in a system. The application of financial loss calculation in the design and retrofit of a sensor network is illustrated in a simple example. © 2006 American Institute of Chemical Engineers *AIChE J,* 52: 3825–3841, 2006
Keywords: instrumentation network design, data reconciliation, plant monitoring

Introduction

Errors in measurements lead to deterioration in plant performance, by deteriorating control performance, leading to off-spec streams, biased on-line optimization set points, and the under/overestimation of production. Data reconciliation—used together with gross error detection techniques—can be used to reduce variability and increase accuracy. However, the performance of data reconciliation relies on the appropriate instrumentation achieving the appropriate level of redundancy.

Recently, the accuracy of estimators was defined as the sum of the precision and the maximum undetectable induced bias in a stream.[1] In addition, a new approach to assess the economic value of precision was presented.[2,3] In a follow-up article, the economic value of accuracy was discussed.[4] Specifically, these articles developed expressions to obtain the expected financial loss associated with the precision and accuracy of the instrumentation; in this way, the gain in economic value (the decrease in financial loss) arising from the addition of an instrumentation can be determined. Bagajewicz[4] derived analytical expressions for the downside financial loss under the presence of one gross error, but he did not develop analytical forms of the integral expressions in the presence of more than one gross error, which consist of integrals with discontinuous integrands that do not reduce to a closed-form solution. Bagajewicz[5] also used Monte Carlo sampling to determine the software accuracy more efficiently. In this latter case, the expected software accuracy is introduced and calculated, whereas in the former[1] accuracy is defined more conservatively using an upper bound. Both definitions are valid and they are of importance for designers in different setups.

This article presents a method that allows the approximate calculation of the integral expressions proposed by Bagajewicz[4] in the presence of more than one gross error. We use and compare two techniques to accomplish this: a system of averaging upper and lower bounds of these integrals, which can be obtained analytically, and a Monte Carlo method to evaluate the integrals.

Background

Performing a statistical analysis, Bagajewicz et al.[3] were able to obtain expressions for assessing the economic value of

Correspondence concerning this article should be addressed to M. Bagajewicz at bagajewicz@ou.edu.

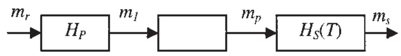

Figure 1. Simplified material balance in a refinery.

precision. A formula was developed to assess this economic value of precision based on the downside expected loss that occurs when an operator adjusts the throughput of a plant when the measurements or estimators obtained through data reconciliation suggest that the targeted production is met or surpassed. However, there is a finite probability that the measurement or estimator is above the target when, in fact, the real flow is below it—thus the expected financial loss calculation. The calculation procedure to obtain the formula is briefly reviewed next.

Bagajewicz et al.[3] argued that a typical refinery consists of several tank units that receive the crude, several processing units, and several tanks where products are stored, summarized in three blocks as in Figure 1, where for simplicity, one raw material and one product are assumed. In this figure m_r and m_s, the raw materials receiving and products delivering "streams" are discontinuous, whereas the other internal streams are continuous. Of importance here is the product tank hold up H_S, which needs to meet a certain target value H_S^* at certain time T.

They argued that the probability of not meeting the targeted production is $P\{H_S(T) \leq H_S^*\}$, which in turn can be rewritten given that the probability of the measured (or estimated) flow rate is smaller than the targeted value, that is $P\{m_p(t) \leq m_p^*\}$ for all $t \leq T$. This last equivalency requires assuming that the true value $m_p(t)$ oscillates around a mean value m_p. Let $\hat{m}_p$ be the estimator of the true value of m_p and consider that production is adjusted to meet the targeted value, based on the estimate. In other words, if $\hat{m}_p < m_p^*$, production is increased and vice versa, if $\hat{m}_p > m_p^*$, production is decreased. They assumed for simplicity (and without loss of generality) that, when $\hat{m}_p > m_p^*$, that is, the measurement indicates that the target has been met or exceeded, the operator would not do any correction to the set points. They argued that the probability of making the wrong decision is given by the conditional probability $P\{m_p \leq m_p^* \mid \hat{m}_p \geq m_p^*\}$, that is, the probability of having missed the target given that the estimator is larger than the target. A statistical analysis was performed to derive the following expression:

$$P(m_p \leq m_p^* | \hat{m}_p \geq m_p^*) = \int_{-\infty}^{m_p^*} \left\{ \int_{m_p^*}^{\infty} g_M(\xi, m_p, \hat{\sigma}_p) d\xi \right\} g_p(m_p, m_p^*, \sigma_p) dm_p \quad (1)$$

where $g_M(\xi, m_p, \hat{\sigma}_p)$ and $g_p(m_p, m_p^*, \sigma_p)$ are the measurement and true value distribution around their means, with standard deviations $\hat{\sigma}_p$ and σ_p, respectively. The downside expected financial loss (*DEFL*, loss incurred for not meeting the production target) was derived as follows:

$$DEFL(\hat{\sigma}_p, \sigma_p) = \int_{-\infty}^{m_p^*} g_p(m_p, m_p^*, \sigma_p) \left\{ K_s T \int_{-\infty}^{m_p^*} (m_p^* - \hat{m}_p) g_M(\hat{m}_p, m_p, \hat{\sigma}_p) d\hat{m}_p \right\} dm_p \quad (2)$$

where K_s is the value of the products sold (or the cost of inventory) and T is the period of time under consideration. Under simplified assumptions of negligible process variations (that is, $\sigma_p/\hat{\sigma}_p \ll 1$) and normal distributions assumed for the measurement $\hat{m}_p$ and the true value m_p, Bagajewicz[3] showed the probability to be 0.25 and Eq. 2 to reduce to $DEFL = 0.19947 K_s T \hat{\sigma}_p$.

Although precision is important, most instruments present biases. To understand how biases corrupt measurement accuracy, Bagajewicz[1] introduced the concept of software accuracy, which is based on the notion that data reconciliation with some test statistics is used to detect biases. If biases are too small to be detected, they smear all the estimators, including those of the variables for which the corresponding instruments have no bias, called *induced bias* ($\hat{\delta}$), which is defined as the difference between expected values of the estimators $E[\hat{x}]$ when gross errors are present and the true value x of process variables[1]:

$$\hat{\delta} = E[\hat{x}] - x = [I - SW]\delta \quad (3)$$

where $W = A^T(ASA^T)^{-1}A$ and S is the variance matrix of measurements, A is the incidence matrix, and δ is the gross errors vector.

Thus, the accuracy of an estimator is defined as the sum of precision of the estimator plus the maximum possible undetected induced bias in that variable arising from sensor biases somewhere in the system

$$\hat{a}_i = \hat{\sigma}_i + \delta_i^* \quad (4)$$

where $\hat{a}_i$, $\hat{\sigma}_i$, and δ_i^* are the accuracy, precision (square root of variance S_{ii}), and the maximum undetected induced bias of the estimator, respectively.

Using the above concept, Bagajewicz[4] developed the theory of economic value of accuracy. He assumed that, when an instrument fails, it happens with a certain probability $f_i(t)$ (a function of time), and that the size of the bias follows a certain distribution $h_i(\theta_i, \bar{\delta}_i, \rho_i)$ with mean $\bar{\delta}_i$ and variance ρ_i^2. Thus, one needs to integrate over all possible values of the gross errors and multiply by the probability of such bias to develop. If it is assumed that one instrument fails at a time, then the probability of n_b instruments failing and the others not is given by

$$\Phi_{i_1,i_2,\ldots,in_b}^{n_T} = f_{i1}(t) \cdots f_{in_b}(t) \prod_{s \neq i_1,\ldots,s \neq in_b} [1 - f_s(t)]$$

Thus, the probability of being wrong in the presence of n_b biases is given by a general expression as follows:

$$P\{\hat{m}_p \geq m_p^* | i_1 \cdots in_b\} = \Phi_{i_1 \cdots in_b}^{n_b} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \{P(\hat{m}_p \geq m_p^* | \theta_1 \cdots \theta_{n_b})\} \times h_1(\theta_1, \bar{\delta}_1, \rho_1) \cdots h_{n_b}(\theta_{n_b}, \bar{\delta}_{n_b}, \rho_{n_b}) d\theta_1 \cdots d\theta_{n_b} \quad (5)$$

where $P(\hat{m}_p \geq m_p^* | \theta_1, \ldots, \theta_{n_b})$ indicates the probability being conditional to the presence of n_b biases. The general expression for the financial loss *DEFL* can be developed from Eq. 2 in the same fashion, that is, one needs to integrate over all possible values of the gross errors. Detailed forms of the expressions for the probability and the financial loss in the presence of biases were given by Bagajewicz.[4] These expressions are broken down into a summation of terms, each one related to the presence and the number of undetected gross errors for particular values of gross errors. Therefore, these expressions are integrals that require integrating a discontinuous function that changes form in different regions of the integrand. For example, the expression for the financial loss *DEFL* in the presence of one bias is broken down into several terms as follows:

$$DEFL^1|i = \int_{-\infty}^{-\infty} \left(\int_{-\infty}^{m_p^*} \left\{ K_s T \int_{-\infty}^{m_p^*} (m_p^* - \xi) g_M(\xi, m_p, \hat{\sigma}_{p,i}^R) d\xi \right\} g_p(m_p, m_p^*, \sigma_p) dm_p \right) h_i(\theta, \delta_i, \rho_i) d\theta$$

$$= \left[\int_{-\infty}^{-\bar{\delta}_i^p} + \int_{\bar{\delta}_i^p}^{\infty} \right] \left(\int_{-\infty}^{m_p^*} \left\{ K_s T \int_{-\infty}^{m_p^*} (m_p^* - \xi) g_M(\xi, m_p, \hat{\sigma}_{p,i}^R) d\xi \right\} g_p(m_p, m_p^*, \sigma_p) dm_p \right) h_i(\theta, \delta_i, \rho_i) d\theta$$

$$+ \int_{-\bar{\delta}_i^p}^{\bar{\delta}_i^p} \left(\int_{-\infty}^{m_p^*} K_s T \left\{ \int_{-\infty}^{m_p^*} (m_p^* - \xi) g_M(\xi, m_p + \hat{\delta}_p^i(\theta_i), \hat{\sigma}_p) d\xi \right\} g_p(m_p, m_p^*, \sigma_p) dm_p \right) h_i(\theta; \bar{\delta}_i, \rho_i) d\theta \quad (6)$$

The first term corresponds to the detection of the gross error (when gross error magnitude is larger than the critical value); the second term corresponds to the case of undetected gross error.

Under simplified assumptions (negligible process variation and normal distributions), Bagajewicz[4] was able to derive an analytical form for the expressions in the presence of one bias, but did not provide analytical forms for the expressions for the presence of more than one bias, which are integrals with discontinuous integrands that do not reduce to a closed-form solution.

New Calculation Procedures

Under simplified assumptions (negligible process variation and normal distributions), the general expressions given by Bagajewicz[4] for the probability and the associated financial loss when two gross errors are present in the system reduce to

$$P = \Phi_{i_1,i_2}^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{4} \left\{ \begin{array}{ll} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) & \text{(both biases are detected)} \\ +[1 + \text{erf}(A_1)] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) & \text{(no bias is detected)} \\ +[1 + \text{erf}(A_2)] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) & \text{(only bias 1 is detected)} \\ +[1 + \text{erf}(A_3)] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) & \text{(only bias 2 is detected)} \end{array} \right\} d\theta_1 d\theta_2 \quad (7)$$

$$DEFL = \frac{K_s T}{2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left\{ \begin{array}{l} \frac{\hat{\sigma}_p}{\sqrt{2\pi}} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) \quad \text{(when both biases are detected)} \\ + \left(\frac{\hat{\sigma}_p e^{-A_1^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\, \hat{\sigma}_p A_1 \{1 - \text{erf}(A_1)\}}{2} \right) \left(\begin{array}{c} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) \\ \text{no bias is detected} \end{array} \right) \\ + \left(\frac{\hat{\sigma}_p e^{-A_2^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\, \hat{\sigma}_p A_2 \{1 - \text{erf}(A_2)\}}{2} \right) \left(\begin{array}{c} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) \\ \text{only bias 1 is detected} \end{array} \right) \\ + \left(\frac{\hat{\sigma}_p e^{-A_3^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\, \hat{\sigma}_p A_3 \{1 - \text{erf}(A_3)\}}{2} \right) \left(\begin{array}{c} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) \\ \text{only bias 2 is detected} \end{array} \right) \end{array} \right\} d\theta_1 d\theta_2 \quad (8)$$

where

$$h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) = \frac{e^{-(\theta_1 - \bar{\delta}_1)^2 / 2\rho_1^2}}{\rho_1 \sqrt{2\pi}} \frac{e^{-(\theta_2 - \bar{\delta}_2)^2 / 2\rho_2^2}}{\rho_2 \sqrt{2\pi}}$$

are the normal probability density functions of biases, and

$$A_1 = \frac{\alpha_1 \theta_1 + \alpha_2 \theta_2}{\sqrt{2}\, \hat{\sigma}_p} \qquad A_2 = \frac{\alpha_2 \theta_2}{\sqrt{2}\, \hat{\sigma}_p} \qquad A_3 = \frac{\alpha_1 \theta_1}{\sqrt{2}\, \hat{\sigma}_p}$$

The constants α_1 and α_2 are parameters that come from the expressions for the induced bias: $\hat{\delta}_p^{i_1 i_2}(\theta_1, \theta_2) = \alpha_1 \theta_1 + \alpha_2 \theta_2$; $\hat{\delta}_p^{i_1}(\theta_1) = \alpha_1 \theta_1$; $\hat{\delta}_p^{i_2}(\theta_2) = \alpha_2 \theta_2$, as developed by Bagajewicz.[4]

To evaluate the above integrals, one needs to identify the different regions according to the existence of undetected gross

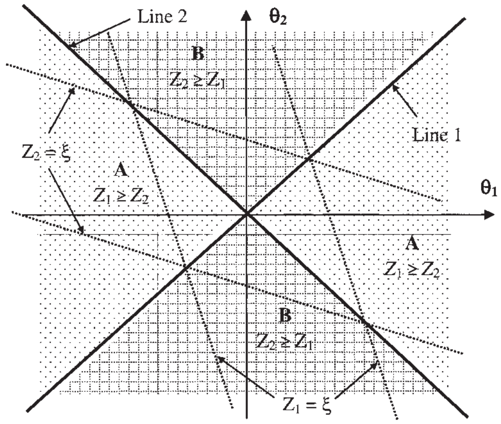

Figure 2. Identifying regions where $Z_1 \geq Z_2$ or vice versa $Z_2 \geq Z_1$.

errors in the system. If a serial elimination strategy is used to detect gross errors based on the maximum power MT test, the test statistics for the two measurements i_1 and i_2 are given by

$$Z_1^{MP} = \frac{|W_{i_1i_1}\theta_1 + W_{i_1i_2}\theta_2|}{\sqrt{W_{i_1i_1}}} \qquad Z_2^{MP} = \frac{|W_{i_2i_1}\theta_1 + W_{i_2i_2}\theta_2|}{\sqrt{W_{i_2i_2}}}$$

where $W = A^T(ASA^T)^{-1}A$. The threshold value for the MT test statistics is ξ (=1.96 at level of confidence of 95%).

In the serial elimination strategy, if one gross error has been detected, the corresponding measurement is eliminated. Which gross error is detected (or detected first) depends on which test statistic is greater (Z_1 or Z_2). Therefore, it is necessary to identify the regions where we have $Z_1 \geq Z_2$ or $Z_2 \geq Z_1$.

Now, $Z_1 = Z_2$ when

$$\frac{W_{i_1i_1}\theta_1 + W_{i_1i_2}\theta_2}{\sqrt{W_{i_1i_1}}} = \frac{W_{i_2i_1}\theta_1 + W_{i_2i_2}\theta_2}{\sqrt{W_{i_2i_2}}}$$

in quadrants I and III, or

$$\frac{(W_{i_1i_1}\theta_1 + W_{i_1i_2}\theta_2)}{\sqrt{W_{i_1i_1}}} = \frac{(W_{i_2i_1}\theta_1 + W_{i_2i_2}\theta_2)}{\sqrt{W_{i_2i_2}}}$$

in quadrants II and IV. Then we have two lines:

$$\left(\sqrt{W_{i_1i_1}} - \frac{W_{i_2i_1}}{\sqrt{W_{i_2i_2}}}\right)\theta_1 = \left(\sqrt{W_{i_2i_2}} - \frac{W_{i_1i_2}}{\sqrt{W_{i_1i_1}}}\right)\theta_2 \qquad \text{(line 1)}$$

and

$$\left(\sqrt{W_{i_1i_1}} + \frac{W_{i_2i_1}}{\sqrt{W_{i_2i_2}}}\right)\theta_1 = -\left(\sqrt{W_{i_2i_2}} + \frac{W_{i_1i_2}}{\sqrt{W_{i_1i_1}}}\right)\theta_2 \qquad \text{(line 2)}$$

These two lines divide the plane into two regions: region A, where we have $Z_1 \geq Z_2$ and region B, where we have $Z_2 \geq Z_1$ (Figure 2). In this figure, the lines for $Z_1 = \xi$ and $Z_2 = \xi$ are added.

The lines $Z_1 = \xi$ and $Z_2 = \xi$ define the rhombus shown in Figure 3. In this figure we assume that $W_{ij} > 0$. Inside the rhombus, we have $Z_1 \leq \xi$ and $Z_2 \leq \xi$, that is, neither of the two gross errors is detected because they are too small to make the MT flag positive. Outside the rhombus, at least one of the two test statistic Z_1, Z_2 is $>\xi$, then at least one gross error is detected.

Now consider the case where gross error θ_1 is detected first, that is, outside the rhombus and in region A where we have $Z_1 \geq Z_2 \geq \xi$ or $Z_1 \geq \xi \geq Z_2$. Following the serial elimination strategy, the corresponding measurement i_1 is eliminated. Then, at the next stage, the new test statistic for the measurement i_2 is $Z_2' = |W''_{i_2i_2}\theta_2|/\sqrt{W''_{i_2i_2}}$ (here, W'' is the updated matrix W after measurement i_1 has been eliminated). Therefore, gross error θ_2 is also detected when

$$Z_2'' = \frac{|W''_{i_2i_2}\theta_2|}{\sqrt{W''_{i_2i_2}}} \geq \xi \qquad \text{or} \qquad |\theta_2| \geq \frac{\xi}{\sqrt{W''_{i_2i_2}}} = \delta_2$$

(that is, both gross errors are detected and gross error θ_1 is detected first). As a consequence, only gross error θ_1 is detected when $|\theta_2| < \delta_2$. Therefore, the lines $|\theta_2| = \delta_2$ divide the regions under consideration (in region A and outside the rhombus, that is, gross error θ_1 is detected first) into two smaller regions: region A_1 where we have $|\theta_2| \geq \delta_2$, that is, both gross errors are detected; and region A_2, where we have $|\theta_2| < \delta_2$, that is, only gross error θ_1 is detected. These regions are depicted in Figure 4.

Similarly, if gross error θ_2 is detected first (outside the rhombus and in region B) and the corresponding measurement i_2 is eliminated, gross error θ_1 is also detected when

$$Z_1' = \frac{|W'_{i_1i_1}\theta_1|}{\sqrt{W'_{i_1i_1}}} \geq \xi \Rightarrow |\theta_1| \geq \frac{\xi}{\sqrt{W'_{i_1i_1}}} = \delta_1$$

(W' is the updated matrix W after measurement i_2 has been eliminated). Finally, with all these analyses, the regions corre-

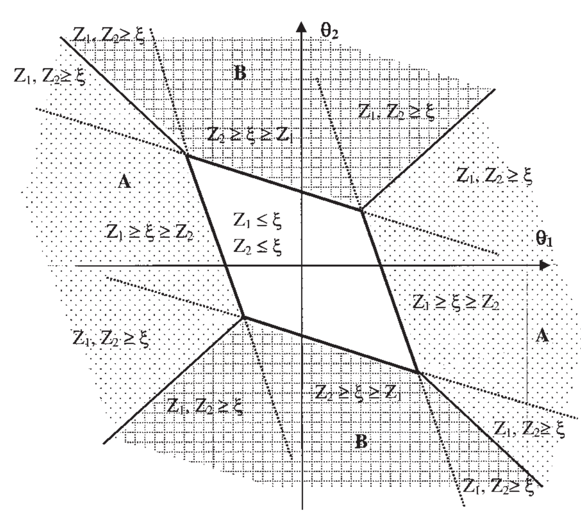

Figure 3. Identifying regions corresponding to values of Z_1 and Z_2.

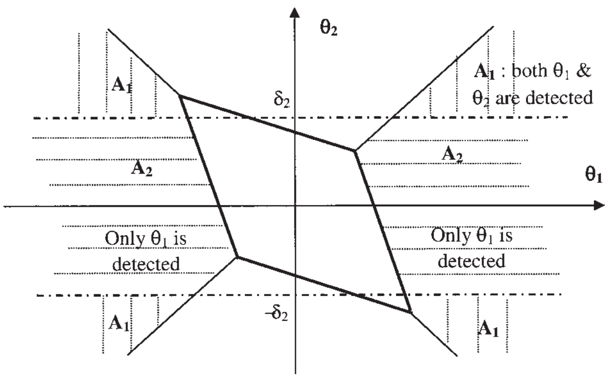

Figure 4. Identifying regions corresponding to the detection of θ_2 given θ_1 is detected.

sponding to the possibilities of the presence of undetected gross errors are shown in the Figure 5 for the case $W_{ij} > 0$.

Note that in Figure 5 we have

$$\delta_2 = \frac{\xi}{\sqrt{W''_{i_2i_2}}} \geq \frac{\xi}{\sqrt{W_{i_2i_2}}} \qquad \text{and} \qquad \delta_1 = \frac{\xi}{\sqrt{W'_{i_1i_1}}} \geq \frac{\xi}{\sqrt{W_{i_1i_1}}}$$

These expressions stem from the fact that $W_{i_1i_1} \geq W'_{i_1i_1}$ and $W_{i_2i_2} \geq W''_{i_2i_2}$ (see the Appendix).

We now describe two methodologies to calculate the probability of occurrence as well as the downside financial loss. We first describe the different regions more fully. Figure 6 depicts a square, outside which the integrals can be defined over semi-infinite regions with borders that are parallel to the axis. In these regions some of the integrals have analytical closed-form solutions, which is convenient. This rectangle is defined by lines $\theta_1 = \pm K_1$ and $\theta_2 = \pm K_2$, where K_1 and K_2 are defined by the vertices of the inside rhombus. In other words, they are the solution of $Z_1 = Z_2 = \xi$. The regions outside the rectangle are identified as follows: In region P_1, both gross errors are detected. In region P_2, because we usually have $\delta_2 \leq K_2$, the lines $|\theta_2| = \delta_2$ divide P_2 further into two smaller regions corresponding to two possibilities: both biases are detected when $\delta_2 \leq |\theta_2| \leq K_2$ (subregion P_{2a} in Figure 7) and only gross error θ_1 is detected when $|\theta_2| < \delta_2$ (subregion P_{2b}). The same thing is observed for region P_3 where both biases are detected when $\delta_1 \leq |\theta_1| \leq K_1$ (subregion P_{3a}) and only gross error θ_2 is detected when $|\theta_1| < \delta_1$ (subregion P_{3b}). These subregions are illustrated in Figure 7 as shadowed regions.

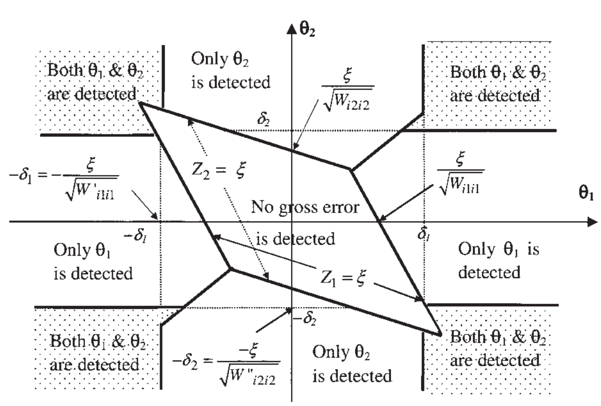

Figure 5. Different regions when two gross errors are present in the system.

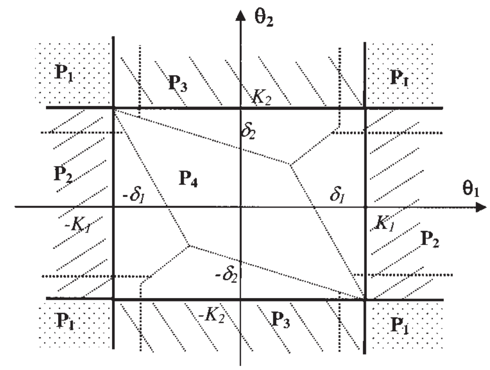

Figure 6. Regions used in analytical calculation.

Finally, region P_4 contains several different subregions, which are depicted in Figure 8. In subregion Ω_1 no bias is detected; in subregion Ω_2, only bias θ_1 is detected; and in subregion Ω_3, only bias θ_2 is detected. Finally, in Ω_4 both biases are detected.

Given the different regions that have been defined, we write

$$\frac{4P}{\Phi^2_{i_1,i_2}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F d\theta_1 d\theta_2 = \int_{P_1} F d\theta_1 d\theta_2 + \int_{P_2} F d\theta_1 d\theta_2 + \int_{P_3} F d\theta_1 d\theta_2 + \int_{P_4} F d\theta_1 d\theta_2 \quad (9)$$

$$\frac{2DEFL}{KT} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G d\theta_1 d\theta_2 = \int_{P_1} G d\theta_1 d\theta_2 + \int_{P_2} G d\theta_1 d\theta_2 + \int_{P_3} G d\theta_1 d\theta_2 + \int_{P_4} G d\theta_1 d\theta_2 \quad (10)$$

where F and G are the discontinuous integrands inside the corresponding integral expressions for P and $DEFL$ as shown in Eqs. 7 and 8.

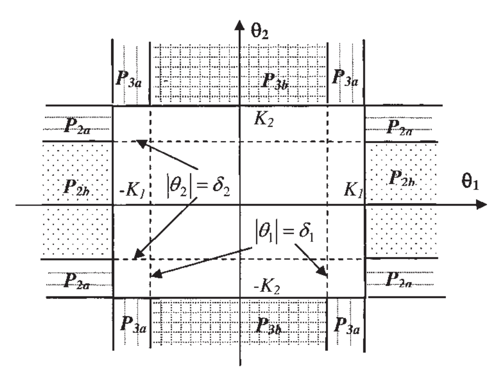

Figure 7. Subregions of P_2 and P_3.

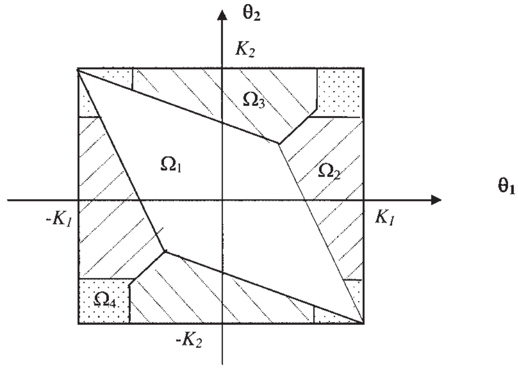

Figure 8. Subregions of P_4.

In region P_1 we obtain the following two closed-form solutions of the integral:

$$\int_{P_1} F d\theta_1 d\theta_2 = \left\{1 - \frac{1}{2}\left[\operatorname{erf}\left(\frac{K_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right) - \operatorname{erf}\left(\frac{-K_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right)\right]\right\} \times \left\{1 - \frac{1}{2}\left[\operatorname{erf}\left(\frac{K_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right) - \operatorname{erf}\left(\frac{-K_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right)\right]\right\} \quad (11)$$

$$\int_{P_1} G d\theta_1 d\theta_2 = \frac{\hat{\sigma}_p}{\sqrt{2\pi}}\left\{1 - \frac{1}{2}\left[\operatorname{erf}\left(\frac{K_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right) - \operatorname{erf}\left(\frac{-K_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right)\right]\right\}\left\{1 - \frac{1}{2}\left[\operatorname{erf}\left(\frac{K_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right) - \operatorname{erf}\left(\frac{-K_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right)\right]\right\} \quad (12)$$

Regions P_2 and P_3 are divided into two integrals as follows:

$$\int_{P_2} F d\theta_1 d\theta_2 = \int_{P_{2a}} F d\theta_1 d\theta_2 + \int_{P_{2b}} F_1 d\theta_1 d\theta_2 \quad (13)$$

$$\int_{P_3} F d\theta_1 d\theta_2 = \int_{P_{3a}} F d\theta_1 d\theta_2 + \int_{P_{3b}} F_1 d\theta_1 d\theta_2 \quad (14)$$

The following two integrals on subregions of P_2 and P_3 have closed-form solutions as follows:

$$\int_{P_{2a}} F d\theta_1 d\theta_2 = \left\{\int_{-\infty}^{-K_1} + \int_{K_1}^{\infty}\right\}\left\{\int_{-K_2}^{-\delta_2} + \int_{\delta_2}^{K_2}\right\} \frac{e^{-(\theta_1-\bar{\delta}_1)^2/2\rho_1^2}}{\rho_1\sqrt{2\pi}}\frac{e^{-(\theta_2-\bar{\delta}_2)^2/2\rho_2^2}}{\rho_2\sqrt{2\pi}} d\theta_1 d\theta_2 = \left\{1 - \frac{1}{2}\left[\operatorname{erf}\left(\frac{K_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right) - \operatorname{erf}\left(\frac{-K_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right)\right]\right\}\frac{1}{2}\left\{\operatorname{erf}\left(\frac{-\delta_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right) - \operatorname{erf}\left(\frac{-K_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right) + \operatorname{erf}\left(\frac{K_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right) - \operatorname{erf}\left(\frac{\delta_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right)\right\} \quad (15)$$

$$\int_{P_{3a}} F d\theta_1 d\theta_2 = \left\{\int_{-K_1}^{-\delta_1} + \int_{\delta_1}^{K_1}\right\}\left\{\int_{-\infty}^{-K_2} + \int_{K_2}^{\infty}\right\} \frac{e^{-(\theta_1-\bar{\delta}_1)^2/2\rho_1^2}}{\rho_1\sqrt{2\pi}}\frac{e^{-(\theta_2-\bar{\delta}_2)^2/2\rho_2^2}}{\rho_2\sqrt{2\pi}} d\theta_1 d\theta_2 = \frac{1}{2}\left\{\operatorname{erf}\left(\frac{-\delta_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right) - \operatorname{erf}\left(\frac{-K_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right) + \operatorname{erf}\left(\frac{K_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right) - \operatorname{erf}\left(\frac{\delta_1 - \bar{\delta}_1}{\sqrt{2}\,\rho_1}\right)\right\}\left\{1 - \frac{1}{2}\left[\operatorname{erf}\left(\frac{K_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right) - \operatorname{erf}\left(\frac{-K_2 - \bar{\delta}_2}{\sqrt{2}\,\rho_2}\right)\right]\right\} \quad (16)$$

Finally, the integral in region P_4 is decomposed in the corresponding integrals for the different subregions, that is,

$$\left.\begin{aligned}\int_{P_4} F d\theta_1 d\theta_2 &= \int_{\Omega_1} F d\theta_1 d\theta_2 + \int_{\Omega_2} F d\theta_1 d\theta_2 + \int_{\Omega_3} F d\theta_1 d\theta_2 + \int_{\Omega_4} F d\theta_1 d\theta_2 \\ &= \int_{\Omega_1} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 + \int_{\Omega_2} [1 + \operatorname{erf}(A_1)] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 \\ &\quad + \int_{\Omega_3} [1 + \operatorname{erf}(A_2)] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 \\ &\quad + \int_{\Omega_4} [1 + \operatorname{erf}(A_3)] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2\end{aligned}\right\} \quad (17)$$

$$
\left.\begin{aligned}
\int_{P_4} G d\theta_1 d\theta_2 &= \int_{\Omega_1} G d\theta_1 d\theta_2 + \int_{\Omega_2} G d\theta_1 d\theta_2 + \int_{\Omega_3} G d\theta_1 d\theta_2 + \int_{\Omega_4} G d\theta_1 d\theta_2 \\
&= \int_{\Omega_1} \frac{\hat{\sigma}_p}{\sqrt{2\pi}} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 \\
&\quad + \int_{\Omega_2} \left(\frac{\hat{\sigma}_p e^{-A_1^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\,\hat{\sigma}_p A_1 \{1 - \mathrm{erf}(A_1)\}}{2} \right) h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 \\
&\quad + \int_{\Omega_3} \left(\frac{\hat{\sigma}_p e^{-A_2^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\,\hat{\sigma}_p A_2 \{1 - \mathrm{erf}(A_2)\}}{2} \right) h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 \\
&\quad + \int_{\Omega_4} \left(\frac{\hat{\sigma}_p e^{-A_3^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\,\hat{\sigma}_p A_3 \{1 - \mathrm{erf}(A_3)\}}{2} \right) h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2
\end{aligned}\right\} \quad (18)
$$

These integrals have no closed analytical form but can be calculated approximately or numerically.

The above procedure to calculate integral expressions can be extended to calculate the probability and the corresponding *DEFL* when three or more gross errors are present in the system. The principle is to partition the space of variables into regions at much the same way, but in more dimensions. We illustrate this, for the case of calculating the probability in the presence of three gross errors, without resort to a figure and without going into the detail of the form of the integrand function. The integral in question for the probability is given by

$$
\frac{4P}{\Phi^3_{i_1,i_2,i_3}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\theta_1, \theta_2, \theta_3) d\theta_1 d\theta_2 d\theta_3 \quad (19)
$$

There are eight ($=2^3$) possibilities that can happen according to the presence and the number of undetected gross errors. They are: all gross errors are detected; only one gross error is detected (three possibilities); two of the three gross errors are detected (three possibilities); and no gross error is detected. Therefore, the integration set is then divided into eight ($=2^3$) regions as follows:

$$
\begin{aligned}
\frac{4P}{\Phi^3_{i_1,i_2,i_3}} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\theta_1, \theta_2, \theta_3) d\theta_1 d\theta_2 d\theta_3 \\
&= \left\{ \int_{-\infty}^{-K_1} + \int_{K_1}^{\infty} \right\} \left\{ \int_{-\infty}^{-K_2} + \int_{K_2}^{\infty} \right\} \left\{ \int_{-\infty}^{-K_3} \right. \\
&\quad \left. + \int_{K_3}^{\infty} \right\} F(\theta_1, \theta_2, \theta_3) d\theta_1 d\theta_2 d\theta_3 (PI_1) \quad + \sum \left\{ \int_{-\infty}^{-K_i} + \int_{K_i}^{\infty} \right\} \left\{ \int_{-\infty}^{-K_j} \right. \\
&\quad \left. + \int_{K_j}^{\infty} \right\} \int_{-K_k}^{K_k} F(\theta_1, \theta_2, \theta_3) d\theta_1 d\theta_2 d\theta_3 (\forall\, i \neq j \neq k) \left(\sum PI_{2i} \right) \quad + \sum \left\{ \int_{-\infty}^{-K_i} \right. \\
&\quad \left. + \int_{K_i}^{\infty} \right\} \int_{-K_j}^{K_j} \int_{-K_k}^{K_k} F(\theta_1, \theta_2, \theta_3) d\theta_1 d\theta_2 d\theta_3 (\forall\, i \neq j \neq k) \left(\sum PI_{3i} \right) \quad + \int_{-K_1}^{K_1} \int_{-K_2}^{K_2} \int_{-K_3}^{K_3} F(\theta_1, \theta_2, \theta_3) d\theta_1 d\theta_2 d\theta_3 (PI_4)
\end{aligned} \quad (20)
$$

Again $\{K_1, K_2, K_3\}$ are the vertices of the intersection $Z_1 = Z_2 = Z_3 = \xi$ (Z_1, Z_2, Z_3 are the MT test statistics for the three biased measurements). Clearly, in the space of variables of the integral PI_1, all three gross errors are detected and PI_1 has a closed-form solution. The spaces of PI_{2i} can be divided further into two smaller regions corresponding to the two possibilities: all three gross errors are detected (when $\xi/\sqrt{W^*_{kk}} = \delta_k \leq |\theta_k|$ [ltequ[K_k) and two gross errors are detected (when $|\theta_k| < \delta_k$). There are several possibilities in the remaining terms, which require the corresponding spaces of variables to be divided into subregions in much the same way as the region P_4 (in Eqs. 9 and 10) is divided into Ω_1, Ω_1, Ω_3, Ω_4.

Now if four gross errors are present, there are 16 ($=2^4$) possibilities that can happen and the integration set is partitioned into 16 different subregions and so on. For the presence of n gross errors, the whole space of variables is partitioned into 2^n regions. The corresponding integral can be summarized as follows:

$$\begin{aligned}
\frac{4P}{\Phi^n_{i_1,i_2,i_3,\cdots i_n}} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} F(\theta_1,\ \theta_2,\ \cdots\ \theta_n)d\theta_1 d\theta_2\cdots d\theta_n \\
&= \left\{\int_{-\infty}^{-K_1} + \int_{K_1}^{\infty}\right\}\left\{\int_{-\infty}^{-K_2} + \int_{K_2}^{\infty}\right\}\cdots\left\{\int_{-\infty}^{-K_n} + \int_{K_n}^{\infty}\right\}F(\theta_1,\ \theta_2,\ \cdots\ \theta_n)d\theta_1 d\theta_2\cdots d\theta_n \qquad (PI_1) \\
&+ \sum\left\{\int_{-\infty}^{-K_i} + \int_{K_i}^{\infty}\right\}\cdots\left\{\int_{-\infty}^{-K_j} + \int_{K_j}^{\infty}\right\}\left\{\int_{-\infty}^{-K_k} + \int_{K_k}^{\infty}\right\}\int_{-K_l}^{K_l} F(\theta_1,\ \theta_2,\ \cdots\ \theta_n)d\theta_1 d\theta_2\cdots d\theta_n \quad \forall\, i \neq j \neq k \neq \cdot \left(\sum PI_{2i}\right) \\
&+ \sum\left\{\int_{-\infty}^{-K_i} + \int_{K_i}^{\infty}\right\}\left\{\int_{-\infty}^{-K_j} + \int_{K_j}^{\infty}\right\}\cdots\int_{-K_k}^{K_k}\int_{-K_l}^{K_l} F(\theta_1,\ \theta_2,\ \cdots\ \theta_n)d\theta_1 d\theta_2\cdots d\theta_n \quad \forall\, i \neq j \neq k \neq \cdots \left(\sum PI_{3i}\right) \\
&+ \sum\left\{\int_{-\infty}^{-K_i} + \int_{K_i}^{\infty}\right\}\cdots\int_{-K_j}^{K_j}\int_{-K_k}^{K_k}\int_{-K_l}^{K_l} F(\theta_1,\ \theta_2,\ \cdots\theta_n)d\theta_1 d\theta_2\cdots d\theta_n \quad \forall\, i \neq j \neq k \neq \cdots \left(\sum PI_{4i}\right) \\
&+ \cdots \\
&+ \int_{-K_1}^{K_1}\int_{-K_2}^{K_2}\cdots\int_{-K_{n-1}}^{K_{n-1}}\int_{-K_n}^{K_n} F(\theta_1,\ \theta_2\cdots\ \theta_n)d\theta_1 d\theta_2\cdots d\theta_n \qquad (PI_f)
\end{aligned} \tag{21}$$

As before, the first term can be calculated analytically because all gross errors are detected. The terms PI_{2i} are to be calculated in the same way as PI_{2i} in Eq. 16 because there exist only two possibilities in the corresponding spaces of variables and so on. For calculating *DEFL*, the same principle is applied; the only difference is the integrand function.

To finish the calculation procedure, the terms that do not have closed-form solution are approximated using an approximation method called *Average of Bounds* or evaluated numerically. We now describe the two different methods to evaluate them.

Average of Bounds Method

To calculate P and *DEFL*, one needs to evaluate integrals of the form

$$J = \int_a^b e^{-z^2}\mathrm{erf}\{z\}dz \tag{22}$$

which do not have a closed-form solution. Thus, the essence of this method is to replace one of the integrand functions by a piecewise linear function. If the error function is replaced by a linear expression, J becomes

$$J^* = \sum \int_{a_i}^{b_i} (y_{1i} + y_{2i}z)e^{-z^2}dz \tag{23}$$

which can be calculated analytically. Certain choices of piecewise linear functions guarantee that the answer is an underestimate J_L^* or an overestimate J_U^* of J (Figure 9).

Thus, the integral J can be approximated as the average of J_L^* and J_U^*, that is,

$$J = (J_L^* + J_U^*)/2 \tag{24}$$

Because J_U^* and J_L^* are bounds of J, then, the maximum error is given by $(J_U^* - J_L^*)$.

Monte Carlo Method

Suppose we need to evaluate the following integral:

$$\mu = \int\cdots\int_{\Omega_X} g(\underline{x})f_{\underline{x}}(\underline{x})d\underline{x} \tag{25}$$

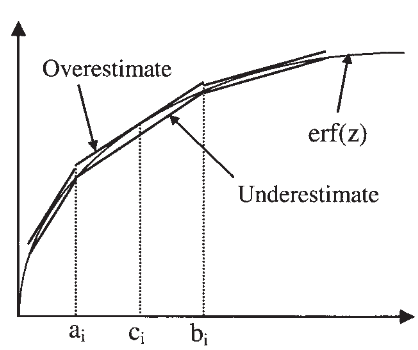

Figure 9. Average of Bounds method.

in which $g(\underline{x})$ is a function of $\underline{x}$, a vector of random variables described by a joint probability density function (pdf) $f_{\underline{x}}(\underline{x})$; Ω_X is the space of $\underline{x}$. If $g(\underline{x})f_{\underline{x}}(\underline{x})$ is bounded on Ω_X and Ω_X is a bounded subset of R^k, the above integral can be replaced by the following integral:

$$\mu = \int \cdots \int_{R^k} I_{\Omega_X}(\underline{x})\, g(\underline{x})\, f_{\underline{x}}(\underline{x})\, d\underline{x} \tag{26}$$

where $I_{\Omega_X}(\underline{x})$ is the indicator function for the set Ω_X,[6] given by

$$I_{\Omega_X}(\underline{x}) = \begin{cases} 1 & \text{if } \underline{x} \in \Omega_X \\ 0 & \text{if } \underline{x} \notin \Omega_X \end{cases} \tag{27}$$

To estimate the integral given by Eq. 25, we randomly sample a sequence of points $\boldsymbol{x}_i$, $i = 1, 2, \ldots, N$ from the density function $f_{\underline{x}}(\underline{x})$ and the space R^k and compute the sample mean

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} I_{\Omega_X}(\underline{x})\, g(\underline{x})_i \tag{28}$$

for computationally reasonable values of N. Given that $\boldsymbol{x}_i$, $i = 1, 2, \ldots, k$ are independent identically distributed random variables, it can be shown that $E[\hat{\mu}] = \mu$, that is, $\hat{\mu}$ is an unbiased estimator of μ.[7]

We now apply this Monte Carlo method to calculate the probability P and the downside financial loss $DEFL$. In this approach we do not need to partition space of variables into smaller regions because we consider P and $DEFL$ as a sum of separate integral terms defined over the whole space of variables as follows:

$$\frac{4P}{\Phi^2_{i_1,i_2}} = \left\{ \begin{array}{ll} \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_1(\theta_1, \theta_2) h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 & (PL_1) \\ \text{(when both gross errors are detected)} & \\ \displaystyle + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_2(\theta_1, \theta_2)[1 + \text{erf}\{A_1\}] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 & (PL_2) \\ \text{(when no gross error is detected)} & \\ \displaystyle + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_3(\theta_1, \theta_2)[1 + \text{erf}\{A_2\}] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 & (PL_3) \\ \text{(when only gross error 1 is detected)} & \\ \displaystyle + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_4(\theta_1, \theta_2)[1 + \text{erf}\{A_3\}] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 & (PL_4) \\ \text{(when only gross error 2 is detected)} & \end{array} \right\} \tag{29}$$

Similarly

$$\frac{2DEFL}{K_sT} = \left\{ \begin{array}{ll} \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_1(\theta_1, \theta_2) \frac{\hat{\sigma}_p}{\sqrt{2\pi}} \begin{cases} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 \\ \text{both gross errors are detected} \end{cases} & (FL_1) \\ \displaystyle + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_2(\theta_1, \theta_2)\left[\frac{\hat{\sigma}_p e^{-(A_1)^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\,\hat{\sigma}_p A_1[1 - \text{erf}(A_1)]}{2}\right] \begin{cases} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 \\ \text{no gross error is detected} \end{cases} & (FL_2) \\ \displaystyle + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_3(\theta_1, \theta_2)\left[\frac{\hat{\sigma}_p e^{-(A_2)^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\,\hat{\sigma}_p A_2[1 - \text{erf}(A_2)]}{2}\right] \begin{cases} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 \\ \text{only gross error 1 is detected} \end{cases} & (FL_3) \\ \displaystyle + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_4(\theta_1, \theta_2)\left[\frac{\hat{\sigma}_p e^{-(A_3)^2}}{\sqrt{2\pi}} - \frac{\sqrt{2}\,\hat{\sigma}_p A_3[1 - \text{erf}(A_3)]}{2}\right] \begin{cases} h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2 \\ \text{only gross error 2 is detected} \end{cases} & (FL_4) \end{array} \right\} \tag{30}$$

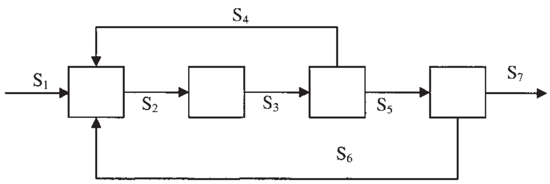

Figure 10. Example process.

where I_1, I_2, I_3, and I_4 are indicator functions for the corresponding integrals; h_1, h_2 and A_1, A_2, and A_3 are given above (Eqs. 7 and 8).

The method then relies on N trials, each trial consisting of two steps:

(1) Generate random numbers θ_1, θ_2 according to the probability distribution function $f(\theta_1, \theta_2)$:

$$f_{\underline{x}}(\underline{x}) \equiv f(\theta_1, \theta_2) = h_1(\theta_1, \bar{\delta}_1, \rho_1)h_2(\theta_2, \bar{\delta}_2, \rho_2) = \frac{e^{-(\theta_1-\bar{\delta}_1)^2/2\rho_1^2}}{\rho_1\sqrt{2\pi}} - \frac{e^{-(\theta_2-\bar{\delta}_2)^2/2\rho_2^2}}{\rho_2\sqrt{2\pi}} \quad (31)$$

Because distributions of gross errors are uncorrelated, this task is equivalent to generating θ_i according to the normal distribution $N(\bar{\delta}_i, \rho_i)$.

(2) With these values of θ_1, θ_2, check for the presence of undetected gross errors (by using the MIMT test) and calculate the values $I_{\Omega_X}(\underline{x})g(\underline{x})_i$ corresponding to separate integral terms PL_i or FL_i, such as $I_{\Omega_X}(\underline{x})g(\underline{x})_i$ in PL_i, are calculated as follows:

$$I_1(\theta_1, \theta_2) = \begin{cases} 1 & \text{if both gross errors are detected} \\ 0 & \text{other cases} \end{cases}$$

$$I_2(\theta_1, \theta_2)[1 + \text{erf}\{A_1\}] = \begin{cases} 1 + \text{erf}\{A_1\}] & \text{if no gross error is detected} \\ 0 & \text{other cases} \end{cases}$$

$$I_3(\theta_1, \theta_2)[1 + \text{erf}\{A_2\}] = \begin{cases} [1 + \text{erf}\{A_2\}] & \text{if only gross error 1 is detected} \\ 0 & \text{other cases} \end{cases}$$

$$I_4(\theta_1, \theta_2)[1 + \text{erf}\{A_3\}] = \begin{cases} [1 + \text{erf}\{A_3\}] & \text{if only gross error 2 is detected} \\ 0 & \text{other cases} \end{cases} \quad (32)$$

We then calculate estimators for PL_i or FL_i using Eq. 28. Then P or $DEFL$ can be calculated as a summation of integral terms PL_i or FL_i, respectively (Eqs. 29 and 30).

Example

Consider the process depicted in Figure 10, which was used by Bagajewicz.[4]

We assume that all variables are measured. The total flow rates are variables of interest, and thus the system is linear. The variance–covariance matrix of measurements is $S = \text{diag}(1.0, 0.2, 1.0, 1.0, 1.0, 1.0, 1.0)$. The incidence matrix is given by

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

Biases are assumed to follow a normal distribution with zero mean and standard deviation $\rho_1 = 3.0$, $\rho_2 = 4.0$, $\rho_3 = 4.0$, $\rho_4 = 5.0$, $\rho_5 = 6.0$, and $\rho_6 = 5.0$, respectively. Stream S_7 is the output stream (or product stream). Therefore, it is the stream for which we assume we calculate the economic value of accuracy. The solution, time of computation, and maximum error are reported and interval sizes are indicated in the following tables. The results were obtained using an Intel 2.4 GHz processor and 1024 MB RAM memory.

The financial loss calculation results for the presence of two biases obtained by using the two methods are shown in Table 1. In this table, the error reported for the average of bounds

Table 1. DEFL/K_sT for Two Gross Errors

Location of Biases	Average of Bounds Method (interval size = 0.1) Solution	Time (1/100th s)	Error	Monte Carlo Method Solution 1 $N = 10^6$	Solution 2 $N = 10^7$	Solution 3 $N = 10^8$	Time (s)	Relative Error (%)
1, 2	0.2162	10	0.6×10^{-5}	0.2162	0.2162	0.2161	509	0.021
1, 3	0.2148	11	0.5×10^{-5}	0.2148	0.2148	0.2147	527	0.041
1, 4	0.2199	16	0.4×10^{-5}	0.2199	0.2198	0.2198	533	0.056
1, 5	0.2208	19	0.8×10^{-5}	0.2208	0.2208	0.2207	517	0.016
1, 6	0.2493	34	1×10^{-5}	0.2491	0.2491	0.249	545	0.101
2, 3	0.2017	69	0.1×10^{-6}	0.2018	0.2017	0.2017	553	0.002
2, 4	0.2005	34	0.1×10^{-6}	0.2005	0.2005	0.2005	493	0.008
2, 5	0.2024	17	0.3×10^{-6}	0.2023	0.2024	0.2024	485	0.026
2, 6	0.2056	14	0.6×10^{-6}	0.2056	0.2056	0.2056	497	0.026
3, 4	0.2006	13	0.1×10^{-6}	0.2006	0.2006	0.2006	491	0.004
3, 5	0.2007	12	0.1×10^{-6}	0.2007	0.2007	0.2007	577	0.005
3, 6	0.2057	12	0.1×10^{-6}	0.2057	0.2057	0.2057	502	0.004
4, 5	0.2098	52	0.5×10^{-6}	0.2098	0.2098	0.2098	560	0.011
4, 6	0.2054	34	1×10^{-6}	0.2054	0.2054	0.2054	542	0.001
5, 6	0.2051	38	0.3×10^{-6}	0.2051	0.2052	0.2051	512	0.002

Table 2. DEFL/K_sT (Two Gross Errors): Average of Bounds Method with Different Interval Sizes

Location of Biases	Interval Size = 0.1			Interval Size = 0.2		
	Solution	Time (1/100th s)	Error	Solution	Time (1/100th s)	Error
1, 2	0.2162	10	0.6×10^{-5}	0.2162	3	2.3×10^{-5}
1, 3	0.2148	11	0.5×10^{-5}	0.2148	2	1.9×10^{-5}
1, 4	0.2199	16	0.4×10^{-5}	0.2197	4	3.4×10^{-5}
1, 5	0.2208	19	0.8×10^{-5}	0.2204	6	3×10^{-5}
1, 6	0.2493	34	1×10^{-5}	0.2487	10	6.2×10^{-5}
2, 3	0.2017	69	0.1×10^{-6}	0.2017	17	0.25×10^{-6}
2, 4	0.2005	34	0.1×10^{-6}	0.2005	10	0.1×10^{-6}
2, 5	0.2024	17	0.3×10^{-6}	0.2024	3	1.0×10^{-6}
2, 6	0.2056	14	0.6×10^{-6}	0.2056	3	2.3×10^{-6}
3, 4	0.2006	13	0.1×10^{-6}	0.2006	3	0.1×10^{-6}
3, 5	0.2007	12	0.1×10^{-6}	0.2007	4	0.3×10^{-6}
3, 6	0.2057	12	0.8×10^{-6}	0.2057	4	3.2×10^{-6}
4, 5	0.2098	52	0.5×10^{-6}	0.2098	13	1.9×10^{-6}
4, 6	0.2054	34	1×10^{-6}	0.2054	8	4.4×10^{-6}
5, 6	0.2051	38	0.3×10^{-6}	0.2052	9	1.3×10^{-6}

method is defined as $(J_U^* - J_L^*)$ and all values $< 0.1 \times 10^{-6}$ are rounded to 0.1×10^{-6}. In addition, N is the number of trial of the Monte Carlo method, and the time reported for the Monte Carlo method is the one corresponding to $N = 10^8$. In general, it was observed that the time is linear with the number of trials, that is, it decreases by 10 when N decreases by 10. Finally, the relative error column corresponds to the difference between methods. Table 2 compares the results for the financial loss calculations using different interval sizes for the piecewise function intervals.

The accuracy of solution and computation time of the Average of Bounds method depends mainly on two factors: interval size chosen and the parameters W_{ij} (to be used in the MT test statistics). When the interval size increases, the solution accuracy decreases (the error increases) but the computation time also decreases. The parameters W_{ij} determine the size of the regions divided (such as P_1, P_2, P_3, P_4), which affect the computation time. The larger the size of the rectangular regions (see P_4 in Figure 4), the longer the computation time. The results show that for the same number of gross errors, the computation time varies significantly. Moreover, this method is also subjected to a round-off problem; for example, the number of subintervals divided (such as the range $[-K_1, K_1]$ divided by interval size) may be rounded off, which will somehow affect the accuracy of solution. Clearly, the results show that at the same number of gross errors, the error, although small, also varies significantly.

For the Monte Carlo method, the accuracy of solution and the computation time of the Monte Carlo method depend on the number of trials N. When we increase N, we obtain a more accurate solution at the expense of longer computation time. It is obvious that we have convergence of solution when the number of trials $N \geq 10^6$. Indeed, the relative error between solution 1 (at $N = 10^6$) and solution 3 (at $N = 10^8$) is usually not more than 0.1% and the relative error between solution 2 (at $N = 10^7$) and solution 3 (at $N = 10^8$) is usually not more than 0.05%. Therefore, we can say that the Monte Carlo method at number of trials $N = 10^6$ has a satisfactory accuracy of solution and relatively fast computation speed (a few seconds; not more than 9 s when four gross errors are present). For the same number of gross errors, the computation time using the Monte Carlo method is virtually unchanged.

The Average of Bounds method is superior to the Monte Carlo method because its solution accuracy is satisfactory (the relative error between its solution and that of the Monte Carlo method at $N = 10^8$ is not more than 0.1%) and the computation time is shorter.

The above arguments regarding how and what factors affect the efficiency (accuracy and computation time) of the Average of Bounds method and the Monte Carlo methods also hold for other cases of the financial loss and the probability calculation in the presence of multiple biases (≥2).

Table 2 also shows that, when two gross errors are present, the two largest financial losses are incurred at locations of biases (1, 6) and (1, 5). This can be explained by the fact that an undetected gross error in stream S_1 causes the largest induced bias in stream S_7 (output stream or product stream), which then follow stream S_6 and S_5. Indeed, streams S_1 and S_7 are part of an equivalent set,[8] and streams S_5 and S_6 are connected directly to stream S_7 through a node. Therefore, undetected gross errors in these streams cause the worst effect on the measurement accuracy of stream S_7 (through data reconciliation). This argument can be verified by looking at the absolute value of the coefficient α_i in the expression $\hat{\delta}_p^i(\theta_i) = \alpha_i\theta_i$ for the induced bias in stream S_7 caused by an undetected gross error anywhere in the system: stream S_1 has the largest absolute value of α_i (0.3774), followed by streams S_6 and S_5 (−0.2453 and 0.1321, respectively). Note that because streams S_1 and S_7 are part of an equivalent set,[8] the results for biases in streams (1, 5), (1, 6), and (1, 2, 3) are the same as results for biases in streams (7, 5), (7, 6), and (7, 2, 3). In other words, if stream S_1 is replaced by stream S_7, the results are the same.

The results also confirm our expectation that that financial loss in the presence of biases is larger than the financial loss without biases.

Under the assumption of normal distributions and the means are zeros, the probabilities for the presence of multiple gross errors were calculated to be $\Phi^n_{i_1,i_2\cdots i_N}/4$, that is, the probability P is the function of $\Phi^n_{i_1,i_2\cdots i_N}$, which is the probability of such set of gross errors at a particular location to develop, only. This result is expected for the following reasons:

(1) The boundaries of integration are straight lines $f(\theta_1, \theta_2, \ldots, \theta_N)$ (as illustrated in Figure 5 for the case of two variables), defined by the expressions $W = \left|\sum_i a_i\theta_i\right| = \xi$. Thus, the boundaries of integration are symmetric around the origin $(0, 0, \ldots, 0)$ as clearly indicated in Figure 5.

Table 3. DEFL/K_sT for Three Gross Errors

Location of Biases	Average of Bounds Method (interval size = 0.2) Solution	Time[a] (s)	Error*	Monte Carlo Method Solution 1 $N = 10^6$	Solution 2 $N = 10^7$	Solution 3 $N = 10^8$	Time[b] (s)	Relative Error** (%)
1, 2, 3	0.2218	10	11×10^{-5}	0.222	0.222	0.222	719	0.099
1, 2, 4	0.2192	4	3.1×10^{-5}	0.2193	0.2194	0.2195	662	0.128
1, 2, 5	0.2279	3	4.8×10^{-5}	0.2277	0.2278	0.2279	641	0.009
1, 2, 6	0.2489	3	11×10^{-5}	0.2484	0.2486	0.2487	649	0.104
1, 3, 4	0.2194	1	2.9×10^{-5}	0.2196	0.2197	0.2197	654	0.133
1, 3, 5	0.2213	1	3.3×10^{-5}	0.2209	0.2211	0.2211	658	0.071
1, 3, 6	0.2489	3	11×10^{-5}	0.2487	0.249	0.249	644	0.065
1, 4, 5	0.2577	19	8.4×10^{-5}	0.2577	0.2581	0.2582	641	0.183
1, 4, 6	0.2482	6	13×10^{-5}	0.2477	0.2479	0.2479	637	0.114
1, 5, 6	0.247	8	12×10^{-5}	0.2471	0.2474	0.2475	652	0.208
2, 3, 5	0.2128	54	0.5×10^{-5}	0.2132	0.2131	0.2131	652	0.125
2, 3, 6	0.2056	26	0.6×10^{-5}	0.2056	0.2056	0.2056	648	0.016
2, 4, 5	0.2073	9	0.5×10^{-5}	0.2072	0.2072	0.2072	628	0.047
2, 4, 6	0.2053	6	0.5×10^{-5}	0.2053	0.2053	0.2053	612	0.012
2, 5, 6	0.2048	17	0.6×10^{-5}	0.2048	0.2048	0.2048	613	0.001
3, 4, 5	0.2097	4	0.4×10^{-5}	0.2097	0.2098	0.2098	623	0.025
3, 4, 6	0.2054	3	0.5×10^{-5}	0.2053	0.2054	0.2054	610	0.003
3, 5, 6	0.2052	4	0.5×10^{-5}	0.2052	0.2052	0.2052	609	0.022

(2) The probability can be written as (from Eq. 7, for the case of two gross errors)

$$P = \frac{1}{4}\Phi^2_{i_1,i_2} \times \left(1 + \int\int_{\Omega} \mathrm{erf}[u(\theta_1, \theta_2)] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2) d\theta_1 d\theta_2\right)$$

where $u(\theta_1, \theta_2)$ is a linear function of θ_1, θ_2: $u(\theta_1, \theta_2)$ is either A_1, A_2, or A_3, given right below Eq. 8; the space of variable Ω is symmetric around the origin (0, 0) as explained above. If the mean values are zero then $h_1(\theta_1, \bar{\delta}_1, \rho_1)$, $h_2(\theta_2, \bar{\delta}_2, \rho_2)$ are even functions, whereas the error function $\mathrm{erf}[u(\theta_1, \theta_2)]$ is odd {that is, $\mathrm{erf}[u(\theta_1, \theta_2)] = -\mathrm{erf}[u(-\theta_1, -\theta_2)]$}, and thus the integrand $G(\theta_1, \theta_2) = \mathrm{erf}[u(\theta_1, \theta_2)] h_1(\theta_1, \bar{\delta}_1, \rho_1) h_2(\theta_2, \bar{\delta}_2, \rho_2)$ is an odd function. Consequently, the integral $\int\int_{\Omega} G(\theta_1, \theta_2) d\theta_1 d\theta_2$ is zero and

$$P = \frac{1}{4}\Phi^2_{i_1,i_2} \qquad \text{(two gross errors case)}$$

or

$$P = \frac{1}{4}\Phi^n_{i_1,i_2,\ldots,i_N} \qquad \text{(in the general case)}$$

The financial loss calculation results in the presence of three biases, obtained by using the two methods, are shown in Table 3. Tables 1, 2, and 3 show that when the number of gross errors is increased by one (from two to three), the computation time of the Average of Bounds method increases significantly. Using the same step size, the computation time can increase 50- to 200-fold. The results also show that when the number of gross errors increases, the error of this method also significantly increases (solution accuracy decreases), although this result is partly attributed to the increase in interval size used. These observations are also applied to the probability calculation.

Regarding the Monte Carlo method, when the number of gross errors is increased by one (from two to three), the computation time increases only slightly (roughly 1.2- to 1.3-fold). Comparing the two methods, when three gross errors are present (the number of variables is three), the computation time of the Average of Bounds method is comparable to that of the Monte Carlo method at number of trials $N = 10^6$ (about a few seconds) but it is still significantly smaller than that of the Monte Carlo method with $N = 10^8$. The above arguments regarding factors affect efficiency of the two methods also hold for probability calculation in the presence of three biases.

From the results shown in Table 1 (two biases) and Table 3 (three biases), it is also obvious that the financial loss increases when the number of gross errors increases. For example, financial losses incurred as a result of the biases at locations (1, 2) and (1, 3) are 0.2162 and 0.2148, respectively, whereas those at locations (1, 2, 3) and (1, 2, 3, 5) are 0.2218 and 0.2591, respectively.

We know that, at the same magnitude, undetected gross errors in streams S_1, S_5, and S_6 cause the largest induced biases in stream S_7. However, Table 3 shows that when three gross errors are present, the three largest financial losses are incurred at bias locations (1, 4, 5), (1, 3, 6), and (1, 4, 6) rather than at (1, 5, 6) (the fourth largest financial loss incurred). The reason is that financial loss depends strongly on the magnitude of the induced bias, which in turn depends not only on the coefficients α_i but also on the power of the gross error detection strategy to detect sets of gross errors at specific locations. Compared with the set of gross errors at location (1, 4, 5), the set of gross errors at location (1, 5, 6) renders larger coefficients α_i ($\alpha_6 > \alpha_4$) but smaller magnitudes of maximum undetected gross errors (that is, the set of gross errors at location (1, 4, 5) is more resistant to gross error detection than that at location (1, 5, 6). This

Table 4. DEFL/K_sT for Four Gross Errors

Location of Biases	Average of Bounds Method (interval size = 0.4) Solution	Time[a] (s)	Error*	Monte Carlo Method Solution 1 $N = 10^6$	Solution 2 $N = 10^7$	Solution 3 $N = 10^8$	Time[b] (s)	Relative Error** (%)
1, 2, 3, 5*	0.2591	146	13×10^{-4}	0.2598	0.2592	0.2592	798	0.027
1, 2, 3, 6	0.2441	112	3.3×10^{-4}	0.2443	0.2443	0.2442	783	0.058
1, 2, 4, 6	0.2474	33	3.7×10^{-4}	0.2481	0.2479	0.2479	789	0.191
1, 3, 4, 5	0.2563	38	4.2×10^{-4}	0.2561	0.256	0.256	810	0.117

*1, 2, 3, 5[c]: interval size = 0.5 was used.

situation arises because a certain set of gross errors is associated with a certain set of parameters W_{ij} (to be used in MIMT test), which in turn allows drawing the rhombus shown in Figure 5 (recall that inside the rhombus, gross errors are undetected). The smaller the rhombus, the smaller the maximum undetected gross errors (given as vertices of the rhombus), and thus the more easily the set of gross errors is detected (that is, less resistance to detection of gross errors). Concluding, the financial loss incurred as the result of a set of gross errors at a particular location (that is, at particular streams) is a function of two factors: (1) the coefficients α_i that indicate the effect of undetected gross errors in these streams to measurement accuracy of product stream and (2) the interaction between undetected biases, which can cancel each other and therefore not be detected, when they would be detected otherwise when they are alone.

The financial loss calculation results for the presence of four biases obtained by using the two methods are shown in Table 4.

Comparing the two methods, when four gross errors are present (number of variables is four), the computation time of the Average of Bounds method is comparable to that of the Monte Carlo method at number of trials $N = 10^7$ (about 80 s) but it is still significantly smaller than that of the Monte Carlo method at number of trials $N = 10^8$.

At high number of gross errors (≥ 6), the Average of Bounds method is no longer superior to the Monte Carlo method because computation time of the Average of Bounds method increases significantly, whereas that of the Monte Carlo method increases insignificantly with the number of gross errors. An alternative choice at high number of gross errors is the Monte Carlo method at low number of trials ($N = 10^6$ or $N = 10^7$) whose solution is sufficiently accurate and the computation time is acceptable. Short computation time is important because the financial loss calculation can be used in sensor network design, which needs to explore many alternatives combinatorially.

Next, the effect of changing parameters to the effectiveness of the Average of Bounds method is investigated. The values of the means of the probability distribution functions of bias sizes were changed to nonzero values and the same standard deviations. Tables 5 and 6 show the results. Interval sizes of 0.1, 0.2, and 0.4 were used for the financial loss calculation in the presence of two, three, and four gross errors, respectively, with the exception of the case (1, 2, 3, 5) where an interval size of 0.5 was used.

From Tables 5 and 6 we can see that when parameters such as the means of the probability distributions of biases change, the computation time and the errors (solution accuracy) of the Average of Bounds method are virtually unchanged for the same number of gross errors and the same locations.

Discussion

We discussed computational issues in the determination of financial loss in the presence of multiple gross errors. The simplified assumptions are that the measurements, the true values follow normal distributions around their means, and

Table 5. DEFL/K_sT for Multiple Gross Errors When the Means Are Changed

Locations of Biases	Zero Means Solution	Zero Means Error	Positive Means (+1) Solution	Positive Means (+1) Error	Negative Means (−1) Solution	Negative Means (−1) Error
1, 2	0.2162	0.6×10^{-5}	0.2033	0.6×10^{-5}	0.228	0.6×10^{-5}
1, 3	0.2148	0.5×10^{-5}	0.2029	0.5×10^{-5}	0.2258	0.4×10^{-5}
1, 5	0.2208	0.8×10^{-5}	0.2064	0.7×10^{-5}	0.2341	0.7×10^{-5}
1, 6	0.2493	1×10^{-5}	0.2333	1.5×10^{-5}	0.263	1.4×10^{-5}
2, 6	0.2056	0.06×10^{-5}	0.2082	0.07×10^{-5}	0.2028	0.09×10^{-5}
4, 5	0.2098	0.05×10^{-5}	0.211	0.07×10^{-5}	0.2083	0.07×10^{-5}
5, 6	0.2051	0.03×10^{-5}	0.2081	0.06×10^{-5}	0.202	0.07×10^{-5}
1, 2, 3	0.2218	11×10^{-5}	0.2016	10.1×10^{-5}	0.2418	10.2×10^{-5}
1, 2, 4	0.2192	3.1×10^{-5}	0.2086	6.14×10^{-5}	0.2285	6.1×10^{-5}
1, 3, 6	0.2489	11×10^{-5}	0.2336	10.4×10^{-5}	0.2619	10.4×10^{-5}
1, 4, 5	0.2577	8.4×10^{-5}	0.2389	12.3×10^{-5}	0.2752	12.2×10^{-5}
3, 4, 5	0.2097	0.4×10^{-5}	0.211	1.0×10^{-5}	0.2081	1.0×10^{-5}
3, 5, 6	0.2052	0.5×10^{-5}	0.2083	0.84×10^{-5}	0.202	0.84×10^{-5}
1, 2, 3, 6	0.2441	3.3×10^{-4}	0.2324	3.15×10^{-4}	0.2534	3.14×10^{-4}
1, 2, 4, 6	0.2474	3.7×10^{-4}	0.232	3.7×10^{-4}	0.2604	3.7×10^{-4}
1, 3, 4, 5	0.2563	4.2×10^{-4}	0.2381	4.13×10^{-4}	0.2734	4.13×10^{-4}

Table 6. **The Probability ($4P/\Phi^2_{i_1}, i_2$), ($4P/\Phi^3_{i_1}, i_2, i_3$), or ($4P/\Phi^4_{i_1}, i_2, i_3, i_4$) for Multiple Gross Errors with Various Values for the Means**

Location of Biases	Zero Means		Positive Means (+1)		Negative Means (−1)	
	Solution	Error	Solution		Solution	Error
1, 2	1	4×10^{-5}	1.036	3.8×10^{-5}	0.964	3.73×10^{-5}
1, 5	1	4.8×10^{-5}	1.0395	4.6×10^{-5}	0.9605	4.62×10^{-5}
2, 6	1	1×10^{-5}	0.9918	0.95×10^{-5}	1.0082	1×10^{-5}
1, 2, 4	1	16×10^{-5}	1.0286	16.2×10^{-5}	0.9714	15.4×10^{-5}
3, 4, 5	1	3×10^{-5}	0.9956	2.8×10^{-5}	1.0044	2.9×10^{-5}
3, 5, 6	1	3×10^{-5}	0.9904	2.61×10^{-5}	1.0096	3×10^{-5}
1, 2, 3, 6	0.999	6.8×10^{-4}	1.025	6.5×10^{-4}	0.9723	6.7×10^{-4}
1, 3, 4, 5	0.9992	9.1×10^{-4}	1.0419	9×10^{-4}	0.9563	9×10^{-4}

negligible process variations. For non-normal distributions, the proposed methodologies can still be applied with appropriate modifications. Indeed, the proposed Monte Carlo method is independent of the density function $f_{\underline{x}}(\underline{x})$ [in this case the density functions $h_1(\theta_1, \bar{\delta}_1, \rho_1)$ and $h_2(\theta_2, \bar{\delta}_2, \rho_2)$]. Thus, the only required modification in the Monte Carlo method is the generation of random numbers θ_1, θ_2 according to the new probability distribution functions (other than normal distribution). In fact, one of the advantages of the Monte Carlo method is its simplicity in preparing the computer codes and its adaptation to different purposes. Thus, "upgrading" computer codes to calculate financial loss for two gross errors to ones that calculate financial loss for several gross errors ($n > 2$) requires little effort. The modification of computer codes to accommodate the new density function can also be easily conducted. On the order hand, the Average of Bounds method requires considerable effort in "upgrading" (increase number of gross errors) or "modifying" (density functions other than normal distribution). With new density functions, the principle of partition of space of variables is still applied, but all the closed-form solutions of integrals (such as P_1) need to be rederived and implemented.

Now if the assumption of negligible process variation is relaxed, the financial loss for the presence of two gross errors is still given by Eq. 8, but now A_1, A_2, and A_3 are different:

$$A_1 = \frac{\alpha_1\theta_1 + \alpha_2\theta_2 + m_p - m_p^*}{\sqrt{2}\,\hat{\sigma}_p}$$

$$A_2 = \frac{\alpha_2\theta_2 + m_p - m_p^*}{\sqrt{2}\,\hat{\sigma}_p} \qquad A_3 = \frac{\alpha_1\theta_1 + m_p - m_p^*}{\sqrt{2}\,\hat{\sigma}_p}$$

(Recall that m_p is the true value of the process variable, which is assumed to follow normal distribution around the meanm_p^*.) An approach to calculate the financial loss consists of two steps: (1) sampling a random number for m_p then substituting this value into the expression for financial loss DEFL; (2) calculating the DEFL using the Average of Bounds methods or Monte Carlo method. It is obvious that, in this approach, the calculated financial loss DEFL depends on the sampled value of m_p. If we sample enough values for m_p and calculate the financial loss for each value of m_p, we can obtain the expected value for financial loss and this expected value is the same as the financial loss DEFL with negligible process variation assumption because the expected value for ($m_p - m_p^*$) is zero. Thus, we can say that the calculated results given above are valid with or without the mentioned assumption with the notation that, without the negligible process variation assumption, the calculated results are the "expected" values rather than the "exact" values for financial losses.

Applications

Having presented and compared two computation procedures for determining the financial losses, we now show some applications of financial losses calculation, that is, the gain in economic value (the decrease in financial loss) when we add an instrumentation to an existing sensor network. Let us consider another example, the ammonia process given in Figure 11.

The simplified ammonia process was used by Ali and Narasimhan.[9] It consists of six nodes and eight edges with node 6 representing the environmental node, as shown in Figure 11. The total flow rates are the variables of interest. The total flow rates for streams $\{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\}$ are {152.6, 152.6, 152.6, 86.4, 66.2, 96.6, 30.4, 56}, respectively.

We assume that the stream S_7 is the main product stream and is the stream for which financial loss is calculated and that the existing sensor network consists of four sensors placed on four streams S_5, S_6, S_7, and S_8. We want to upgrade the existing sensor network by adding one more sensor on one of the streams S_1, S_2, S_3, or S_4. Our objective is to find the best sensor location by calculating the net present value (NPV) for each candidate sensor location, which is given by

$$NPV = d_n\{\text{Change in } DEFL\} - \text{Cost of new instrumentation} \tag{33}$$

In Eq. 33, the *complete* expression for financial loss *DEFL* is given by[4]

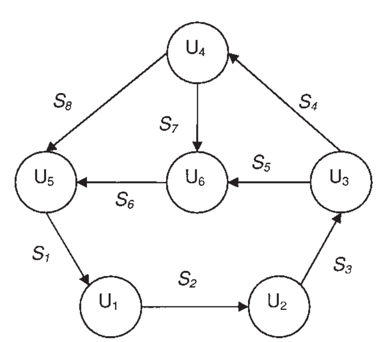

Figure 11. Example 2, the simplified ammonia process.

Table 7. Financial Loss as a Function of Sensor Network

System of Sensors	Sensor Added to the Original Sensor System	Financial Losses (DEFL/ K_sT)	Improvement in Financial Loss Due to Added Sensor
$\{S_5, S_6, S_7, S_8\}$		0.15813	
$\{S_1, S_5, S_6, S_7, S_8\}$	S_1	0.15419	0.00394
$\{S_2, S_5, S_6, S_7, S_8\}$	S_2	0.15406	0.00407
$\{S_3, S_5, S_6, S_7, S_8\}$	S_3	0.15414	0.00399
$\{S_4, S_5, S_6, S_7, S_8\}$	S_4	0.14025	0.01788

$$DEFL = \Psi^0 DEFL^0 + \sum_i \Psi_i^1 DEFL^1\Big|_i + \sum_{i_1,i_2} \Psi_{i_1,i_2}^2 DEFL^2\Big|_{i_1,i_2} + \cdots \sum_{i_1,i_2\cdots i_N} \Psi_{i_1,i_2\cdots i_N}^n DEFL^N\Big|_{i_1,i_2\cdots i_N} \quad (34)$$

In Eq. 34, $\Psi^n_{i_1,i_2\cdots i_N}$ and $DEFL^N|_{i_1,i_2\cdots i_N}$ represent the average fraction of time the system is in the state containing n gross errors $i_1, i_2, \ldots, i_N$ and its associated financial losses, respectively. The values Ψ are in fact equal to the probability of each state, that is, $\Psi^n_{i_1,i_2\cdots i_N}$ is equal to the probability of n instruments failing and the others not:

$$\Psi^n_{i_1,i_2,\ldots,i_N} = \Phi^n_{i_1,i_2,\ldots,i_N} = f_{i_1}(t)\cdots f_{i_N}(t) \prod_{s\neq i_1,\cdots s\neq i_N} [1 - f_s(t)]$$

(Recall that $f_i(t)$ is the probability of failure of sensor i.) For the ammonia process example, we use $f_i(t) = 0.1$ for all sensors. Assume next that all sensor costs are equal; then the best location for the added sensor would be the one with the largest reduction of financial loss. The calculated financial losses are given in Table 7.

Clearly, by adding a sensor to stream S_4, one more *independent* material balance equation ($F_4 + F_7 = F_8$) that involves the flow rate of the product stream (F_7) is obtained to be used for data reconciliation and gross error detection. This is not the case for other candidates (that is, with other candidates, the new equation obtained does not involve F_7).

Conclusion

Two new methods have been proposed to calculate the downside expected financial loss and the associated probability for the presence of gross errors. Both methods are accurate and can be used for repetitive calculations in the framework of sensor network design, where several locations of measurements have to be evaluated. With the same sensor network, financial loss has been shown to be a function of the number and location of biases, whereas different sensor network configurations render different values of financial loss of the system. This characteristic can be used in the field of sensor network design and retrofit.

Acknowledgments

We gratefully acknowledge the financial support of this work from National Science Foundation Grant CTS-0350501.

Literature Cited

1. Bagajewicz M. On the definition of software accuracy in redundant measurement systems. *AIChE J.* 2005;51:1201-1206.
2. Bagajewicz M, Markowski M. Instrumentation design and upgrade using an unconstrained method with pure economical objectives. Proceedings of FOCAPO (Foundations of Computer Aided Process Operations), Coral Springs, FL; 2003.
3. Bagajewicz M, Markowski M, Budek A. Economic value of precision in the monitoring of linear systems. *AIChE J.* 2005;51:1304-1309.
4. Bagajewicz M. Value of accuracy in linear systems. *AIChE J.* 2006; 52:638-650.
5. Bagajewicz M. On a new definition of a stochastic-based accuracy concept of data reconciliation-based estimators. Proceedings of the ESCAPE 2005, Barcelona, Spain; 2005.
6. Evans M, Swartz T. *Approximating Integrals via Monte Carlo and Deterministic Methods.* Oxford, UK: Oxford Univ. Press; 2000.
7. Lu Z, Zhang D. On importance sampling Monte Carlo approach to uncertainty analysis for flow and transport in porous media. *Adv Water Resourc.* 2003;26:1177-1188.
8. Bagajewicz M, Jiang Q. Gross error modeling and detection in plant linear dynamic reconciliation. *Comput Chem Eng.* 1998;22:1789-1810.
9. Ali Y, Narasimhan S. Sensor network design for maximizing reliability of linear processes. *AIChE J.* 1993;39:820-828.
10. Noble B, Daniel JW. *Applied Linear Algebra*. 3rd Edition. Englewood Cliffs, NJ: Prentice Hall; 1988.

Appendix

This appendix proves that $W_{i_1i_1} > W'_{i_1i_1}$, where $\boldsymbol{W}'$ is the updated matrix $\boldsymbol{W}$ after a redundant measurement has been eliminated and $W_{i_1i_1}$ and $W'_{i_1i_1}$ are diagonal elements of matrices $\boldsymbol{W}$, $\boldsymbol{W}'$. More specifically, matrix $\boldsymbol{W}$ is given by: $\boldsymbol{W} = \boldsymbol{A^T}(\boldsymbol{ASA^T})^{-1}\boldsymbol{A}$ and matrix $\boldsymbol{W}'$ is given by $(A')^T[A'S'(A')^T]^{-1}A'$, where $\boldsymbol{A}'$, $\boldsymbol{S}'$ are the updated matrices $\boldsymbol{A}$ (constraint matrix) and $\boldsymbol{S}$ (covariance matrix of measurements) after a redundant measurement has been removed.

Without loss of generality, we assume that the column that corresponds to the measurement to be eliminated contains only one nonzero element, that is, this measurement is present in only one material balance. Then matrices $\boldsymbol{A}$ and $\boldsymbol{A}'$ are related through the following expression:

$$\boldsymbol{A} = \begin{pmatrix} \begin{bmatrix} \boldsymbol{A}' \\ n\times m \end{bmatrix} & \begin{bmatrix} \boldsymbol{0} \\ n\times 1 \end{bmatrix} \\ [\boldsymbol{B}_i,\ 1\times m] & \delta \end{pmatrix} \quad (A1)$$

where vector $[\boldsymbol{B[b]}_i,\ \delta] = [1 \times (m+1)]$ is the last row corresponding to the constraint relating the redundant measurement that has been eliminated with some other redundant measurements in the systems; the vector $[\boldsymbol{0}, \delta] = [(n+1) \times 1]$ is the last column corresponding to the eliminated redundant measurement (δ can take a value of either $+1$ or -1).

In turn, the covariance matrices of measurements $\boldsymbol{S}$ and $\boldsymbol{S}'$ are related by (we assume that $\boldsymbol{S}$ and $\boldsymbol{S}'$ are diagonal matrices)

$$\boldsymbol{S} = \begin{pmatrix} \begin{bmatrix} \boldsymbol{S}' \\ m\times m \end{bmatrix} & \begin{bmatrix} \boldsymbol{0} \\ m\times 1 \end{bmatrix} \\ [\boldsymbol{0},\ 1\times m] & s_i \end{pmatrix} \quad (A2)$$

Then

$$(\boldsymbol{ASA^T})^{-1} = \begin{pmatrix} \begin{bmatrix} \boldsymbol{A'S'(A')^T} \\ n\times n \end{bmatrix} & \begin{bmatrix} \boldsymbol{A'S'B_i^T} \\ n\times 1 \end{bmatrix} \\ [\boldsymbol{B_iS'(A')^T},\ 1\times n] & B_iS'B_i^T + s_i\delta^2 \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{P} & \boldsymbol{Q} \\ \boldsymbol{R} & \boldsymbol{L} \end{pmatrix}^{-1} = \begin{bmatrix} \boldsymbol{X} & -\boldsymbol{P^{-1}QY} \\ -\boldsymbol{YRP^{-1}} & \boldsymbol{Y} \end{bmatrix} \quad (A3)$$

where $Y = (L - RP^{-1}Q)^{-1}$ and $X = P^{-1} + P^{-1}QYRP^{-1}$.[10] In this case L and Y are scalars. Then

$$A^T(ASA^T)^{-1}A = \begin{pmatrix} [([(A')^T X] + [B_i^T[-YRP^{-1}]])[A'] + ([(A')^T[-P^{-1}QY]] + Y[B_i^T])[B_i]] & \delta[[(A')^T[-P^{-1}QY]] + Y[B_i^T]] \\ [\delta[-YRP^{-1}][A'] + \delta Y[B_i]] & \delta^2 Y \end{pmatrix} \quad (A4)$$

We now concentrate only on diagonal elements of matrices $W = A^T(ASA^T)^{-1}A$ and $W' = (A')^T[A'S'(A')^T]^{-1}A'$. First,

$$[([(A')^T X] + [B_i^T[-YRP^{-1}]])[A'] + ([(A')^T[-P^{-1}QY]] + Y[B_i^T])[B_i]] = [(A')^T XA'] + [B_i^T[-YRP^{-1}]A'] + [(A')^T[-P^{-1}QY]B_i] + Y[B_i^T B_i] \quad (A5)$$

Now we have that $Y = (L - RP^{-1}Q)^{-1}$ (a scalar) > 0 because matrix $(ASA^T)^{-1}$ or $[A'S'(A')^T]^{-1}$ is positive definite. In addition

$$-YRP^{-1} = -Y[B_iS'(A')^T][A'S'(A')^T]^{-1}$$

$$-P^{-1}QY = -[A'S'(A')^T]^{-1}[A'S'B_i^T]Y = -Y[A'S'(A')^T]^{-1}[A'S'B_i^T]$$

$$X = P^{-1} + P^{-1}QYRP^{-1} = [A'S'(A')^T]^{-1} + Y[A'S'(A')^T]^{-1}[A'S'B_i^T][B_iS'(A')^T][A'S'(A')^T]^{-1} \quad (A6)$$

Therefore

$$[B_i^T][-YRP^{-1}][A'] = -YB_i^T B_i S'[(A')^T][A'S'(A')^T]^{-1}[A']$$

$$[(A')^T][-P^{-1}QY][B_i] = -Y[(A')^T][A'S'(A')^T]^{-1}[A']S'B_i^T B_i \quad (A7)$$

Then

$$[(A')^T XA'] + [B_i^T[-YRP^{-1}]A'] + [(A')^T[-P^{-1}QY]B_i] + Y[B_i^T B_i] = [(A')^T][A'S'(A')^T]^{-1}[A'] + Y[(A')^T] \times [A'S'(A')^T]^{-1}[A'][S'B_i^T B_i S'][(A')^T][A'S'(A')^T]^{-1}[A'] - YB_i^T B_i S'[(A')^T][A'S'(A')^T]^{-1}[A'] - Y[(A')^T] \times [A'S'(A')^T]^{-1}[A']S'B_i^T B_i + Y[B_i^T B_i] \quad (A8)$$

Now

$$[(A')^T][A'S'(A')^T]^{-1}[A'][S'B_i^T B_i S'][(A')^T][A'S'(A')^T]^{-1}[A'] = W'S'(B_i^T B_i)S'W' = W'S'BS'W' \quad (A9)$$

where $B = B_i^T B_i$ {because $W' = (A')^T[A'S'(A')^T]^{-1}A'$}. Now, the diagonal elements B_{ii} of matrix B ($m \times m$) are b_i^2, where b_i represents elements of the vector $B[b]_i$ ($1 \times m$). Matrix B is also a symmetric matrix whose element $B_{ij} = B_{ji} = b_i b_j$ (b_i, b_j are elements of the vector $B[b]_i$ and can be 0, +1, or −1). Then, the right-hand side of Eq. A8 becomes

$$[(A')^T][A'S'(A')^T]^{-1}[A'] + Y(W'S'BS'W' - BS'W' - W'S'B + B) = W' + Y(W'S'BS'W' - BS'W' - W'SB' + B) \quad (A10)$$

We have $Y > 0$. Therefore, if diagonal elements of the matrix $W'S'BS'W' - BS'W' - W'S'B + B$ are positive then we have: $W_{i_1i_1} > W'_{i_1i_1}$ because $W_{i_1i_1} = W'_{i_1i_1} + Y \times$ Diagonal elements ii of matrix $(W'S'BS'W' - BS'W' - W'S'B + B)$.

If we assume that S and S' are diagonal matrices, then diagonal element 11 of matrix $W'S'BS'W'$ is given by

$$(W'S'BS'W')_{11} = \sum_j (W'S'B)_{1j}(S'W')_{j1} = (W'_{11}S'_{11}B_{11} + W'_{12}S'_{22}B_{21} + \cdots + W'_{1m}S'_{mm}B_{m1})S'_{11}W'_{11} + (W'_{11}S'_{11}B_{12} + W'_{12}S'_{22}B_{22} + \cdots + W'_{1m}S'_{mm}B_{m2})S'_{22}W'_{21} + \cdots + (W'_{11}S'_{11}B_{1m} + W'_{12}S'_{22}B_{2m} + \cdots + W'_{1m}S'_{mm}B_{mm})S'_{mm}W'_{m1} \quad (A11)$$

Because matrix B is a symmetric matrix whose elements $B_{ij} = B_{ji} = b_i b_j$, we have

$$(W'S'BS'W')_{11} = (W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots + W'_{1m}S'_{mm}b_m)b_1S'_{11}W'_{11} + (W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots + W'_{1m}S'_{mm}b_m)b_2S'_{22}W'_{21} + \cdots + (W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots + \cdot W'_{1m}S'_{mm}b_m)b_mS'_{mm}W'_{m1} = (W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots + W'_{1m}S'_{mm}b_m)(b_1S'_{11}W'_{11} + b_2S'_{22}W'_{21} + \cdots + b_mS'_{mm}W'_{m1}) \quad (A12)$$

It can be easily verified that matrix W' or W is symmetric (given that ASA^T is symmetric, $(ASA^T)^{-1}$ is therefore also symmetric and thus $W = A^T(ASA^T)^{-1}A$ is also a symmetric matrix); in other words: $W_{ij} = W_{ji}$ or $W'_{ij} = W'_{ji}$. Therefore, we obtain

$$(W'S'BS'W')_{11} = (W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots + W'_{1m}S'_{mm}b_m)^2 \quad (A13)$$

Similarly, diagonal element 11 of matrix $\boldsymbol{W'S'B}$ is given by

$$(\boldsymbol{W'S'B})_{11} = \sum_j (\boldsymbol{W'S'})_{1j}\boldsymbol{B}_{j1} = W'_{11}S'_{11}B_{11} + W'_{12}S'_{22}B_{21} + \cdots W'_{1m}S'_{mm}B_{m1} = b_1(W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots W'_{1m}S'_{mm}b_m) \tag{A14}$$

Diagonal element 11 of matrix $\boldsymbol{B'S'W}$ is given by

$$(\boldsymbol{B'S'W})_{11} = \sum_j (\boldsymbol{B})_{1j}(\boldsymbol{S'W'})_{j1} = B_{11}S'_{11}W'_{11} + B_{12}S'_{22}W'_{21} + \cdots B_{1m}S'_{mm}W'_{m1} = b_1(W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots W'_{1m}S'_{mm}b_m) \tag{A15}$$

The diagonal elements B_{11} of matrix $\boldsymbol{B}$ are b_1^2. Therefore, the diagonal element 11 of matrix $\boldsymbol{W'S'BS'W'} - \boldsymbol{BS'W'} - \boldsymbol{W'S'B} + \boldsymbol{B}$ are given by

$$\begin{aligned}&(W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots W'_{1m}S'_{mm}b_m)^2 - b_1(W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots W'_{1m}S'_{mm}b_m)\\ &\quad - b_1(W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots W'_{1m}S'_{mm}b_m) + b_1^2\\ &\quad = (W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots W'_{1m}S'_{mm}b_m)^2 - 2b_1(W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots W'_{1m}S'_{mm}b_m) + b_1^2\\ &\quad = [(W'_{11}S'_{11}b_1 + W'_{12}S'_{22}b_2 + \cdots W'_{1m}S'_{mm}b_m) - b_1]^2 \geq 0\end{aligned} \tag{A16}$$

In general, the diagonal element ii of matrix $\boldsymbol{W'S'BS'W'} - \boldsymbol{BS'W'} - \boldsymbol{W'S'B} + \boldsymbol{B}$ is

$$[(W'_{i1}S'_{11}b_1 + W'_{i2}S'_{22}b_2 + \cdots + W'_{ii}S'_{ii}b_i + \cdots + W'_{im}S'_{mm}b_m) - b_i]^2 \geq 0$$

Therefore, the following holds:

$$W_{ii} = W'_{ii} + Y[(W'_{i1}S'_{11}b_1 + W'_{i2}S'_{22}b_2 + \cdots + W'_{ii}S'_{ii}b_i + \cdots + W'_{im}S'_{mm}b_m) - b_i]^2 \geq W'_{ii} \tag{A17}$$

Q.E.D.

Manuscript received Sept. 14, 2005, and revision received July 6, 2006.