Integral Approach to Plant Linear Dynamic Reconciliation

Miguel J. Bagajewicz and Qiyou Jiang
School of Chemical Engineering and Material Science, University of Oklahoma, Norman, OK 73019

An integral method is proposed that performs dynamic data reconciliation on linear systems, in contrast with recent methods that utilize differential algebraic equations. The differential equations representing this system are first rearranged to obtain a system of equations containing only redundant measurements. These equations are formally integrated using polynomial approximations, and the reconciliation is then performed using analytical solutions. A new statistic to detect gross errors is proposed, and a procedure to detect biased measurement is presented.

Introduction

Available commercial software for plant data reconciliation uses a steady-state model for data reconciliation. To use this software, it is common practice to use straightforward averages of several measurements of the different plant variables taken in a certain period of time. In doing this, several penalties are paid. System fluctuations are lumped with instrument biases and inherent instrument white noise. The detection of outliers, biases, and leaks is thus difficult. Additionally, variance estimation is obscured. Instrument variance can be obtained directly from serially correlated data, but it cannot be directly used in steady-state data reconciliation because departures from steady state affect the variance estimates. A few studies have addressed this problem (Almasy and Mah, 1984; Keller et al., 1992). Also, the issue of the probability distribution of data cannot be addressed satisfactorily when averaging is performed and steady state is assumed (Bagajewicz, 1996). Finally, the steady-state assumption of current data-reconciliation technology has been disputed in several forums, from practitioners to academia (Bagajewicz and Mullick, 1995).

Even though steady-state data reconciliation has been able to perform rather well in practice, it is in the realm of gross-error detection that the steady-state model has not been able to perform successfully, in spite of all the different tests that have been proposed. In view of this frustration with the low power of these tests, recent work in this field is moving in the direction of sequential analysis (Tong and Crowe, 1996), where data from different days is sequentially analyzed with different statistical techniques, aiming at the identification of biases and leaks.

The obvious alternative to steady-state models is to rely on reconciliation methods based on dynamic models. For the case of linear systems one-step integration-reconciliation procedures, many of them rooted on Kalman filtering (Kalman, 1960), have been proposed. Stanley and Mah (1977) showed how Kalman filtering could be adapted to take advantage of spatial and temporal redundancy in a quasi-steady-state condition. Darouach and Zasadzinski (1991) proposed a backward difference approximation and recursive technique to solve the constrained least-square optimization problem. Rollins and Devanathan (1993) improved on the estimation accuracy using a maximum-likelihood function and proposing two estimators that are later averaged.

The preceding summarized line of work on local estimators is useful for control and monitoring purposes, and in the absence of gross errors (biases and leaks) they all perform rather well. However, the detection of gross errors shares some of the difficulties of tests based on steady-state models. Narasimhan and Mah (1988) proposed applying the generalized likelihood ratio (GLR) to dynamic situations with small departures from steady-state values. Kao et al. (1992) studied the effect of serially correlated data on gross-error detection. They proposed composite test procedures based on window averages, prewhitening procedures, and the GLR.

Setting aside the computational volume challenges, these methods, as will any one-step procedure, at best ameliorate the fluctuations of measurement data, but rarely produce a smooth profile. For plant day-to-day management and economics such fluctuating patterns are not satisfactory. Indeed,

Correspondence concerning this article should be addressed to Miguel J. Bagajewicz.

a fluctuating rather than a smooth reconciled pattern does not provide a simple data description, so, for example, daily balances cannot be performed without further processing. Smooth stream profiles, based on a few parameters, are amenable for integration, and are thus easy to use and inexpensive to store.

Other methods have been presented for dynamic data reconciliation. Ramamurthi and Bequette (1991) proposed a technique based on a successively linearized horizon; Liebman et al. (1992) used orthogonal collocation; Karjala and Himmelblau (1996) rely on neural networks; and Albuquerque and Biegler (1995) on a discretization of the DAE system using Runge-Kutta methods. Albuquerque and Biegler (1996) also proposed gross-error detection techniques as an extension of their discretization approach. All these methods have the same aforementioned lack of smoothness.

This article focuses on presenting an integral approach to dynamic data reconciliation. First, the model differential equations are integrated formally. Then, the variables are parametrized and the reconciliation is performed using all time serially correlated measurements. Bias detection is discussed and the detection of one biased instrument is presented. Future work will address other important aspects of the problem.

One of the myths that circulate in the software industry is that dynamic data reconciliation is difficult, computationally too intensive, and impractical. This article addresses the first two issues, as it shows that the solution offered is straightforward and computationally reasonable. The issue of practicality stems from the improved precision of data and the effectiveness of gross error detection. Although results presented in this article are encouraging, no final conclusion can be made until this technique is tested in practice. Some of the issues facing future work are briefly discussed toward the end of the article.

Integral Model for Linear Reconciliation

The dynamic model of a material balance in a process plant can be represented by the following differential-algebraic system of equations (DAE):

$$\frac{dv}{dt} = Af \tag{1}$$

$$Cf = 0. \tag{2}$$

In practice, there are usually four different kinds of variables in data processing. Measured variables are classified as redundant and nonredundant (in the spatial sense), whereas nonmeasured variables are classified as observable and unobservable. This classification, which was developed for steady-state data reconciliation, holds also for the dynamic case. Values for redundant variables are estimated using data-reconciliation, typically maximum-likelihood procedures. It is a common practical (but not theoretical) assumption in steady-state data reconciliation that measurements are independent, that is, there is no covariance between different variables. In those cases, nonredundant variables have to be accepted at measured face value. When covariances are not zero, adjustments of nonredundant variables can be made. In the dynamic case, since several measurements are available it is said that they present temporal redundancy. As we shall see later, this redundancy allows certain fitting or filtering to be performed. We call these variables, self-redundant.

Observable variables can be calculated from the redundant and/or nonredundant quantities, while the unobservable variables cannot. It is desired to break down the system of Eqs. 1 and 2, in subsystems of equations, so that a system of equations containing only redundant variables is obtained. Let

$$D = \begin{bmatrix} A & -I \\ C & 0 \end{bmatrix} \tag{3}$$

and

$$x = \begin{pmatrix} f \\ \dfrac{dv}{dt} \end{pmatrix}. \tag{4}$$

Thus the system of differential algebraic equations (Eqs. 1 and 2) can be rewritten as follows:

$$Dx = 0. \tag{5}$$

Madron (1992) showed that by using simple linear combination and rearrangement of rows as well as column reordering in matrix D, the system can be rewritten in the following way:

$$\begin{bmatrix} I & 0 & -E_{RO} & -E_{SRO} \\ 0 & E_{UO} & E_{RUO} & E_{SRUO} \\ 0 & 0 & E_R & 0 \end{bmatrix} \begin{bmatrix} x_O \\ x_{UO} \\ x_R \\ x_{SR} \end{bmatrix} = 0. \tag{6}$$

This matrix is called the canonical form of D. Data reconciliation can be performed on redundant variables only, that is, using

$$E_R x_R = 0. \tag{7}$$

However, the matrix E_R has the following structure

$$E_R = \begin{bmatrix} A_R & -B_R \\ C_R & 0 \end{bmatrix}. \tag{8}$$

Therefore Eq. 7 is rewritten in terms of its specific parts as follows:

$$B_R \frac{dv_R}{dt} = A_R f_R \tag{9}$$

$$C_R f_R = 0. \tag{10}$$

The next step is to resolve the temporal redundancy of self-redundant variables. In general, this can be done using any smoothing or filtering technique. The last step, called *cooptation*, consists of computing the observable quantities using the reconciled values of the redundant variables and

the nonredundant variables. Cooptation can be performed using

$$x_O = E_{RO}x_R + E_{SRO}x_{SR}. \tag{11}$$

Example 1

Consider the system depicted in Figure 1. It contains nine streams, four tanks, and one splitter. Assume that all volumes and flow rates are measured except the volume of Tank 3 and the flow rates of $S2$ and $S6$. Matrix D is then given by

$$D = \begin{array}{c} \begin{matrix} S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & V1 & V2 & V3 & V4 \end{matrix} \\ \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}. \tag{12}$$

The canonical form of D is

$$\begin{array}{c} \begin{matrix} S2 & S6 & V3 & S4 & S5 & S7 & S8 & S9 & V4 & V1 & V2 & S1 & S3 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 1 & -1 \\ 0 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}. \tag{13}$$

As a result of the procedure, only two equations (last two rows) represent the redundant system of equations. Additionally, in this case all variables are observable. The corresponding matrices are

$$E_R = \begin{bmatrix} 1 & 0 & -1 & -1 & -1 & -1 \\ 0 & 1 & 0 & -1 & -1 & 0 \end{bmatrix} \tag{14}$$

$$E_{RO} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{15}$$

$$E_{SRO} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 0 \end{bmatrix}. \tag{16}$$

Thus $V4$, $S4$, $S5$, $S7$, $S8$, and $S9$ have been identified as redundant variables. In particular:

$$A_R = [1 \quad 0 \quad -1 \quad -1 \quad -1] \qquad B_R = 1$$
$$C_R = [0 \quad 1 \quad 0 \quad -1 \quad -1], \tag{17}$$

which correspond to the following system of equations:

$$\frac{dv_4}{dt} = f_4 - f_7 - f_8 - f_9 \tag{18}$$

$$f_5 - f_8 - f_9 = 0. \tag{19}$$

The redundant system can be shown as Figure 2.

Thus, the nonredundant (self-redundant) variables are $V1$, $V2$, $S1$, and $S3$, and the observable variables are $V3$, $S2$, and $S6$, as is indicated by the first three columns of the canonical form of D.

Polynomial representation

Consider now the following s-order polynomial representation of f_R and v_R:

$$f_R \approx \sum_{k=0}^{s} \alpha_k^R t^k \tag{20}$$

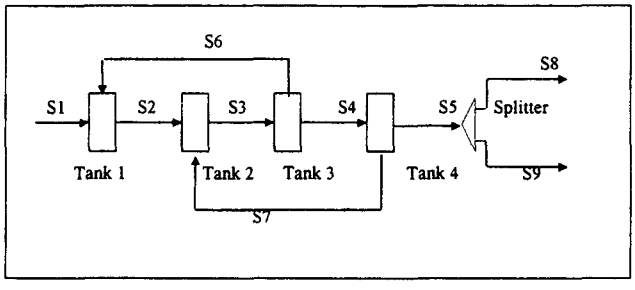

Figure 1. Example 1.

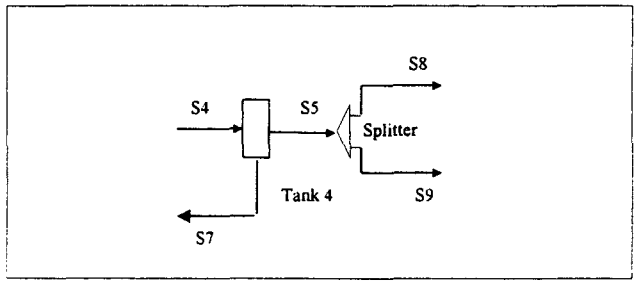

Figure 2. Redundant system of Example 1.

$$v_R \approx v_{R0} + \sum_{k=0}^{s} \omega_{k+1}^R t^{k+1}, \tag{21}$$

therefore

$$\int_0^t f_R(\xi)\, d\xi \approx \sum_{k=0}^{s} \frac{\alpha_k^R}{k+1} t^{k+1}. \tag{22}$$

Thus, Eq. 9 is equivalent to writing

$$B_R[v_R - v_{R0}] = B_R \sum_{k=0}^{s} \omega_{k+1}^R t^{k+1} = A_R \sum_{k=0}^{s} \frac{\alpha_k^R}{k+1} t^{k+1}. \tag{23}$$

Since this equation is valid for all t, then

$$B_R \omega_{k+1}^R = \frac{A_R \alpha_k^R}{k+1} \qquad k = 0, \ldots, s. \tag{24}$$

In turn, Eq. 10 is equivalent to the following set of equations:

$$C_R \alpha_k^R = 0 \qquad k = 0, \ldots, s. \tag{25}$$

Now consider the case on $n+1$ measurements and assume that the measurement errors are normally distributed around a mean value, which in turn varies with time. With this assumption the maximum of the likelihood function is obtained by solving:

$$\text{Min} \sum_{i=0}^{n} \left\{ (v_{Ri} - v_{Ri}^+)^T S_V^{-1} (v_{Ri} - v_{Ri}^+) + (f_{Ri} - f_{Ri}^+)^T S_F^{-1} (f_{Ri} - f_{Ri}^+) \right\} \tag{26}$$

s.t.

$$B_R(v_{Ri} - v_{R0}) = A_R \int_0^t f_R(\xi)\, d\xi \quad (i = 0, \ldots, n)$$
$$C_R f_{Ri} = 0 \quad (i = 0, \ldots, n),$$

or, in terms of the polynomial coefficients,

$$\text{Min} \left\{ (Jv_{R0} + T_\omega \omega^R - v_R^+)^T R_V^{-1} (Jv_{R0} + T_\omega \omega^R - v_R^+) + (T_\alpha \alpha^R - f_R^+)^T R_F^{-1} (T_\alpha \alpha^R - f_R^+) \right\} \tag{27}$$

s.t.

$$D_m \omega^R = R_m \alpha^R$$
$$C_\alpha \alpha^R = 0,$$

where

$$\alpha^R = \begin{bmatrix} \alpha_0^R \\ \alpha_1^R \\ \vdots \\ \alpha_s^R \end{bmatrix} \qquad \omega^R = \begin{bmatrix} \omega_1^R \\ \omega_2^R \\ \vdots \\ \omega_{s+1}^R \end{bmatrix} \qquad f_R^+ = \begin{bmatrix} f_{R0}^+ \\ f_{R1}^+ \\ \vdots \\ f_{Rn}^+ \end{bmatrix}$$

$$v_R^+ = \begin{bmatrix} v_{R0}^+ \\ v_{R1}^+ \\ \vdots \\ v_{Rn}^+ \end{bmatrix} \qquad J = \begin{bmatrix} I \\ I \\ \vdots \\ I \end{bmatrix}$$

$$D_m = \begin{bmatrix} B_R & 0 & \cdots & 0 \\ 0 & B_R & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_R \end{bmatrix} \qquad R_M = \begin{bmatrix} A_R & 0 & \cdots & 0 \\ 0 & \frac{A_R}{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{A_R}{s+1} \end{bmatrix}$$

$$C_\alpha = \begin{bmatrix} C_R & 0 & \cdots & 0 \\ 0 & C_R & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_R \end{bmatrix} \qquad R_V^{-1} = \begin{bmatrix} S_V^{-1} & 0 & \cdots & 0 \\ 0 & S_V^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_V^{-1} \end{bmatrix}$$

$$R_F^{-1} = \begin{bmatrix} S_F^{-1} & 0 & \cdots & 0 \\ 0 & S_F^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_F^{-1} \end{bmatrix} \qquad T_\alpha = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & t_1 I & \cdots & t_1^s I \\ \vdots & \vdots & \ddots & \vdots \\ I & t_n I & \cdots & t_n^s I \end{bmatrix}$$

$$T_\omega = \begin{bmatrix} 0 & \cdots & 0 \\ t_1 I & \cdots & t_1^{s+1} I \\ \vdots & \ddots & \vdots \\ t_n I & \cdots & t_n^{s+1} I \end{bmatrix}.$$

Rewrite this problem as follows:

$$\text{Min } z^T Q^{-1} z + w^T z \tag{28}$$

s.t.

$$Mz = 0,$$

where

$$z = \begin{pmatrix} v_{R0} \\ \omega^R \\ \alpha^R \end{pmatrix} \qquad w = -2 \begin{pmatrix} (R_V^{-1} J)^T v_R^+ \\ (R_V^{-1} T_\omega)^T v_R^+ \\ (R_F^{-1} T_\alpha)^T f_R^+ \end{pmatrix}$$

$$Q^{-1} = \begin{pmatrix} J^T R_V^{-1} J & J^T R_V^{-1} T_\omega & 0 \\ T_\omega^T R_V^{-1} J & T_\omega^T R_V^{-1} T_\omega & 0 \\ 0 & 0 & T_\alpha^T R_F^{-1} T_\alpha \end{pmatrix}$$

$$M = \begin{pmatrix} 0 & -D_m & R_m \\ 0 & 0 & C_\alpha \end{pmatrix}.$$

The solution to this problem is

$$z = \left[I - QM^T (MQM^T)^{-1} M \right] \left(-\frac{1}{2} Qw \right). \tag{29}$$

The matrix MQM^T does not depend on the measured data and for a given system can be inverted *a priori*. In the case of measurements performed at regular intervals, which is the assumption of this article, scaling can be used to perform this inversion.

Scaling

Scaling of time allows a data-independent expression for matrix T. Consider the following stretching: $t = \beta t_r$. Then

$$f_R = \sum_{k=0}^{s} \alpha_k^R \beta^k t_r^k \tag{30}$$

$$v_R = v_{R0} + \sum_{k=0}^{s} \omega_{k+1}^R \beta^{k+1} t_r^{k+1}. \tag{31}$$

Thus, on rescaled coordinates one can write

$$\alpha_{sk}^R = \alpha_k^R \beta^k \tag{32}$$

$$\omega_{sk}^R = \omega_k^R \beta^k. \tag{33}$$

Assume that the measurements are taken at equal intervals Δ. Then:

$$T_\alpha = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & \Delta I & \cdots & \Delta^s I \\ \vdots & \vdots & \ddots & \vdots \\ I & n\Delta I & \cdots & n^s \Delta^s I \end{bmatrix} \tag{34}$$

$$T_\omega = \begin{bmatrix} 0 & \cdots & 0 \\ \Delta I & \cdots & \Delta^{s+1} I \\ \vdots & \ddots & \vdots \\ n\Delta I & \cdots & n^{s+1} \Delta^{s+1} I \end{bmatrix}. \tag{35}$$

The scaling factor β is chosen to make $\Delta = 1$, that is,

$$\beta = \frac{t_n}{n}. \tag{36}$$

In this way, the inversion of $T_\alpha^T R_F^{-1} T_\alpha$ and $T_\omega^T R_V^{-1} T_\omega$ can be done analytically and independently of the data. Details of this inversion are shown in the Appendix.

Example 1 (continued)

About 100 measurements for each stream and tank were created with a random Gaussian number generator, and only 50 of them were used for data reconciliation. The coefficient of the polynomial representation of holdups and flow rates of redundant variables $V4$, $S4$, $S5$, $S7$, $S8$, and $S9$ were obtained through reconciliation. Figures 3 to 8 show the comparison between the reconciled values and the corresponding measured values for each of the redundant variables.

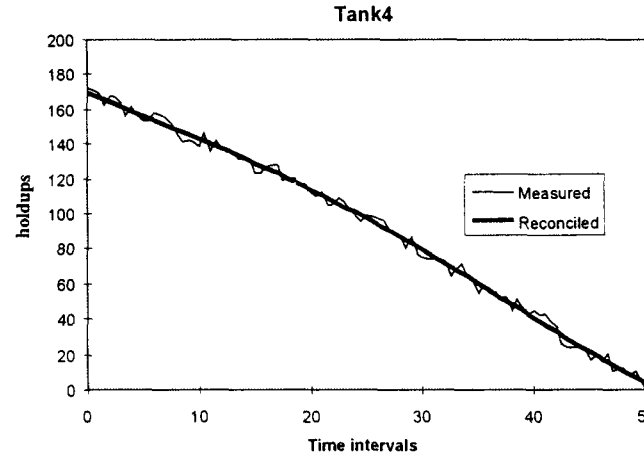

Figure 3. Example 1: reconciled and measured holdups for Tank 4.

Role of the polynomial order

The order of the polynomial plays an important role in data reconciliation. Observing the data for the stream flow rates, it is obvious that any order smaller than $s = 3$ will not lead to a successful data reconciliation, simply because the data seem to have a maximum and a minimum. The question remains whether an order higher than $s = 3$ will allow a better data reconciliation. The answer to this question relies on observing the behavior of the objective function upon reconciliation. For $s = 3$ its value of the objective function is 353.8325. As the order is increased to $s = 4$, the value of the objective function is 340.7043, a slightly lower value. This suggests the simple procedure of incrementing the order until small reductions in the objective function are seen. In addition, no bias is introduced and the differences between the two solutions are significantly smaller than those between any of these solutions and the measurements. One must be aware that, in addition, as in fitting, increasing the polynomial order unnecessarily introduces additional inflexion points in the curve. The important question of what order should be chosen and when it is proper to stop will be addressed in future work.

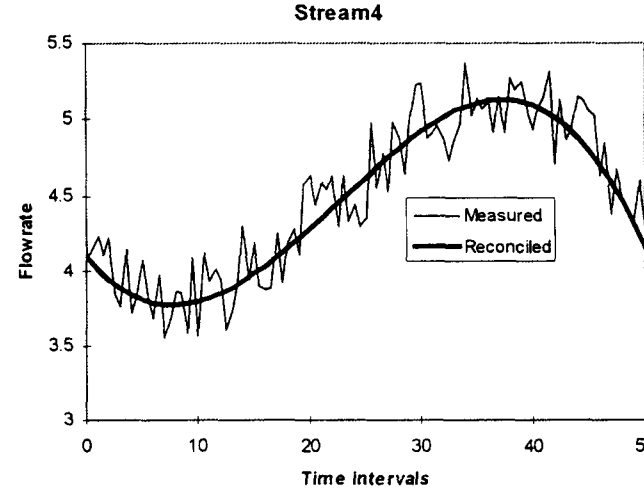

Figure 4. Example 1: reconciled and measured flow rates for Stream 4.

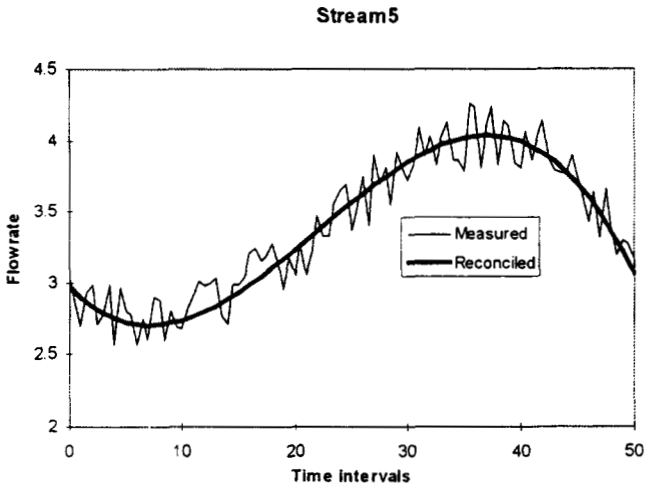

Figure 5. Example 1: reconciled and measured flow rates for Stream 5.

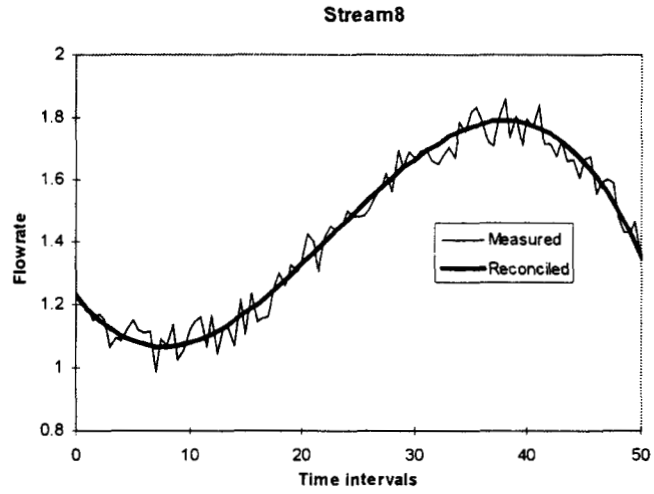

Figure 7. Example 1: reconciled and measured flow rates for Stream 8.

Self-redundant variables

Different techniques can be used to filter or smooth the temporal redundant data of each variable. For example, exponential smoothing can be used to obtain a better estimate. In this article polynomial fitting is used. The fitting model is given by

$$\text{Min}\,(U_F\alpha^{SR} - f_{SR}^{+})^T(U_F\alpha^{SR} - f_{SR}^{+}), \tag{37}$$

where

$$U_F = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & t_1 I & \cdots & t_1^s I \\ \vdots & \vdots & \ddots & \vdots \\ I & t_n I & \cdots & t_n^s I \end{bmatrix}.$$

This matrix and T_α have similar structure. Note, however, that the dimensions could be different. The solution is

$$\alpha^{SR} = [U_F^T U_F]^{-1} U_F^T f_{SR}^{+}. \tag{38}$$

A similar result can be obtained for self-redundant holdups, that is, solving

$$\text{Min}\,(U_V\omega^{SR} - v_{SR}^{+})^T(U_V\omega^{SR} - v_{SR}^{+}), \tag{39}$$

where

$$U_V = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & t_1 I & \cdots & t_1^{s+1} I \\ \vdots & \vdots & \ddots & \vdots \\ I & t_n I & \cdots & t_n^{s+1} I \end{bmatrix}.$$

In this case, for simplicity ω^{SR} contains the initial holdup. The solution is

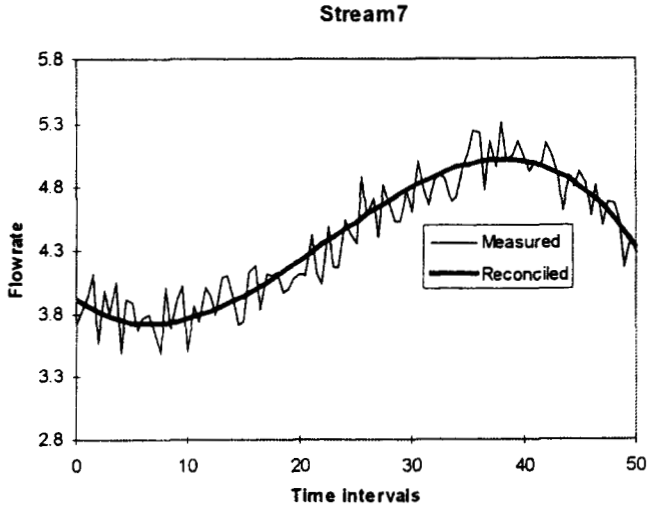

Figure 6. Example 1: reconciled and measured flow rates for Stream 7.

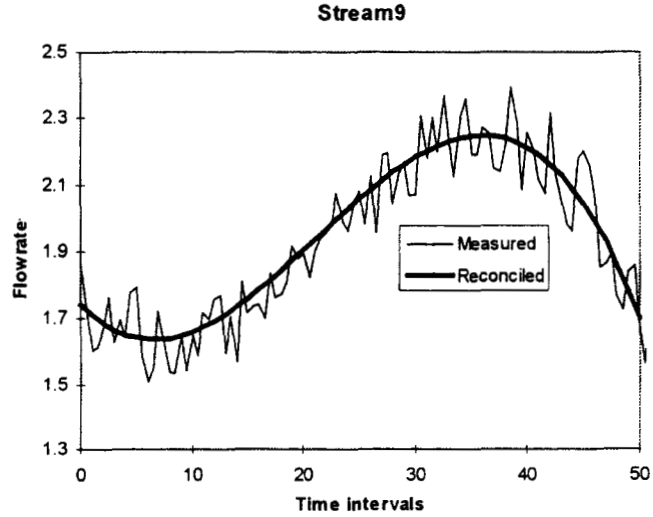

Figure 8. Example 1: reconciled and measured flow rates for Stream 9.

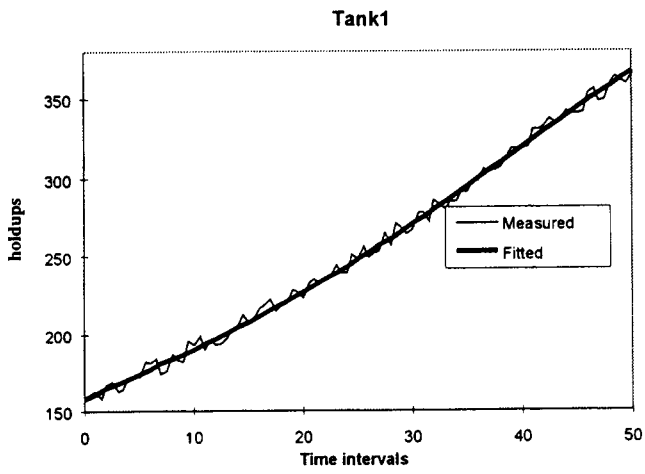

Figure 9. Example 1: fitted and measured holdups for Tank 1.

$$\omega^{SR} = [U_V^T U_V]^{-1} U_V^T v_{SR}^+. \tag{40}$$

Figures 9 to 12 show the fitting results for nonredundant variables $V1$, $V2$, $S1$, and $S3$.

Observable variables

Observable variables can be obtained from Eq. 11. This equation can be rewritten in more detail as follows:

$$\begin{pmatrix} f_O \\ \dfrac{dv_O}{dt} \end{pmatrix} = \begin{pmatrix} A_{RO1} & B_{RO1} \\ A_{RO2} & B_{RO2} \end{pmatrix} \begin{pmatrix} f_R \\ \dfrac{dv_R}{dt} \end{pmatrix} + \begin{pmatrix} A_{SRO1} & B_{SRO1} \\ A_{SRO2} & B_{SRO2} \end{pmatrix} \begin{pmatrix} f_{SR} \\ \dfrac{dv_{SR}}{dt} \end{pmatrix}, \tag{41}$$

or

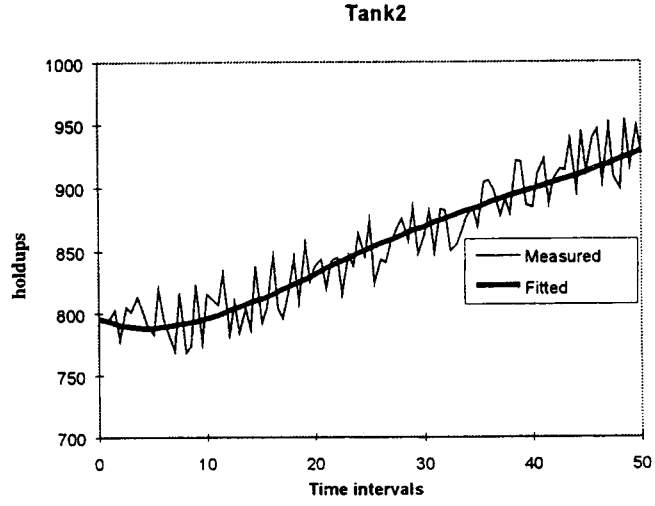

Figure 10. Example 1: fitted and measured holdups for Tank 2.

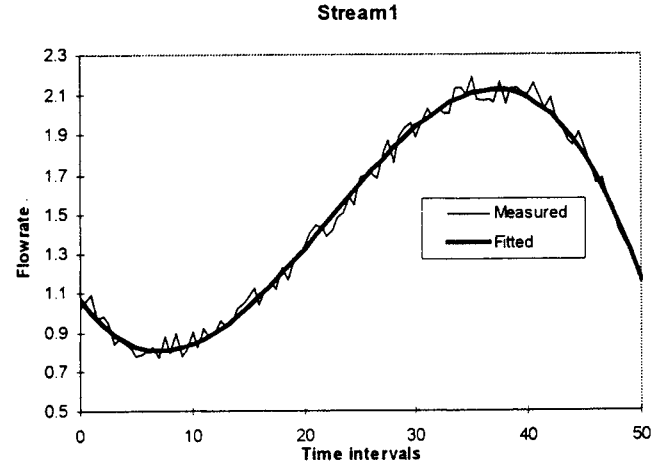

Figure 11. Example 1: fitted and measured flow rates for Stream 1.

$$f_O = A_{RO1} f_R + B_{RO1} \frac{dv_R}{dt} + A_{SRO1} f_{SR} + B_{SRO1} \frac{dv_{SR}}{dt} \tag{42}$$

$$\frac{dv_O}{dt} = A_{RO2} f_R + B_{RO2} \frac{dv_R}{dt} + A_{SRO2} f_{SR} + B_{SRO2} \frac{dv_{SR}}{dt}. \tag{43}$$

Furthermore, Eq. 42 can be rewritten in terms of the polynomial expressions as follows:

$$\sum_{k=0}^{s} \alpha_k^O t^k = A_{RO1} \sum_{k=0}^{s} \alpha_k^R t^k + B_{RO1} \sum_{k=0}^{s} (k+1)\omega_{k+1}^R t^K + A_{SRO1} \sum_{k=0}^{s} \alpha_k^{SR} t^k + B_{SRO1} \sum_{k=0}^{s} (k+1)\omega_{k+1}^{SR} t^k. \tag{44}$$

Thus

$$\alpha_k^O = A_{RO1}\alpha_k^R + (k+1)B_{RO1}\omega_{k+1}^R + A_{SRO1}\alpha_k^{SR} + (k+1)B_{SRO1}\omega_{k+1}^{SR}. \tag{45}$$

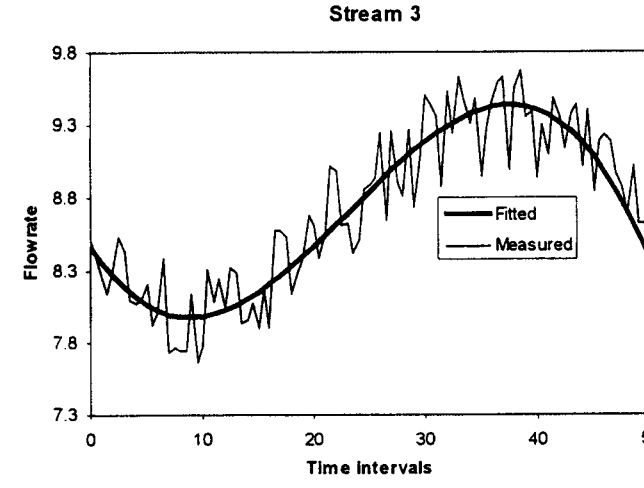

Figure 12. Example 1: fitted and measured flow rates for Stream 3.

Similarly, from Eq. 43 one obtains:

$$\sum_{k=0}^{s}(k+1)\omega_{k+1}^{O}t^{k} = A_{RO2}\sum_{k=0}^{s}\alpha_{k}^{R}t^{k} + B_{RO2}\sum_{k=0}^{s}(k+1)\omega_{k+1}^{R}t^{k} + A_{SRO2}\sum_{k=0}^{S}\alpha_{k}^{SR}t^{k} + B_{SRO2}\sum_{k=0}^{s}(k+1)\omega_{k+1}^{SR}t^{k}, \quad (46)$$

or

$$\omega_{k}^{O} = A_{RO2}\frac{\alpha_{k}^{R}}{k+1} + B_{RO2}\omega_{k+1}^{R} + A_{SRO2}\frac{\alpha_{k}^{SR}}{k+1} + B_{SRO2}\omega_{k+1}^{SR}. \quad (47)$$

Figures 13 to 15 show the results for observable variables $V3$, $S2$, and $S6$.

Gross-Error Detection

We classify gross errors in three categories: true outliers, leaks, and biased instrumentation.

Outliers have been defined as values that depart from the expected distribution of values. Since the distribution of values is not known in the first place, assuming it to be normal has been the easy escape route. If outliers are to be identified without assuming any particular data distribution, it should be based on some criteria that takes into account what the distribution of the majority of data is. Typically, these gross errors correspond to sudden surges in the power supply to the instrumentation, or other isolated events.

Leaks can be classified in two kinds: predictable and unpredictable. Tank evaporation can be predicted if certain conditions, typically temperature, are known. Such leaks can be incorporated into the dynamic model as streams leaving the corresponding unit and treated as redundant variables if temperature is measured. Flare discharges are predictable, but can hardly be modeled. So they have to be modeled as observable known quantities. Other leaks, like tank leakage to the ground, or flange leaks cannot be predicted and must be detected. We leave the treatment of all these leaks to future work.

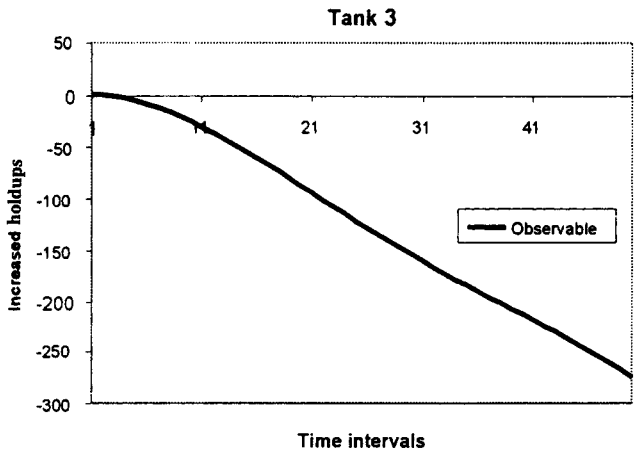

Figure 13. Example 1: calculated changes of holdups for Tank 3.

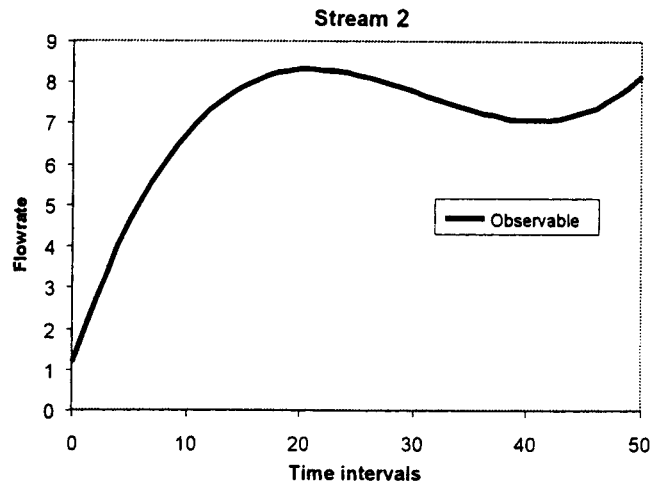

Figure 14. Example 1: calculated flow rates of Stream 2.

Biased instrumentation is typically a constant value added to the measurement signal due to miscalibration or a constant drift in the signal. Before we attempt to modify our integral approach to detect these biases, let us show the effect they have on reconciled data.

Example 2

Consider the data for Example 1. A constant bias +1 was added to all measurements of stream 4. The results of reconciling in this case are shown in Figures 16 to 21. Since it is common to all data-reconciliation methods based on maximum-likelihood procedures, the bias has spread to other variables, making its identification difficult. To create a list of suspect measurements we propose a measurement test.

Dynamic integral measurement test

Considering the measurement adjustments for instrument k at time interval j:

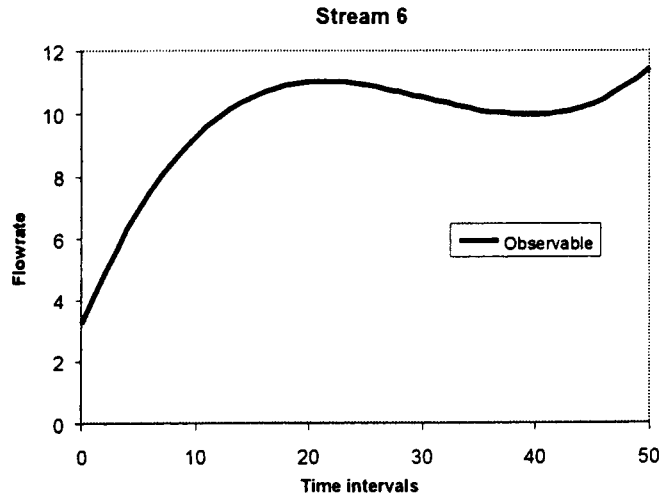

Figure 15. Example 1: calculated flow rates of Stream 6.

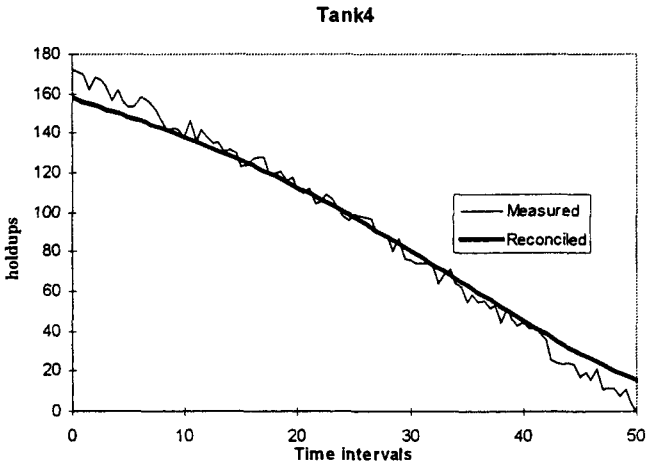

Figure 16. Example 2: Reconciled holdup for Tank 4.

$$Z_{jk} = x_{jk} - \hat{x}_{jk}, \tag{48}$$

where x_{jk} is the measurement for instrument k at time interval j, $\hat{x}_{jk}$ is the corresponding reconciled value. Assume that the errors in measurement are random. Therefore under the null hypothesis H_0 (no gross error is present), Z_{jk} should follow a normal distribution with zero mean:

$$E(Z_{jk}) = 0. \tag{49}$$

The standard deviation for Z_{jk} is hard to obtain, but the sample standard deviation is given by

$$S = \sqrt{\frac{1}{n}\sum_{j=0}^{n}\left(Z_{jk} - \overline{Z}_k\right)^2}, \tag{50}$$

where $\overline{Z}_k$ is the mean value of all adjustments. Thus, the following variable:

$$y_k = \frac{\overline{Z}_k}{S/\sqrt{n+1}} \tag{51}$$

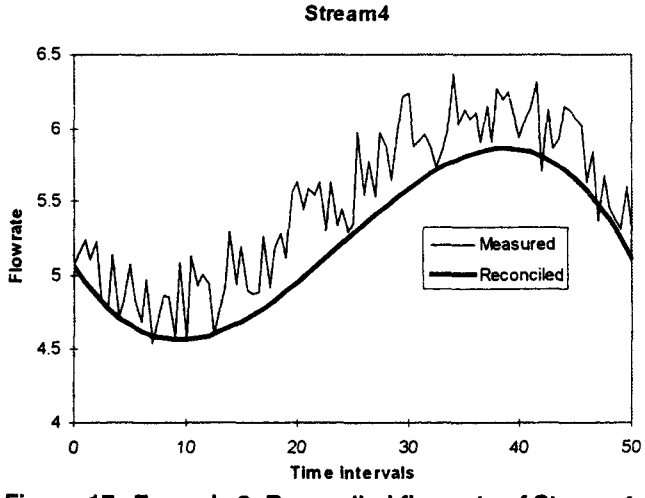

Figure 17. Example 2: Reconciled flow rate of Stream 4.

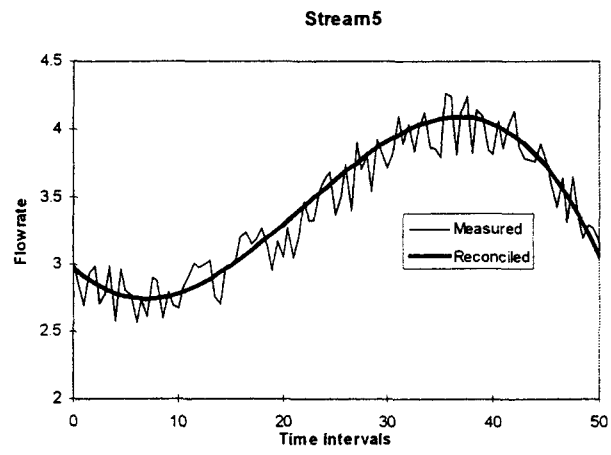

Figure 18. Example 2: Reconciled flow rate of Stream 5.

follows a t-student distribution:

$$y_k \sim t(n). \tag{52}$$

Thus, under the test, variable k will be suspected of containing a gross error if the following holds

$$\frac{|\overline{Z}_k|}{S/\sqrt{n+1}} > t_{1-(\eta/2)}, \tag{53}$$

where η is usually selected as 0.05 (95% confidence level).

Table 1 shows the measurement test applied to all redundant measurements in Example 2 and pinpoints which are the candidates for bias. The critical value at $\eta = 0.05$ is $t_{1-(\eta/2)} = 2.01$.

Effect of bias

Before any attempt is made to develop a model for bias detection a few numerical experiments will be presented. To see the effect of a bias in the value of the objective function the suspected streams measured data for Example 2 were

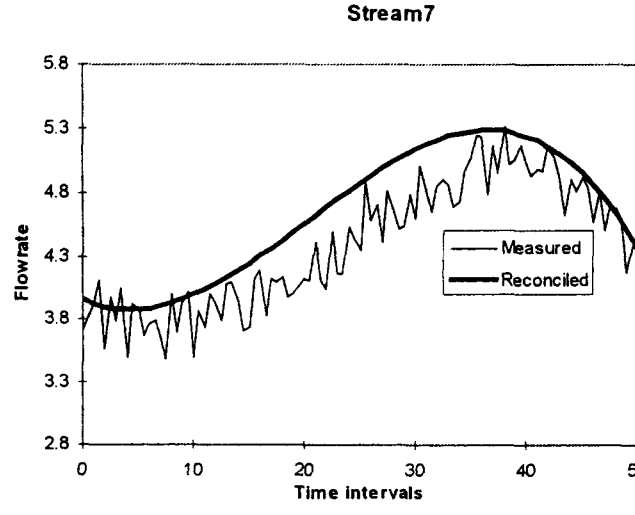

Figure 19. Example 2: Reconciled flow rate of Stream 7.

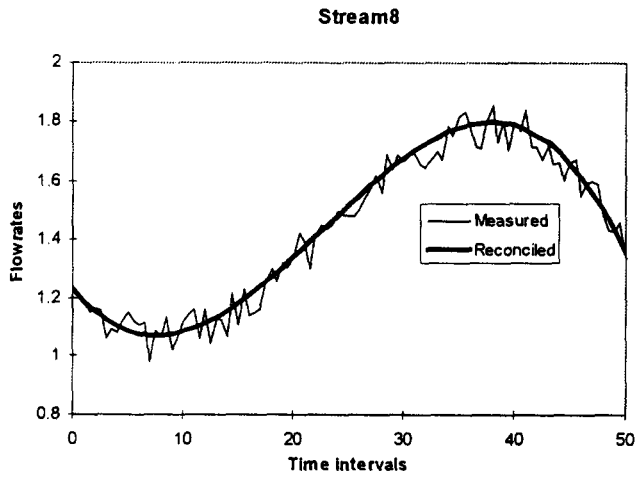

Figure 20. Example 2: Reconciled flow rate of Stream 8.

corrected by different values and the reconciliation was run. This addition, if it has the proper value and if it is performed on the right stream, is equivalent to suppressing the bias. Figure 22 shows the effect of these additions in the data to the value of the objective function. From this figure it is apparent that stream 4, as well as stream 7, can have opposite biases, that is, the level of redundancy of this network will not be enough to distinguish one from the other. This result can be generalized in the following statement: *For every unit in a flow sheet that is connected to the environmental node with more than one stream, detecting biased instrumentation on streams connecting the environmental node is impossible.* This result is important to notice, as it demonstrates the need for higher redundancy in the systems if bias instrumentation is to be detected; it also predicts the existence of multiple solutions if naive approaches are used for bias detection.

Bias-detection model

In this section, a model is proposed to detect one biased instrument. Consider then that for a particular biased stream candidate k the following holds:

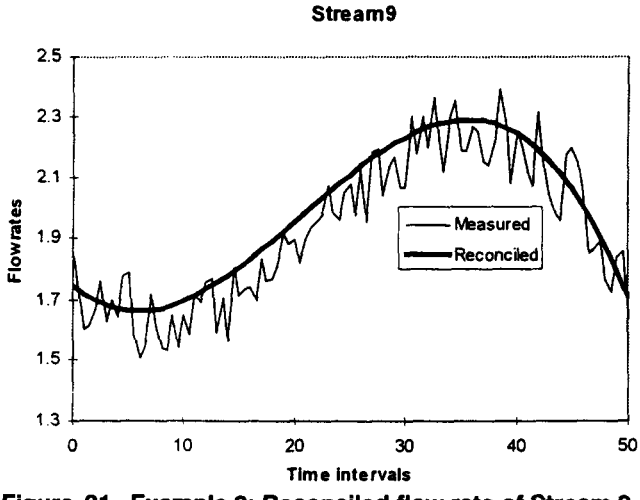

Figure 21. Example 2: Reconciled flow rate of Stream 9.

Table 1. Measurement Test Results

Variables	$\frac{\lvert\bar{Z}_k\rvert}{S/\sqrt{n+1}}$	Variables in Suspect
$V4$	0.0000	
$S4$	8.9178	×
$S5$	1.9305	
$S7$	9.7482	×
$S8$	1.9733	
$S9$	4.1877	×

$$f_{R,k} = \sum_{k=0}^{s} \alpha_k t^k + \delta \tag{54}$$

$$\text{Min}\left\{(J\upsilon_{R0} + T_\omega \omega^R - \upsilon_R^+)R_V^{-1}(J\upsilon_{R0} + T_\omega \omega^R - \upsilon_R^+) + (T_\alpha \alpha^R + E_k\delta - f_R^+)^T R_F^{-1}(T_\alpha \alpha^R + E_k\delta - f_R^+)\right\} \tag{55}$$

s.t.

$$D_m \omega^R = R_m \alpha^R$$

$$C_\alpha \alpha^R = 0,$$

where

$$E_k = \begin{bmatrix} e_k \\ e_k \\ \vdots \\ e_k \end{bmatrix},$$

and e_k is a vector with unity in position k and zero elsewhere.

If we apply the necessary conditions of optima to this problem, we obtain

$$z = [I - PM^T(MPM^T)^{-1}M]\left(-\frac{1}{2}P(w+u)\right) \tag{56}$$

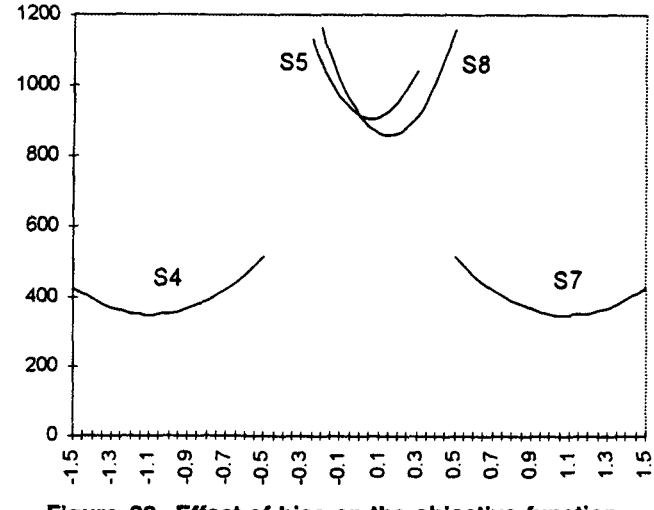

Figure 22. Effect of bias on the objective function.

$$\delta = -\frac{1}{2r_k}[p_k^T z - q_k], \tag{57}$$

where

$$p_k^T = (0 \quad 0 \quad 2E_k^T R_F^{-1} T_a) \qquad q_k = -2(f_R^+)^T R_F^{-1} E_k$$

$$r_k = E_k^T R_F^{-1} E_k \qquad u = -\frac{p_k q_k}{2r_k} \qquad P = Q + \frac{Q p_k p_k^T Q}{4r_k - p_k^T Q p_k}.$$

Unfortunately, the determinant of P is zero, which means that P is singular and a different approach needs to be used. Rewrite (Eq. 55) as follows:

$$\text{Min}\{x^T Q^{-1} x + w^T x + \delta p_k^T x + q_k \delta + r_k \delta^2\} \tag{58}$$

s.t.

$$Mx = 0$$

where

$$x = \begin{pmatrix} v_{R0} \\ \omega^R \\ \alpha^R \end{pmatrix}.$$

The optimization problem can be rewritten in the following way:

$$\min_{\delta} \left[\left\{ \begin{matrix} \min\limits_{\forall x} x^T Q^{-1} x + w^T x + \delta p_k^T x \\ \text{s.t.} \\ Mx = 0 \end{matrix} \right\} + q_k \delta + r_k \delta^2 \right]. \tag{59}$$

Let

$$F(\delta) = \left\{ \begin{matrix} \min\limits_{\forall x} x^T Q^{-1} x + w^T x + \delta p_k^T x \\ \text{s.t.} \\ Mx = 0 \end{matrix} \right\} + q_k \delta + r_k \delta^2. \tag{60}$$

Thus, the original problem becomes

$$\min_{\delta} F(\delta). \tag{61}$$

We first solve the following problem:

$$\min_{\forall x} \{x^T Q^{-1} x + (w^T + \delta p_k^T) x\}$$

s.t.

$$Mx = 0. \tag{62}$$

The solution is given by

$$x^*(\delta) = \left[I - QM^T (MQM^T)^{-1} M \right] \times \left(-\frac{1}{2} Q(w + p_k \delta) \right) = \tilde{x} + h_k \delta, \tag{63}$$

where $\tilde{x}$ is the solution for the reconciliation problem assuming no biases (given by Eq. 29). In turn, h is given by

$$h_k = -\frac{1}{2} Q p_k + \frac{1}{2} Q M^T (MQM^T)^{-1} MQ p_k \tag{64}$$

Substitute in Eq. 60 to obtain

$$F(\delta) = \tilde{x}^T Q^{-1} \tilde{x} + w^T \tilde{x} + (2\tilde{x}^T Q^{-1} h_k + w^T h + p_k^T \tilde{x} + q_k)\delta + (h_k^T Q^{-1} h_k + p_k^T h_k + r_k)\delta^2. \tag{65}$$

Thus, the solution of Eq. 61 is

$$\delta^* = -\frac{(2\tilde{x}^T Q^{-1} h_k + w^T h_k + p_k^T \tilde{x} + q_k)}{2(h_k^T Q^{-1} h_k + p_k^T h_k + r_k)}. \tag{66}$$

Bias detection is then performed as follows:

(a) The reconciliation is performed using the bias-detection model for each of the suspected candidates.
(b) The solution with the lowest value of the objective function is chosen.

Example 2 (continued)

Table 2 shows the biases calculated and the corresponding values of the objective function for each of the candidates detected by the dynamic integral measurement test. For comparison, the biases calculated when one selects other streams not flagged by the measurement test are shown. As is observed, the value of the objective function is much larger in these last cases.

Obviously, $S4$ and $S7$ are equally identified as having a bias. The value obtained is very close to the introduced bias. As was explained before, without a higher redundancy no further progress can be made. Methods to detect multiple biases will be addressed in future work.

Future Work

Several issues have been left for future work: leak detection, outlier identification, and multiple biased instrumentation. Other extensions, such as the inclusion of bounds as well as extension to bilinear systems will be investigated. Finally, as this article does not cover asymptotic behavior or periodic signals, a special technique to partition the measurement window automatically to allow for local polynomial representations will be presented.

Table 2. Bias Sizes

Stream No.	δ	Value of Objective Function	Streams Flagged
4	1.0943	349.6305	None
5	−0.0604	908.5810	*S4, S7, S8, S9*
7	−1.0943	349.6305	None
8	−0.1519	858.0663	*S4, S5, S7*
9	−0.1519	858.0663	*S4, S5, S7*

Conclusions

This article has presented an integral approach to dynamic data reconciliation of linear systems. It has been proven that through this approach a large volume of data can be handled effectively through an analytical solution that is computationally cheap. Gross errors have been discussed and an effective method to find one biased flow rate has been proposed.

Acknowledgments

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

Notation

A = system matrix, Eq. 1
B = system matrix, Eq. 8
C = system matrix, Eq. 2
f = vector of flow rates
I = identity matrix
S_F = variance matrix of flow rates
S_V = variance matrix of holdups
v = vector of volumes
α = coefficients in flow-rate polynomial
ω = coefficients in holdup polynomial
ξ = variable
β = scaling factor
δ = bias

Subscripts and superscripts

F = flow-rate-related quantities
NR = nonredundant quantities
O = observable quantities
R = redundant quantities
SR = self-redundant (temporally redundant) quantities
UO = unobservable quantities
V = holdup-related quantities
$+$ = measured quantities

Literature Cited

Almasy, G. A., and R. S. H. Mah, "Estimation of Measurement Error Variances from Process Data," *Ind. Eng. Chem. Process Des. Dev.*, **23**, 779 (1984).
Albuquerque, J. S., and L. T. Biegler, "Decomposition Algorithms for On-line Estimation with Nonlinear Models," *Comput. Chem. Eng.*, **19**, 1031 (1995).
Albuquerque, J. S., and L. T. Biegler, "Data Reconciliation and Gross-error Detection for Dynamic Systems," *AIChE J.*, **42**, 2841 (1996).
Bagajewicz, M., "On the Probability Distribution and Reconciliation of Process Plant Data," *Comput. Chem. Eng.*, **20**, 813 (1996).
Bagajewicz, M., and S. Mullick, "Reconciliation of Plant Data. Application and Future Trends," AIChE Meeting, Houston (1995).
Darouach, M., and M. Zasadzinski, "Data Reconciliation in Generalized Linear Dynamic Systems," *AIChE J.*, **37**, 193 (1991).
Kalman, R. E., "New Approach to Linear Filtering and Prediction Problems," *J. Basic Eng., ASME*, **82D**, 35 (1960).
Kao, C. S., A. C. Tamhane, and R. S. H. Mah, "Gross Error Detection in Serially Correlated Process Data. 2. Dynamic Systems," *Ind. Eng. Chem. Res.*, **31**, 254 (1992).
Karjala, T. W., and D. M. Himmelblau, "Dynamic Rectification of Data via Recurrent Neural Nets and the Extended Kalman Filter," *AIChE J.*, **42**, 2225 (1996).
Keller, J. Y., M. Zasadzinski, and M. Darouach, "Analytical Estimator of Measurement Error Variances in Data Reconciliation," *Comput. Chem. Eng.*, **16**, 185 (1992).
Liebman, M. J., T. F. Edgar, and L. S. Lasdon, "Efficient Data Reconciliation and Estimation for Dynamic Process Using Nonlinear Programming Techniques," *CES*, **16**, 963 (1992).
Madron, F., "Process Plant Performance. Measurement and Data Processing for Optimization and Retrofits," *Ellis Horwood Series in Chemical Engineering*, Chichester, England (1992).
Narasimhan, S., and R. S. H. Mah, "Generalized Likelihood Ratios for Gross Error Identification in Dynamic Processes," *AIChE J.*, **34**, 1321 (1988).
Ramamurthi, Y., P. B. Sistu, and B. W. Bequette, "Control-relevant Dynamic Data Reconciliation and Parameter Estimation," *Comput. Chem. Eng.*, **17**, 41 (1993).
Rollins, D. K., and S. Devanathan, "Unbiased Estimation in Dynamic Data Reconciliation," *AIChE J.*, **39**, 1330 (1993).
Tong, H., and C. M. Crowe, "Detecting Persistent Gross Errors by Sequential Analysis of Principal Components," *Comput. Chem. Eng.*, **20**, S733 (1996).
Stanley, G. M., and R. S. H. Mah, "Estimation of Flows and Temperatures in Process Networks," *AIChE J.*, **23**, 642 (1977).

Appendix

In this appendix it is shown how the inverse of $(T_\alpha^T R_F^{-1} T_\alpha)$ can be performed analytically. Although this matrix is sparse, its inverse is not, as will be shown below. Additionally, its elements have an order of magnitude of difference. Thus, it is advantageous to perform its inversion analytically, instead of numerically.

First note that

$$(T_\alpha^T R_F^{-1} T_\alpha) = \begin{bmatrix} I & I & \cdots & I \\ 0 & I & \cdots & nI \\ \vdots & \vdots & \ddots & \vdots \\ 0 & I & \cdots & n^s I \end{bmatrix} \begin{bmatrix} S_F^{-1} & 0 & \cdots & 0 \\ 0 & S_F^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_F^{-1} \end{bmatrix} \begin{bmatrix} I & 0 & \cdots & 0 \\ I & I & \cdots & I \\ \vdots & \vdots & \ddots & \vdots \\ I & nI & \cdots & n^s I \end{bmatrix}$$

$$= \begin{bmatrix} I & I & \cdots & I \\ 0 & I & \cdots & nI \\ \vdots & \vdots & \ddots & \vdots \\ 0 & I & \cdots & n^s I \end{bmatrix} \begin{bmatrix} S_F^{-1} & 0 & \cdots & 0 \\ S_F^{-1} & S_F^{-1} & \cdots & S_F^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ S_F^{-1} & nS_F^{-1} & \cdots & n^s S_F^{-1} \end{bmatrix} = \begin{bmatrix} (n+1)S_F^{-1} & S_F^{-1}\sum_{i=1}^{n} j & \cdots & S_F^{-1}\sum_{i=1}^{n} j^s \\ S_F^{-1}\sum_{i=1}^{n} j & S_F^{-1}\sum_{i=1}^{n} j^2 & \cdots & S_F^{-1}\sum_{i=1}^{n} j^{s+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_F^{-1}\sum_{i=1}^{n} j^s & S_F^{-1}\sum_{i=1}^{n} j^{s+1} & \cdots & S_F^{-1}\sum_{i=1}^{n} j^{2s} \end{bmatrix}$$

$$= \begin{bmatrix} S_F^{-1} & 0 & \cdots & 0 \\ 0 & S_F^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_F^{-1} \end{bmatrix} \begin{bmatrix} (n+1)I & I\sum_{i=1}^{n} j & \cdots & I\sum_{i=1}^{n} j^s \\ I\sum_{i=1}^{n} j & I\sum_{i=1}^{n} j^2 & \cdots & I\sum_{i=1}^{n} j^{s+1} \\ \vdots & \vdots & \ddots & \vdots \\ I\sum_{i=1}^{n} j^s & I\sum_{i=1}^{n} j^{s+1} & \cdots & I\sum_{i=1}^{n} j^{2s} \end{bmatrix}.$$

Then

$$(T_\alpha^T R_F^{-1} T_\alpha)^{-1} = V\Gamma_F,$$

where

$$\Gamma_F = \begin{bmatrix} S_F & 0 & \cdots & 0 \\ 0 & S_F & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_F^1 \end{bmatrix}$$

$$V = \begin{bmatrix} (n+1)I & I\sum_{i=1}^{n} j & \cdots & I\sum_{i=1}^{n} j^s \\ I\sum_{i=1}^{n} j & I\sum_{i=1}^{n} j^2 & \cdots & I\sum_{i=1}^{n} j^{s+1} \\ \vdots & \vdots & \ddots & \vdots \\ I\sum_{i=1}^{n} j^s & I\sum_{i=1}^{n} j^{s+1} & \cdots & I\sum_{i=1}^{n} j^{2s} \end{bmatrix}^{-1}$$

Matrix V^{-1} is sparse, consisting of several diagonal submatrices, each with all elements in the diagonal being the same. Additionally for n large the elements in the matrix can become large. For example, for $s = 3$ and $n = 100$ we have

$$V = \frac{1}{101} \begin{bmatrix} I & 50I & 3350I & 2.53\,10^5 I \\ 50I & 3350I & 2.53\,10^5 I & 2.03\,10^7 I \\ 3350I & 2.53\,10^5 I & 2.03\,10^7 I & 1.70\,10^9 I \\ 2.53\,10^5 I & 2.03\,10^7 I & 1.70\,10^9 I & 1.46\,10^{11} I \end{bmatrix}^{-1}.$$

We propose to use the Householder formula for matrix inversion recursively. The Householder formula is as follows: If

$$A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix},$$

then

$$A^{-1} = \begin{bmatrix} E & -EF \\ -GE & S^{-1} + GEF \end{bmatrix},$$

where

$$F = QS^{-1} \qquad G = S^{-1}R \qquad E = (P - QG)^{-1}$$

$$P = (n+1)I \qquad S = \begin{bmatrix} I\sum_{i=1}^{n} j^2 & \cdots & I\sum_{i=1}^{n} j^{s+1} \\ \vdots & \ddots & \vdots \\ I\sum_{i=1}^{n} j^{s+1} & & I\sum_{i=1}^{n} j^{2s} \end{bmatrix}$$

$$R = \begin{bmatrix} I\sum_{i=1}^{n} j \\ \vdots \\ I\sum_{i=1}^{n} j^s \end{bmatrix} \qquad Q = R^T.$$

In turn, S can be inverted using the same procedure. Now, this procedure can be applied recursively.

Manuscript received Jan. 6, 1997, and revision received May 23, 1997.