Pergamon

Computers chem. Engng Vol. 20, No. 6/7, pp. 813–819, 1996

0098-1354(95)00180-8
0098-1354/96 $15.00 + 0.00

ON THE PROBABILITY DISTRIBUTION AND RECONCILIATION OF PROCESS PLANT DATA†

MIGUEL J. BAGAJEWICZ‡
Simulation Sciences Inc., 601 S. Valencia Ave, Brea, CA 92621, U.S.A.

(Received 11 April 1995)

Abstract—In all methods of plant data reconciliation, measurement errors have been assumed to be random variables with normal distribution. However, the propagation of measurement errors through different devices is a nonlinear process that distorts the probability distribution of the original signal. The reconciliation problem mathematical formulation is here redefined to take into account these distortions. Comparison with state of the art techniques shows that there are sizable differences of practical importance.

INTRODUCTION

Data Reconciliation is usually performed using a constrained quadratic optimization problem (Mah, 1990; Madron, 1992). In particular, material balance reconciliation is a linearly constrained quadratic problem for which an analytical solution can be obtained. The problem is usually represented by:

$$\text{P1} = \text{Min}(\bar{y}-x)^T S_y^{-1}(\bar{y}-x)$$
$$\text{s.t. } f(x,u)=0 \qquad (1)$$

where $\bar{y}$ are the measurements of the variables x, u are the unmeasured (observable and unobservable) variables and S_y is the covariance matrix. In turn $f(x, u)$ represents the plant model.

Problem (P1) relies on the assumption that $e_y = (\bar{y}-x)$ is a random variable with normal distribution and it can be derived using the maximum likelihood principle (Mah *et al.*, 1976; Madron, 1992). This assumption is based on the argument that measurement errors are a composite of disturbances ("noise") from many sources and as the number of these sources is large, the Central Limit Theorem applies. Nevertheless Mah (1990) recognizes that this assumption "is necessitated by the lack of data" and that "it appears to be a reasonable first assumption, unless the data indicate otherwise". Maldron (1992) adds that "... in the case that the distribution of errors is not known (which is a frequent case, indeed), assuming normality of distribution, one is adding only minimum information (which moreover need not be correct) to the problem". Lately, concerns have been raised about the validity of this assumption.

† Presented at the Process Systems Engineering Conference, PSE'94 Kyongju, Korea.
‡ Present Address: University of Oklahoma, 100 E Boyd, T-335, Norman OK 73019-0628, USA.

Processing of normally distributed electronic signals involves very often nonlinear transformations included in the chain of instruments, which render their distribution non-normal. As a consequence, even in the case where the originally measured signal is normally distributed, problem (P1) may no longer be the appropriate model for data reconciliation. This is the central focus of this paper. Unfortunately, most of the flow measurement is performed using orifice flow meters, which involve taking the square root of a pressure drop signal, so that even the simplest of the reconciliation problems, the material balance reconciliation problem, needs revision.

ERROR PROPAGATION AND PROBABILITY DISTRIBUTION

Chemical engineers have been trained to think about errors in measurments in the framework of control theory. Within this theory, the magnitude of the errors is of importance, and seldom its distribution is considered. Additionally, as errors in measurment are usually small in relation to the magnitude of the measured quantity itself, linearization has been proven to be a useful tool, both conceptually and practically. The distribution of errors does not get distorted if the transformations are linear. However, signal processing is nonlinear, and therefore the probability distributions of measured variables do get distorted. Madron (1992), recognizes the difficulty, but justifies ignoring it on the basis that "... errors usually are small, and quite a number of functions may be viewed as approximately linear within this limited region". Hitherto, there is no analysis in the literature regarding when is the normal distribution assumption correct and what to do when it is not. In this section of the paper the propagation of errors as the signal is processed is analyzed with emphasis on the changes in distribution.

Measured variables are classified into two categories: Directly Measured or Primary Variables and Indirectly "Measured" or Secondary Variables. Primary variables are those for which the signal obtained from the measuring device is subject only to linear or "almost linear" transformations. An example is temperature measured by a thermocouple. The signal is only amplified and linearly transformed. Indirectly measured variables are those that rely on the measurement of the other variables and involve a nonlinear transformation of the signals obtained by the measuring devices. One example is a liquid flow rate obtained using an orifice plate flow meter. A signal representing pressure drop is amplified and the flow rate is obtained by taking a square root of the signal. Multivariable nonlinear transformations are used in the case where temperature, pressure and pressure drop are used to calculate the flow rate of a gas stream, or when temperature is used "to compensate" liquid flows.

Errors of directly measured variables

Let us turn our attention to the case of the direct measurement of a single state variable. Since state variables are stochastic variables, we say that $\tilde{x}_i = x_i + e_{xi}$, i.e. the instantaneous value $\tilde{x}_i$ fluctuates around a mean value x_i following a normal distribution with variance σ_{xi}. This mean value x_i is what the literature accepts as "true" value. The departure e_{xi} from this mean ("true") value has various sources and it is assumed to follow a normal distribution with variance σ_{xi}. These are true process fluctuations. For example, in the case of pressure, its value will be affected by small fluctuations originated in pumps and/or compressors' vibrations, as well as other factors. Process temperatures are affected by ambient temperature fluctuations, etc. Turbulent flow is modeled in this way and even laminar flows are subject to such variations, since flow is driven by pressure differences and depends on density.

These flucutations are classified into two categories: deterministic and stochastic (Madron, 1992). In this paper, e_{xi} will be consisdered to be random with normal distribution, i.e., deterministic fluctuations as those obtained from poorly designed control loops, are ignored. Randomness with normal distribution is assumed to arise from the assumption that e_{xi} is the product of innumerable sources and consequently the central limit theorem applies. Furthermore, as σ_{xi} is usually not known, it will be neglected, assuming that bigger errors are introduced in the measurment process and the signal processing that follows. This has been the assumption, often implicit, of all reconciliation methods. Examination of the validity of these two assumptions is left for future work.

Measurement values $\tilde{y}_i$ are related to state variables x_i through a series of signal transformations. These transformations involve the use of measurement devices, transducers, electronic amplifiers, and final reading instruments, among others. Each of these devices adds a random error, sometimes called "noise" in signal processing. This noise is usually assumed to have a normal distribution. This is also assumed in this paper, but sometimes it is not true. For example, when noise has a narrow frequency band width it has a Rayleigh distribution (Brown and Glazier, 1964). The consequences of this type of noise error probability distribution are not explored here.

Figure 1 shows schematically a block diagram which generalizes conceptually the signal processing involved in the delivery of process measurement data. A new variable $\tilde{z}_i$ is obtained as an output of the *measuring device*. Let us call these *measurement* variables. The relationship between process variables and measurement variables is known from first principles and/or experiments. As noise is "added" in the measuring device, the output is

$$\tilde{r}_i = \tilde{z}_i + e_{zi} = m_i(\tilde{x}_i) + e_{zi} \quad (2)$$

where e_{zi}, is random with zero mean and variance σ_{zi}. As the operator, a computer, the electronic circuit, or a combination thereof, obtain $\tilde{y}_i$ given $\tilde{z}_i$, they perform the following mathematical operation:

$$\tilde{y}_i = \tilde{s}_i + e_{si} = h(\tilde{r}_i) + e_{si} \quad (3)$$

where e_{si} is the *converter and final reading error*.

Usually $h(\tilde{r}_i) = K_i\tilde{r}_i + r_{oi}$. During calibration K_i and r_{oi} are set to values such that $\tilde{y}_i$ "reads" the same values as those independently measured through procedures that are more accurate.

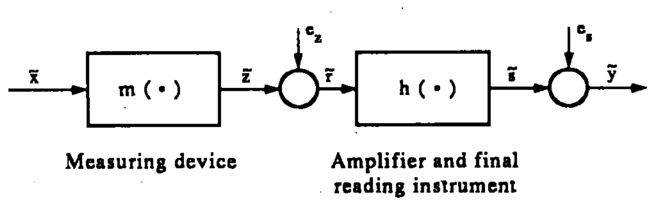


Fig. 1. Chain of instruments for directly measured data.

Linearizing $m(\bullet)$ and $h(\bullet)$, and assuming that:

- $h[m_i(x_i)] \approx x_i$
- $h'[m_i(\bar{x}_i)]m_i'(\bar{x}_i) \approx 1$
- $m_i'(\bar{x}_i)\sigma_{xi} \ll \sigma_{zi}$

one obtains

$$e_{yi} \approx K_i e_{zi} + e_{si} \tag{4}$$

$$\sigma_{yi}^2 \approx K_i^2 \sigma_{zi}^2 + \sigma_{si}^2. \tag{5}$$

It is based on this kind of linearization that e_{yi} is regarded as a sum of normally distributed errors.

Probability distribution of directly measured variables

The measuring function $m_i(x_i)$ can be modeled as a linear transformation within a certain interval (x_i^L, x_i^U) (the measuring window) and a constant outside this interval as follows:

$$m_i(x_i) = \begin{cases} z_i^L & \forall x_i \in (-\infty, x_i^L) \\ b_i x_i + c_i & \forall x_i \in (x_i^L, x_i^U) \\ z_i^U & \forall x_1 \in (x_i^U, \infty). \end{cases} \tag{6}$$

That is, below a certain value of the measured variable the instrument does not respond and gives a baseline signal z_i^L, whereas above a certain value it becomes saturated and it responds with a maximum signal z_i^U.

If the measuring window is sufficiently wide, i.e., the variance is smaller than the distance of x_i to the window bounds, we write:

$$\begin{cases} \sigma_{xi} \ll |x_i - x_i^L| \\ \sigma_{xi} \ll |x_i - x_i^U|. \end{cases} \tag{7}$$

Then the probability distribution $p_y(\bar{y})$ is nearly normal. Indeed, the probability distribution of $m_i(\bar{x}_i)$ is (O'Flynn, 1982):

$$p_{zi}(\bar{z}_i) = \frac{1}{b_i} p_{xi}\left(\frac{\bar{z}_i - c_i}{b_i}\right) H(\bar{z}_i - z_i^L) H(z_i^U - \bar{z}_i) + \Phi(x_i^L)\delta(\bar{z}_i - z_i^L) + [1 - \Phi(x_i^U)]\delta(z_i^U - \bar{z}_i) \tag{8}$$

where $H(\bullet)$ is the step function, $\delta(\bullet)$ is the delta function and $\Phi(\bullet)$ is the cumulative distribution of $\bar{x}$

$$\Phi(\bar{x}_i) = \int_{-\infty}^{x_i} p_{xi}(\xi)\,\mathrm{d}\xi. \tag{9}$$

The mean value of $\bar{z}_i$ is given by:

$$z_i = b_i x_i + c_i + \Phi(x_i^L) z_i^L + [1 - \Phi(x_i^U)] z_i^U. \tag{10}$$

As the measuring window is large, this expression tends to the usual value: $z_i^* = b_i x_i + c_i$. *This shows how an instrument with narrow width can distort the mean value of a signal.*

As e_{zi} has a normal probability distribution with zero mean $N(e_{zi}; 0, \sigma_{zi})$, the probability distribution of r_i is the convolution of the two distributions.

$$p_{ri}(\bar{r}_i) = \int_{-\infty}^{\infty} p_{xi}(\xi) N(\bar{r}_i - \xi; 0, \sigma_{zi})\,\mathrm{d}\xi. \tag{11}$$

As is shown in the appendix this distribution can be trimodal but in the case of a large measuring window it is approximated by a normal distribution, i.e.

$$p_{ri}(\bar{r}_i) \approx N[\bar{r}_i; z_i^*, (\sigma_{zi}^2 + b_i^2 \sigma_{xi}^2)^{1/2}]. \tag{12}$$

Similarly,

$$p_{yi}(\bar{y}_i) = N\{\bar{y}_i; K_i z_i^* + r_{0i}, [K_i^2(\sigma_{zi}^2 + b_i^2\sigma_{xi}^2)^{1/2} + \sigma_{si}^2]^{1/2}\}. \tag{13}$$

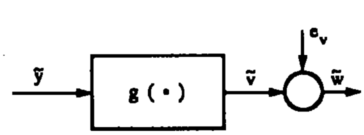
As a conclusion of the previous analysis one can say that *for directly measured state variables the assumption of normal distribution is fairly adequate as long as the measurement window is large [in the sense defined by (7)]*. In the rest of the paper it will be assumed that measurement windows are sufficiently large so that the normal distribution assumption holds. Relaxation of this assumption is left for future work.

Errors of single variable indirect measurement

Indirect measurements are very often present in process plants. One of the best examples is the measurement of liquid mass flows. In order to obtain the mass flow, one measures other variables, such as pressure differences. If density variations are negligible, one uses the expression $F = k\sqrt{\Delta p}$ to obtain flows. Extent of reactions are obtained from composition data through a linear relationship. Evaluation of densities of pure component liquids, which is often obtained from temperature measurements, is another example.

Let $\bar{w}_i$ be the *indirectly obtained measured variable*. Then (Fig. 2)

$$\bar{w}_{il} = \bar{v}_i + e_{vi} = g_i(\bar{y}_i) + e_{vl} \tag{14}$$



Nonlinear processing device

Fig. 2. Nonlinear transformation of measured data.

where $g_i(\bullet)$ is the corresponding functional relationship. Linearization and calibration conditions provide the following:

$$\tilde{w}_i \approx g_i(x_i) + g_i'(x_i)e_{yi} + e_{vi}$$
$$\approx g_i(x_i) + g_i'(x_i)\{K_i e_{zi} + e_{si}\} + e_{vi} \quad (15)$$
$$\sigma_{wi}^2 \approx [g_i'(x_i)]^2 \sigma_{yi}^2 + \sigma_{vi}^2$$
$$\approx [g_i'(x_i)]^2 [K_i \sigma_{zi}^2 + \sigma_{si}^2] + \sigma_{vi}^2. \quad (16)$$

The true value of σ_{wi}, however, should be obtained from the proper integration of the probability distribution function.

Probability distribution of single variable indirect measurement

In turning our attention to the probability distribution we find that $p_{vi}(\tilde{w}_i)$ can no longer be approximated by a normal distribution. Indeed

$$p_{vi}(\tilde{v}_i) = \left[\frac{\partial g_i^{-1}(\tilde{v}_i)}{\partial \tilde{v}_i}\right] p_{yi}[g_i^{-1}(\tilde{v}_i)] \quad (17)$$

and

$$p_{wi}(\tilde{w}_i) = \int_{-\infty}^{\infty} p_{vi}(\xi) N(\tilde{w}_i - \xi; 0, \sigma_{wi})\, d\xi. \quad (18)$$

Even if $p_{yi}(\tilde{v}_i)$ is assumed normal, unless $g_i(\bullet)$ is linear, $p_{vi}(\tilde{w}_i)$ is not normal. Moreover, it may not be available in closed analytical form. Finally, it can be even more complicated if windows for these nonlinear transformations are introduced, as done for the measuring device.

Unfortunately, as flow rates are measured indirectly, this renders the simplest of the reconciliation problems, the material balance reconciliation problem no longer analytically solvable. Not only is this a conceptual difference, but, as we shall see later, it leads to alternate solutions of the reconciliation problem that are of practical significance.

Multivariable indirect measurement

The analysis becomes even more complex if the indirectly measured variables are obtained by manipulating more than one state variable. This is the case of flow rate measurement when temperature corrections are made. In such case:

$$\tilde{w}_i = \tilde{v}_i + e_{vi} = G(\tilde{y}) + e_{vi} \quad (19)$$

where $G(\bullet)$ is a multivariable function. The probability distribution of these variables are then obtained by integrating all the appropriate signal distributions in a form similar to (17), using now the jacobian of $G(\bullet)$ instead of simple derivatives.

Variable classification

Variables such as temperature and pressure are usually direct measurements, and provided their measuring window is large, they can be treated as normally distributed. Level measurements can have intrinsic non-normal distributions: Madron (1992) gives an example of a uniform distribution for level measurements. Concentrations are usually indirectly measured. In many cases, laboratory procedures involving many steps are used and nonlinear transformations (logarithmic, for example) are often applied.

THE RECONCILIATION PROBLEM

When indirectly measured variables are present (P1) is not correct. Therefore an alternative representation that makes use of a proper likelihood function is needed. To obtain this generalized formulation, let $\tilde{\alpha} = (\alpha_y \alpha_w)^T$ be the set of all measurements (direct and indirect), i.e.,

$$\tilde{\alpha} = \begin{cases} \alpha_y = \tilde{y} & \text{for directly measured variables} \\ \alpha_w = \tilde{w} & \text{for indirectly measured variables.} \end{cases} \quad (20)$$

Let also $W(\alpha)$ be defined as follows:

$$W(\tilde{\alpha}) = \begin{pmatrix} \tilde{\alpha}_y \\ G(\tilde{\alpha}_y) \end{pmatrix}. \quad (21)$$

Therefore the following is a more appropriate formulation of the reconciliation problem

$$\text{P2} = \text{Max } L_\alpha(x \mid \tilde{\alpha})$$
$$\text{s.t. } f[W(x), u] = 0 \quad (22)$$

where L_α is the likelihood function of x given $\tilde{\alpha}$.

This form of the reconciliation problem has the advantage of having a quadratic objective function, as the likelihood function can be approximated by a product of error functions [as in (P1)], provided the measuring windows are large. In practice however, not all the information is readily available, i.e., $\tilde{\alpha}_y$ may not be known completely, $\tilde{\alpha}_w$ being given instead. In such cases the folowing is the proper reconciliation problem:

$$\text{P3} = \text{Max } L_w(W \mid \tilde{w})$$
$$\text{s.t. } f(W, u) = 0. \quad (23)$$

Problem (P3) involves the maximization of L_w, and even in the case of independent probability distributions, where L_w is obtained as a product of all probability distribution functions given by (19), its analytic expression may be hard or even impossible to obtain. The ideal situation is that all measurements be provided before any nonlinear transformation of $\tilde{y}$, so that (P2) can be used. In process plant data reconciliation, the state variables that are usually reconciled are: flow rate, pressure, temperature, and concentration. All other variables such as enthalpies, enthropies and fouling factors and other equipment parameters, are unmeasured and therefore lumped in u. Pressures and temperatures are usually directly measured. Concentrations can be considered as directly measured, and usually there is no signal processing involved and rather a laboratory procedure. These values can be considered having random errors with normal distribution, as they involve a statistical calibration procedure. This leaves only flow rates as indirectly measured variables. In very simple cases, temperature and pressure corrections are ignored. This is the case of liquid flow rate measurment of streams for which the composition and the temperature range are known $(F_i = k_i\sqrt{\Delta p_i})$.

Therefore, we can classify $G_i(\bar{a}_y)$ into one to one transformations $G_i(\bar{a}_i) = G_i(\bar{y}_i)$, and multivariable transformations. Assume that for each $G_i(\bar{a}_y) = G_i(\bar{a}_{y1}, \bar{a}2, \ldots, \bar{a}_{yn})$ all but one, say $\bar{a}_{yi}$ are known. Then we can define

$$G_i(\bar{a}_y) = G_i(\bar{a}_{yi}, \bar{w}_i) \quad \forall j \neq i. \tag{24}$$

This is the case of gas flow measurements, where the directly measured variables, temperature and pressure are normally known, and pressure drop is usually not recorded. For such cases, one can define the following approximation (P2)

$$\begin{aligned} &\text{Max } L_{yw}[x \mid \beta(\bar{w}, \bar{y})] \\ &\text{s.t. } f[W(x), u] = 0 \end{aligned} \tag{25}$$

where

$$\beta(\bar{w}, \bar{y}) = \begin{cases} \bar{a}_{yi} & \text{if } y_i \text{ is known} \\ a_{yi} = G_i^{-1}(\bar{a}_{yi}, \bar{W}_i) & \text{if } y_i \text{ is not known.} \end{cases} \tag{26}$$

Finally:

$$L_{yw}[x \mid \beta(\bar{w}, \bar{y})] = \frac{\exp(-\frac{1}{2} e_y^T S_{yw}^{-1} e_y)}{(2\pi)^{1/n} |S_{wy}|^{1/2n_y}}. \tag{27}$$

The variance matrix's diagonal elements are:

$$[S_{wy}]_{ii} = \sigma_{yi}^2 = \frac{\sigma_{wi}^2 - \sigma_{vi}^2}{\left(\dfrac{\partial W(\bar{a}_i)}{\partial y_i}\right)^2} \tag{28}$$

as suggested by (16).

However, (28) is not usually available ($\bar{a}$ is not completely known), so that one needs to use $G_i^{-1}(\bullet)$, as follows:

$$[S_{wy}]_{ii} = \frac{(\sigma_{wi}^2 - \sigma_{vi}^2)}{\left(\dfrac{\partial W[G_i^{-1}(\bar{a}_j, w_i)]}{\partial y_i}\right)^2} \tag{29}$$

or

$$[S_{wy}]_{ii} = \frac{\sigma_{vi}^2}{\left(\dfrac{\partial W[G_i^{-1}(\bar{a}_j, w_i)]}{\partial y_i}\right)^2}. \tag{30}$$

Equations (29) and (30) offer two alternatives. Sometimes one has knowledge of σ_{yi} so that the second form can be used, but usually σ_{wi} is the only value given in practice. In those cases the first form is used and an estimate of σ_{vi} is needed.

Comparisons between this new approach and classical results are given next for the case of flow rate measurements. Other variables (temperature, pressure, level and concentrations) are not explored in this paper.

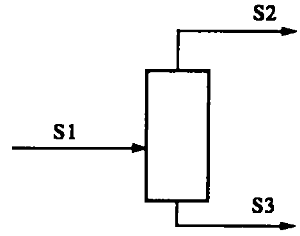


Fig. 3. Example.

RESULTS

Consider a simple splitter or mixer as in Fig. 3. For a material balance reconciliation problem (P1) is written as follows:

$$\begin{aligned} &\text{PF1} = \text{Min} \sum_{\forall i} \left(\frac{\bar{F}_i - F_i}{\sigma_{fi}}\right)^2 \\ &\text{s.t. } \sum_{i=2}^{3} F_i - F_1 = 0. \end{aligned} \tag{31}$$

Assume for simplicity that the flow rates are obtained through $F_i = k_i\sqrt{\Delta p_i}$. If the measurements $\Delta\bar{p}_i$ are available and they are normally distributed, as can be assumed since they are primary variables, then they can be reconciled using:

$$\begin{aligned} &\text{PF2} = \text{Min} \sum_{\forall i} \left(\frac{\Delta\bar{p}_i - \Delta p_i}{\sigma_{\Delta pi}}\right)^2 \\ &\text{s.t. } \sum_{i=2}^{3} k_i\sqrt{\Delta p_i} - k_1\sqrt{\Delta p_1} = 0. \end{aligned} \tag{32}$$

In the absence of pressure drop measured values, $\Delta\bar{p}_i$, one can approximate them using the measured flow values, i.e. $\Delta\bar{p} \approx (F_i/k_i)^2$. If on top of it $\sigma_{\Delta pi}$ is not known, one can use (28) and a reasonable estimate of the added noise standard deviation σ_{vi} to

Table 1. Example 1. Comparative results of problems (PF1) and (PF3)

Measured flows			(PF1)			(PF3)		
$\tilde{S}1$	$\tilde{S}2$	$\tilde{S}3$	S1	S2	S3	S1	S2	S3
85.21	48.70	39.90	86.87	47.75	39.12	86.29	47.32	38.97
85.21	21.78	65.56	86.26	21.51	64.75	85.89	20.89	65.00
85.21	10.89	76.23	86.15	10.77	75.38	85.81	10.08	75.73

write $\sigma_{\Delta pi} \approx 2(\sigma_{Fi}^2 - \sigma_{vi}^2)^{1/2}$. Then the following problem, called here the *Approximation Problem*, is obtained by substituting the above approximations in (PF2). The result, expressed in terms of flows, rather than pressure drops is:

$$\mathrm{PF3} = \mathrm{Min} \sum_{\forall i} \left(\frac{\tilde{F}_i^2 - F_i^2}{2(\sigma_{Fi}^2 - \sigma_{vi}^2)^{1/2} \tilde{F}_i} \right)^2$$

$$\text{s.t.} \ \sum_{i=2}^{3} F_i - F_1 = 0. \tag{33}$$

Comparison of (PF1) and (PF3) reveals that they have the same set of constraints differing only in their objective function. In the case of (PF3) the individual terms of the objective function are the square of the difference of the squares of the measured and reconciled flow rates, whereas in the case of (PF1) these terms are the square of the difference of these values. If all weights are equal, one can easily show that larger flow rates will be subject to smaller corrections in the case of (PF3) than in the case of (PF1).

Table 1 illustrates the results. Three cases of the same example are displayed. In each case $\tilde{S}1$ was maintained the same, whereas $\tilde{S}2$ was varied. The same value of relative variance was maintained for all streams in all three cases. Results of (PF1) and (PF3) are displayed confirming the aforementioned tendency of higher flow rates to be less corrected in the case of (PF3). Additionally one can see that as $\tilde{S}2$ is smaller, the relative difference between S2 obtained from (PF1) and (PF3) grows from around 1% to around 8%.

The practical impact of the above results is significant. For example, consider the case where S2 is a valuable product, and S3 is a cheap by-product. In such a case S2 is usually small compared to S1 and S3 and the differences in cost may be even greater.

CONCLUSIONS

It was shown in this paper that normally random signals are non linearly distorted when processed to obtain the final reading values. It has been assumed that for directly measured variables (primary variables) the assumption of normal distribution can still be used, as long as the measuring device has an acceptably *quasi*-linear input–output function with a wide measuring window. However, non-linear transformations, such as taking the square root of a pressure drop on an orifice flow meter distort the probability distribution and the assumption does not hold any longer.

The Process Data Reconciliation problem was then discussed showing that the state of the art models should be revisited. An alternative approach, which takes into account some of the most relevant aspects of the distortions is proposed and illustrated on a material balance reconciliation problem. The examples show that the differences in results obtained can be of practical importance.

REFERENCES

Brown J. and E. V. D. Glazier, *Signal Analysis*. Reinhold, New York (1964).

Madron F., *Process Plant Performance, Measurement Data Processing for Optimization and Retrofits*. Ellis Horwood, West Sussex (1992).

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

Mah R. S. H., G. M. Stanley and D. M. Downing, *Reconciliation and Rectification of process flow and inventory data. Ind. Engng Chem. Process Des. Develop.* **15**, 175–183 (1976).

O'Flynn M., *Probability Random Variables and Processes*. Harper & Row, New York (1982).

APPENDIX

Using (8) one obtains:

$$p_{ri}(\tilde{r}_i) = \int_{-\infty}^{\infty} p_{xi}\left(\frac{\xi - c_i}{b_i}\right) N(\tilde{r}_i - \xi; 0, \sigma_{zi}) \, d\xi$$
$$+ \Phi(x_i^L) N(\tilde{r}_i; z_i^L, \sigma_{zi}) + [1 - \Phi(x_i^U)] N(\tilde{r}_i; z_i^U, \sigma_{zi}). \tag{A1}$$

As x_i has a normal distribution $p_{ri}(\tilde{x}_i) = N(\tilde{x}_i; x_i, \sigma_{xi})$ it reduces to

$$p_{ri}(\tilde{r}_i) = \frac{\exp\left(\frac{(\tilde{r}_i - b_i x_i - c_i)^2}{2(\sigma_{zi}^2 + b_i^2 \sigma_{xi}^2)}\right)}{\sqrt{2\pi}(\sigma_{zi}^2 + b_i^2 \sigma_{xi}^2)^{1/2}} \{\mathrm{erfc}(A_i^U) - \mathrm{erfc}(A_i^L)\}$$
$$+ \Phi(x_i^L) N(\tilde{r}_i; z_i^L, \sigma_{zi})$$
$$+ [1 - \Phi(x_i^U)] N(\tilde{r}_i; z_i^U, \sigma_{zi}) \tag{A2}$$

where

$$A_i^j=(x_i^j-x_i)\frac{(\sigma_{zi}^2+b_i^2\sigma_{xi}^2)^{1/2}}{\sigma_{xi}\sigma_{zi}}-\frac{(\bar{r}_i-b_ix_i-c_i)b_i\sigma_{xi}}{\sigma_{zi}(\sigma_{zi}^2+b_i^2\sigma_{xi}^2)^{1/2}} \quad j=L,U. \tag{A3}$$

As $p_{ri}(\bar{r}_i)$ is now given by the sum of three terms with maximum values near z_i^L, z_i^* and z_i^U, this distribution is trimodal. Moreover, due to the erfc terms, it is not symmetric. However, when σ_{xi} is small compared to the measuring window and x_i is away from the window boundaries, then $\mathrm{erfc}(A_i^L)\ll 1$, $\mathrm{erfc}(A_i^U)\approx 1$, $\Phi(x_i^L)\ll 1$ and $\Phi(x_i^U)\approx 1$, which renders $p_{ri}(\bar{r}_i)$ nearly normal, i.e.,

$$p_{ri}(\bar{r}_i)\approx N(\bar{r}_i; z_i^*, (\sigma_{zi}^2+b_i^2\sigma_{xi}^2)^{1/2}) \quad \text{if} \begin{cases}\sigma_{xi}\ll|x_i-x_i^L| \\ \sigma_{xi}\ll|x_i-x_i^U|\end{cases}. \tag{A4}$$

A similar manipulation can be done to show that for a large measuring window:

$$p_{ri}(\bar{y}_i)\approx N\{\bar{y}_i; K_iz_i^*+r_{0i}, [K_i^2(\sigma_{zi}^2+b_i^2\sigma_{xi}^2)^{1/2}+\sigma_{si}^2]^{1/2}\}. \tag{A5}$$