# Gross error modeling and detection in plant linear dynamic reconciliation*

Miguel J. Bagajewicz† and Qiyou Jiang

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

## Abstract

This paper presents a method to identify and estimate gross errors in plant linear dynamic data reconciliation. An *integral dynamic data reconciliation method* presented in a previous paper (Bagajewicz and Jiang, 1997) is extended to allow multiple gross error estimation. The *dynamic integral measurement test* is extended to identify hold-up measurements as suspects of gross error. A series of theorems are used to show the equivalencies of gross errors and to discuss the issue of exact identification. A serial approach for gross error identification and estimation is then presented. Gross errors are identified without the need for measurement elimination. The strategy is capable of effectively identifying a large number of gross errors. © 1998 Published by Elsevier Science Ltd. All rights reserved.

*Keywords*: gross error modeling; gross error detection; dynamic data reconciliation

## Introduction

Reliable process data are the basis for efficient process operation and control. All instrument readings are subject to errors. Random errors are inherent to the measurement procedure and are also called noise. Nonrandom events are also present. They are caused by instrument biases, process leaks and inaccurate process models and received the name of gross errors. Random errors can be eliminated by using data reconciliation technique that adjusts measurements to be consistent to the corresponding conversation laws. However, the presence of gross error will invalidate the statistical basis of data reconciliation, and therefore gross errors must be detected and either eliminated or corrected before valid data reconciliation can be performed. When gross errors are corrected, the process is called compensation.

Gross error detection and size estimation has been for many years an elusive issue. Several test statistics and methods have been proposed for the identification and size estimation of biases and leaks. Methods based on steady-state linear data reconciliation are known as the *global test* (Reilly and Carpani, 1963), the *measurement test* (Mah and Tamhane, 1982;

Crowe *et al.*, 1983), the *nodal test* (Reilly and Carpani, 1963; Mah *et al.*, 1976), *generalized likelihood ratios* (GLR) (Narasimhan and Mah, 1987), *unbiased estimation* of *gross error* (UBET) (Rollins and Davis, 1992), *principal component tests* (PCT) (Tong and Crowe, 1994), etc. Gross error size estimation methods were proposed by Madron (1985) and Narasimhan and Mah (1987). Multiple gross errors are detected and eventually estimated through graph-theoretic analysis (Mah *et al.*, 1976) or serial elimination procedures or serial compensation procedures (Serth and Heenan, 1986; DATACON, 1993). Serial compensation, which consists of fixing one bias at a time, is evolving into serial identification with collective estimation (Rollins and Davis, 1992; Keller *et al.*, 1994; Kim *et al.*, 1997; Sanchez, 1996). This technique consists in successively identifying one gross error at a time and solving a model that will include the compensation for all the identified gross errors.

The performance of all these steady-state based methods has been reported to be dependent on the flow sheet and the size of the gross error. For example, the power of the measurement test varies from poor values around 10% to encouraging values of 99% (Iordache *et al.*, 1985). To ameliorate this limited power, attention has shifted recently to sequential analysis (Tong and Crowe, 1996, 1997), where data from different days is sequentially analyzed with different statistical techniques.

Detection of gross errors using dynamic methods has been also proposed (Narasimhan and Mah, 1988; Rollins and Devanathan, 1993; Albuquerque and Biegler, 1996). In a recent paper, a method for dynamic data reconciliation based on an integral form of the system equations was presented (Bagajewicz and Jiang, 1997). In this paper, the system equations are first rewritten in a canonic form to separate measured from unmeasured variables. Then a polynomial representation of flowrates and tank/unit holdups is proposed and reconciliation is performed. Finally, the *dynamic integral measurement test* was introduced and a model for dynamic data reconciliation in the presence of one gross error was introduced.

This paper is devoted to the issue of gross error detection in the context of the *dynamic integral data reconciliation* method and is organized as follows: A model of data reconciliation in the presence of multiple gross errors, including flow and holdup bias as well as leaks is presented first. Next, gross error equivalencies that affect gross error detection are discussed and a method to avoid introducing equivalent gross errors in the model is presented. The issue of exact gross error detection and size estimation is presented. The method is based on serial identification with collective compensation at each identification stage.

## A model for dynamic data reconciliation and estimation of multiple gross errors

Let $f$ be the flowrates and $v$ the tank holdups. The dynamic material balance around tanks and other units without holdups can be respectively represented by equation (1) and (2) (Bagajewicz and Jiang, 1997):

$$\mathrm{d}v/\mathrm{d}t = Af, \tag{1}$$

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

or

$$D\begin{pmatrix} f \\ \mathrm{d}v/\mathrm{d}t \end{pmatrix} = 0, \tag{3}$$

where

$$D = \begin{bmatrix} A & -I \\ C & 0 \end{bmatrix}. \tag{4}$$

The initial condition for equation (1) is $v|_{t=0} = v_0$.

Let $f_i^*$, $i = 0, \ldots, n$, be the vector of measurements of all stream flows and $v_i^*$ the vector of tank holdup measurements at a particular time $t_i$. Since the system (equations (1) and (2)) may contain unmeasured variables, a Gauss–Jordan procedure, which is called co-optation (Madron, 1992), is used to obtain the redundant system:

$$B_R \frac{\mathrm{d}v_R}{\mathrm{d}t} = A_R f_R, \tag{5}$$

$$C_R f_R = 0. \tag{6}$$

where only redundant measured streams are included. A measured stream is redundant if its elimination makes it observable (Madron, 1992).

The following polynomial representation was proposed for the flowrates and holdups as a function of time (Bagajewicz and Jiang, 1997):

$$f_R \approx \sum_{k=0}^{s} \alpha_k^R t^k, \tag{7}$$

$$v_R \approx v_{R0} + \sum_{k=0}^{s} \omega_{k+1}^R t^{k+1}, \tag{8}$$

where $s$ is the polynomial order. Assuming that there are $n + 1$ measurements of each stream at equally spaced time intervals with normally distributed errors, the corresponding maximum likelihood problem that is associated with data reconciliation is:

$$\mathrm{Min} \quad \sum_{i=0}^{n} \{ (v_{Ri} - v_{Ri}^+)^T S_v^{-1} (v_{Ri} - v_{Ri}^+)$$

$$+ (f_{Ri} - f_{Ri}^+)^T S_F^{-1} (f_{Ri} - f_{Ri}^+) \} \tag{9}$$

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

$$C_R f_{Ri} = 0 \quad (i = 0, \ldots, n).$$

or, in terms of the polynomial coefficients,

$$\mathrm{Min} \ \{ Jv_{R0} + T_\omega \omega^R - v_R^+)^T R_v^{-1} \ (Jv_{R0} + T_\omega \omega^R - v_R^+)$$

$$+ (T_\alpha \alpha^R - f_R^+)^T R_F^{-1} (T_\alpha \alpha^R - f_R^+) \} \tag{10}$$

s.t. $\quad D_m \omega^R = R_m \alpha^R,$

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

where $S_F$ and $S_V$ are variance matrix of flowrates and holdups and the rest of the vectors and matrices are given by

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

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

$$D_m = \begin{bmatrix} B_R & 0 & \cdots & 0 \\ 0 & B_R & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_R \end{bmatrix},$$

$$R_m = \begin{bmatrix} A_R & 0 & \cdots & 0 \\ 0 & \dfrac{A_R}{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \dfrac{A_R}{s+1} \end{bmatrix},$$

$$C_\alpha = \begin{bmatrix} C_R & 0 & \cdots & 0 \\ 0 & C_R & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_R \end{bmatrix},$$

$$R_V^{-1} = \begin{bmatrix} S_V^{-1} & 0 & \cdots & 0 \\ 0 & S_V^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_V^{-1} \end{bmatrix},$$

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

$$T_\alpha = \begin{bmatrix} I & 0 & \cdots & 0 \\ I & t_1 I & \cdots & t_1^s I \\ \vdots & \vdots & \ddots & \vdots \\ I & t_n I & \cdots & t_n^s I \end{bmatrix},$$

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

This model has been used as a basis to investigate the effect of a single flowrate bias (Bagajewicz and Jiang, 1997). In this paper, we extend this model to include the prediction of multiple biased flowrate measurements, biased holdup measurements and leaks.

Three kinds of gross errors are considered in this paper. Flowrate biases, holdup biases and tank leaks. Biases are typically classified as zero-order (constant) and nonzero order (Dunia *et al.*, 1996). Figure 1 illustrates a zero-order and a first-order bias. In turn, leaks can be constant (zero order) or in the case when they depend on height of liquid in a tank, they can be of any order.

Gross errors are typically considered as zero-order errors and lumped into steady-state models in practice. In this paper, a model with arbitrary order biases and leaks will be derived first. However, only zero-order flowrate biases and leaks are addressed in the examples.

*Flowrate biases*

Assume a set of gross errors is present. This is modeled as follows.

$$f_R = \sum_{k=0}^{s} \alpha_k^R t^k + \delta_F \tag{11}$$

where $\delta_F$ is a vector that contains biases for bias candidates. We are interested in deleting the zero elements of $\delta_F$. Thus we write $\delta_F$ as follows:
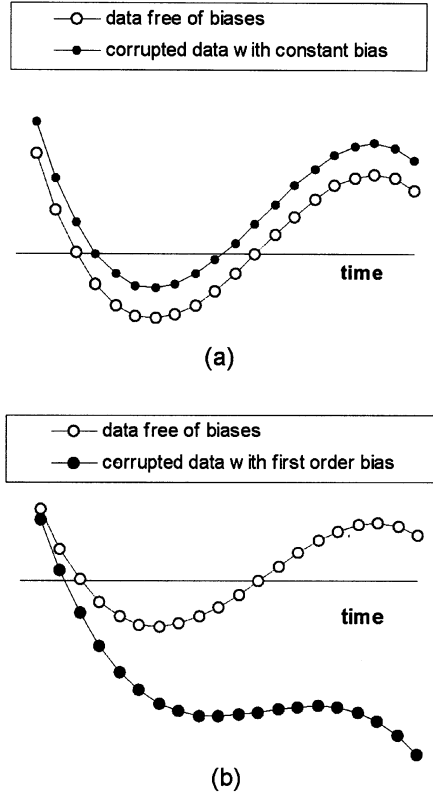
$$\delta_F = E_F \hat{\delta}_F \tag{12}$$



Fig. 1. Examples of different order biases.

where

$$E_F = \begin{bmatrix} \psi & & & \\ & \psi & & \\ & & \ddots & \\ & & & \psi \end{bmatrix}, \qquad \psi = \begin{bmatrix} I \\ 0 \end{bmatrix} \tag{13}$$

and

$$\hat{\delta}_F = \begin{bmatrix} \hat{\delta}_{F0} \\ \hat{\delta}_{F1} \\ \vdots \\ \hat{\delta}_{Fn} \end{bmatrix}, \qquad \hat{\delta}_{Fi} = \sum_{k=0}^{SF} \rho_k t^k. \tag{14}$$

Thus to simultaneously reconcile the system and estimate the size of the biases we solve:

$$\text{Min} \quad \{ (Jv_{R0} + T_\alpha \omega^R - v_R^+)^T R_V^{-1} (Jv_{R0} + T_\alpha \omega^R - v_R^+)$$
$$+ (T_\alpha \alpha^R + E_F T_\rho \rho - f_R^+)^T$$
$$\times R_F^{-1} (T_\alpha \alpha^R + E_F T_\rho \rho - f_R^+) \} \tag{15}$$

$$\text{s.t.} \quad D_m \omega^R = R_m \alpha^R,$$
$$C_\alpha \alpha^R = 0.$$

Let

$$P_F^T = (0 \quad 0 \quad 2 T_\rho^T E_F^T R_F^{-1} T_\alpha),$$
$$q_F^T = -2 (f_R^+)^T R_F^{-1} E_F T_\rho,$$
$$U_F = T_\rho^T E_F^T R_F^{-1} E_F T_\rho. \tag{16}$$

Then

$$\text{Min} \quad \{x^T Q^{-1} x + w^T x + \rho^T P_F^T x + q_F^T \rho + \rho^T U_F \rho\} \tag{17}$$

s.t. $\quad Mx = 0$

where

$$x = \begin{Bmatrix} v_{R0} \\ \omega^R \\ \alpha^R \end{Bmatrix}, \quad M = \begin{pmatrix} 0 & D_m & -R_m \\ 0 & 0 & C_\alpha \end{pmatrix}. \tag{18}$$

The optimization problem can be written in the following way:

$$\underset{\rho}{\text{Min}} \left[ \left\{ \begin{matrix} \text{Min } x^T Q^{-1} x + w^T x + \rho^T P_F^T x \\ \forall x \\ \text{s.t.} \quad Mx = 0 \end{matrix} \right\} + q_F^T \rho + \rho^T U_F \rho \right]. \tag{19}$$

let

$$F(\rho) = \left\{ \begin{matrix} \text{Min } x^T Q^{-1} x + w^T x + \rho^T P_F^T x \\ \forall x \\ \text{s.t.} \quad Mx = 0 \end{matrix} \right\} + q_F^T \rho + \rho^T U_F \rho. \tag{20}$$

Thus, the original problem becomes:

$$\underset{\rho}{\text{Min}} \; F(\rho). \tag{21}$$

Thus, we first solve the following problem:

$$\underset{\forall x}{\text{Min}} \quad \{x^T Q^{-1} x + (w^T + \rho^T P_F^T) x\} \tag{22}$$

s.t. $\quad Mx = 0$.

The solution is given by

$$\dot{x}(\rho) = [I - QM^T(MQM^T)^{-1} M]\left( -\frac{1}{2} Q(w + P_F \rho) \right)$$

$$= \tilde{x} + H_F \rho \tag{23}$$

where $\tilde{x}$ is the solution for the reconciliation problem assuming no biases and

$$H_F = \tfrac{1}{2}[ -1 + QM^T(MQM^T)^{-1}M]QP_F. \tag{24}$$

Thus,

$$F(\rho) = (\tilde{x} + H_F \rho)^T Q^{-1} (\tilde{x} + H_F \rho)$$

$$+ w^T (\tilde{x} + H_F \rho) + \rho^T P_F^T (\tilde{x} + H_F \rho)$$

$$+ q_F^T \rho + \rho^T U_F \rho, \tag{25}$$

$$\frac{\partial F(\rho)}{\partial p} = (2H_F^T Q^{-1} \tilde{x} + H_F^T w + P_F^T \tilde{x} + q_F)$$

$$+ 2(H_F^T Q^{-1} H_F + P_F^T H_F + U_F)\rho = 0 \tag{26}$$

and the solution is

$$\rho = -\tfrac{1}{2}(H_F^T Q^{-1} H_F + P_F^T H_F + U_F)^{-1}$$

$$\times (2H_F^T Q^{-1} \tilde{x} + H_F^T w + P_F^T \tilde{x} + q_F). \tag{27}$$

*Tank holdup biases*

Now, in addition to flowrate biases, assume also that

$$v_R = \sum_{k=0}^{s+1} \omega_k^R t^k + \delta_V. \tag{28}$$

Let

$$\delta_V = E_V \hat{\delta}_V \tag{29}$$

where

$$E_V = \begin{bmatrix} \psi & & & \\ & \psi & & \\ & & \ddots & \\ & & & \psi \end{bmatrix}, \quad \psi = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \hat{\delta}_V = \begin{bmatrix} \hat{\delta}_{V0} \\ \hat{\delta}_{V1} \\ \vdots \\ \hat{\delta}_{Vn} \end{bmatrix},$$

$$\hat{\delta}_{Vi} = \sum_{k=0}^{SV} \mu_k t^k. \tag{30}$$

The reconciliation and gross error estimation problem becomes

$$\text{min} \quad (J v_{R0} + T_\omega \omega^R + E_V T_\mu \mu - v_R^+)^T$$

$$\times R_V^{-1} (J v_{R0} + T_\omega \omega^R + E_V T_\mu \mu - v_R^+$$

$$+ (T_\alpha \alpha^R + E_F T_p \rho - f_R^+)^T$$

$$R_F^{-1} (T_\alpha \alpha^R + E_F T_p \rho - f_R^+) \tag{31}$$

s.t. $\quad D_m \omega^R = R_m \alpha^R,$

$\quad\quad C_\alpha \alpha^R = 0.$

Let

$$P^T = 2\begin{pmatrix} T_\mu^T E_V^T R_V^{-1} J & T_\mu^T E_V^T R_V^{-1} T_\omega & 0 \\ 0 & 0 & T_\rho^T E_F^T R_F^{-1} T_\alpha \end{pmatrix}, \tag{32}$$

$$q^T = -2((v_R^+)^T R_V^{-1} E_V T_\mu \;\; (f_R^+)^T R_F^{-1} E_F T_\rho), \tag{33}$$

$$U = \begin{pmatrix} T_\mu^T E_V^T R_V^{-1} E_V T_\mu & 0 \\ 0 & T_p^T E_F^T R_F^{-1} E_F T_\rho \end{pmatrix}, \quad \tau = \begin{pmatrix} \mu \\ \rho \end{pmatrix}. \tag{34}$$

Then

$$\text{min} \; \{x^T Q^{-1} x + w^T x + \tau^T P^T x + q^T \tau + \tau^T U \tau\} \tag{35}$$

s.t. $\quad Mx = 0$

where

$$x = \begin{Bmatrix} v_{R0} \\ \omega^R \\ \alpha^R \end{Bmatrix} \tag{36}$$

and the solution is similar to expressions (22) and (26).

*Leaks*

In the presence of leaks, the original model $Dx = 0$ is now $Dx = L$, where $L$ is a vector of leaks for each

tank. Thus after transforming the system equations into their canonical form

$$B_R \frac{dv_R}{dt} = A_R f_R - B_R E_l l, \qquad (37)$$

$$C_R f_R = 0, \qquad (38)$$

where

$$E_l = [e_1 \quad e_2 \quad \cdots \quad e_{sl}] \qquad (39)$$

and $l$ is a vector of proposed leaks and $e_i$ is a vector with unity in a position corresponding to a leak and zero elsewhere. Let

$$l_i = \sum_{k=0}^{sl} \lambda_k t^k. \qquad (40)$$

Then

$$B_R \sum_{k=0}^{s} \omega_{k+1}^R t^{k+1} = A_R \sum_{k=0}^{s} \frac{\alpha_k^R}{k+1} t^{k+1}$$
$$- B_R E_l \sum_{k=0}^{sl} \frac{\lambda_k}{k+1} t^{k+1}. \qquad (41)$$

And since this equation is valid for all $t$, then

$$B_R \omega_{k+1}^R = \frac{A_R \alpha_k^R}{k+1} - G_R E_1 \frac{\lambda_k}{k+1}, \quad k = 0, \ldots, s \qquad (42)$$

where

$$\begin{cases} G_R = B_R, & k = 0, \ldots, sl, \\ G_R = 0, & k = sl + 1, \ldots l. \end{cases} \qquad (43)$$

Thereofre, the reconciliation and gross error estimation model is:

$$\min \quad \{(Jv_{R0} + T_\omega \omega^R + \delta_V - v_R^+)^T R_V^{-1}$$
$$(Jv_{R0} + T_\omega \omega^R + \delta_V - v_R^+) + (T_\alpha \alpha^R + \delta_F - f_R^+)^T$$
$$\times R_F^{-1}(T_\alpha \alpha^R + \delta_F - f_R^+)\} \qquad (44)$$

s.t. $\quad D_m \omega^R = R_m \alpha^R - G_m \lambda,$

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

where

$$G_m = \begin{bmatrix} B_R E_l & & & \\ & \dfrac{B_R E_l}{2} & & \\ & & \ddots & \\ & & & \dfrac{B_R E_l}{sl+1} \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \qquad (45)$$

We rewrite the problem as follows:

$$\min \quad \{x^T Q^{-1} x + w^T x + \tau^T P^T x + q\tau + \tau^T U\tau\} \qquad (46)$$

s.t. $\quad Mx = K_m \lambda$

where

$$K_m = \begin{pmatrix} G_m \\ 0 \end{pmatrix}. \qquad (47)$$

In general, a solution to

$$\min_{\forall x} \quad \{x^T Q^{-1} x + (w^T + \tau^T P^T)x\} \qquad (48)$$

s.t. $\quad Mx = K_m \lambda$

is linear in $\tau$ and $\lambda$, i.e.

$$x^* = \tilde{x} + h_1 \tau + h_2 \lambda \qquad (49)$$

where

$$h_1 = -\tfrac{1}{2}QP + \tfrac{1}{2}QM^T(MQM^T)^{-1} MQP, \qquad (50)$$

$$h_2 = QM^T(MQM^T)^{-1}K_m. \qquad (51)$$

Let

$$F(\tau, \lambda) = (\tilde{x} + h_1 \tau + h_2 \lambda)^T Q^{-1}(\tilde{x} + h_1 \tau + h_2 \lambda)$$
$$+ (w^T + \tau^T P^T)(\tilde{x} + h_1 \tau + h_2 \lambda)$$
$$+ q\tau + \tau^T U\tau. \qquad (52)$$

Then solve the system

$$\partial F(\tau^*, \lambda^*)/\partial \lambda = 0, \qquad (53)$$

$$\partial F(\tau^*, \lambda^*)/\partial \tau = 0, \qquad (54)$$

and solve for $\tau$ and $\lambda$:

$$\tau^* = (C2 - B2*B1^{-1}*C1)^{-1}(B2*B1^{-1}*a1 - a2), \qquad (55)$$

$$\lambda^* = B1^{-1}(-a1 - C1*\tau^*), \qquad (56)$$

where

$$a1 = 2h_2^T Q^{-1}\tilde{x} + h_2^T w, \quad B1 = 2h_2^T Q^{-1}h_2, \qquad (57)$$

$$C1 = h_2^T Q^{-1}h_1 + h_2^T P$$

$$a2 = 2h_1^T Q^{-1}\tilde{x} + h_1^T w + P^T\tilde{x} + q,$$
$$B2 = 2h_1^T Q^{-1}h_2 + P^T h_2 \qquad (58)$$

$$C2 = 2(h_1^T Q^{-1}h_1 + P^T h_1 + U). \qquad (59)$$

Thus, the solution to data reconciliation with simultaneous bias and leak estimation is:

$$x^* = \tilde{x} + h_1 \tau^* + h_2 \lambda^*. \qquad (60)$$

### A dynamic integral measurement test for holdups

The *dynamic integral measurement test* (DIMT) was introduced in a previous paper (Bagajewicz and Jiang, 1997). The method is based on introducing the following statistics:

$$y_k = \bar{Z}_k/(S/\sqrt{n+1}) \qquad (61)$$

where $\bar{Z}_k$ is the average of all deviations of the measurements from the reconciled values and $S$ is the sample standard deviation. It is assumed that $y_k$ follows a $t$-student distribution:

$$y_k \sim t(n). \qquad (62)$$

Thus, under the test, variable $i$ will be declared being suspect of containing a gross error if the following holds:

$$\frac{|\bar{Z}_k|}{S/\sqrt{n+1}} > t_{1-\eta/2} \qquad (63)$$

where $\eta$ is usually selected as 0.05 (95% confidence level).

For holdups, however, we have that $\bar{Z}_k = 0$. This can be easily proved by realizing that for any tank holdup the initial holdup is contained in the objective function of the reconciliation problem as a variable, but is not present in the constraints. The statement is proved by contradiction, that is, assuming that the average is not zero and showing that there exist a different value of the initial holdup that makes the objective function smaller. Suppose by contradiction that $\bar{Z}_k = \Delta \neq 0$, i.e.

$$\bar{Z}_k = \Delta = \sum_{j=0}^{n} (v_{ij}^+ - v_{ij}) = \sum_{j=1}^{n} [v_{ij}^+ - v_{i0} - p_i(t_j)]$$
$$+ (v_{i0}^+ - v_{i0}) \neq 0 \qquad (64)$$

where we have used $v_{ij}$ to indicate the reconciled holdup of tank $i$ at time $t_j$ and $p_i(t_j)$ to indicate the nonconstant polynomial part of the holdup. Now, consider that the contribution of these holdups to the objective function:

$$\Pi = \sum_{j=0}^{n} (v_{ij}^+ - v_{ij})^2 = \sum_{j=1}^{n} [v_{ij}^+ - v_{i0} - p_i(t_j)]^2$$
$$+ (v_{i0}^+ - v_{i0})^2. \qquad (65)$$

As the initial holdup is only present in the objective function, consider using a new value

$$\hat{v}_{i0} = v_{i0} + \Delta/(n+1). \qquad (66)$$

This value will now make the average $\bar{\bar{Z}}_k = 0$. The contribution of these holdups to the objective function is now

$$\hat{\Pi} = \Pi - \Delta^2/(n+1) \qquad (67)$$

making, thus, the objective function lower, which proves the statement.

In spite of this fact, large deviations from the measurements are still possible. Consider the case of a single tank with constant input and constant output stream and a first-order holdup bias. Figure 2 shows the result of reconciliation.

As it is displayed, the measurements display a positive adjustment for half the interval and a negative adjustment for the other half, making the average of all adjustments zero, but exhibiting a large deviations. Thus the following DIMT for holdup biases is proposed.

*DIMT for Hold-up Bias (DIVMT)*: Apply the DIMT using a sample covering only the first half of time period in consideration for the reconciliation.

As proposed the DIVMT is able to detect large deviations, compensating thus for the shifting performed at reconciliation time.

*Example*

Consider the process network shown in Fig. 3, which was first introduced by Bagajewicz and Jiang
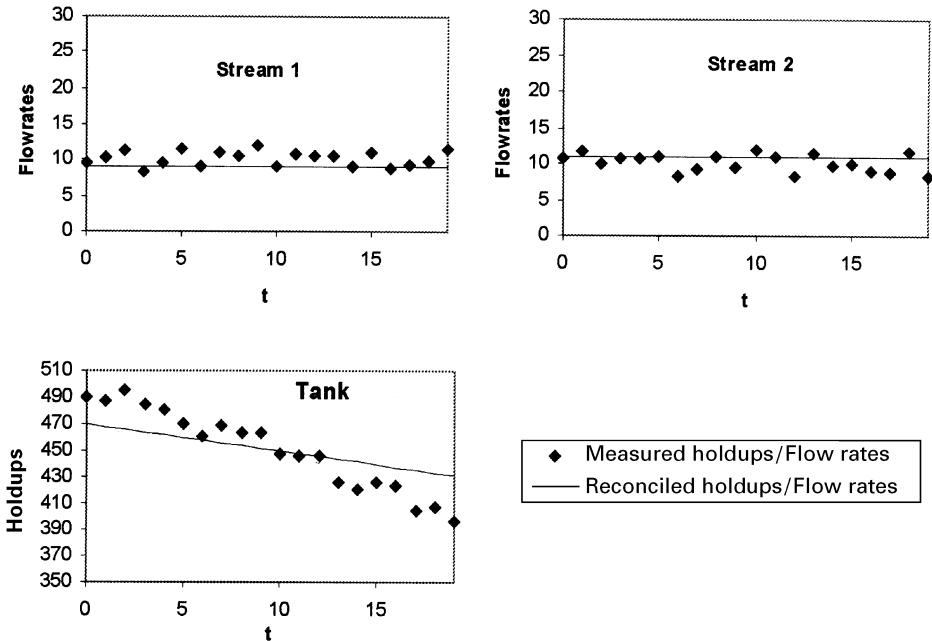


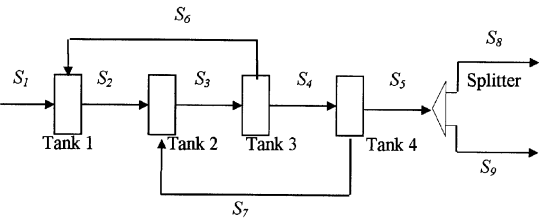Fig. 2. Reconciliation of a single tank with holdup bias.

Fig. 3. Example DDR1/DDR2.

(1997). There are 9 streams, 4 tanks and 1 splitter. In the case that all holdups and flowrates are measured except the holdup of Tank 3 and the flowrates of $S_2$ and $S_6$, we will call this example Problem DDR1. When all variables are measured, we call the example Problem DDR2.

Let us first confine our attention to Problem DDR1. The redundant holdups and flowrates are $V_4$, $S_4$, $S_5$, $S_7$, $S_8$ and $S_9$. A set of simulation runs was performed on this system. For each simulation run, a single gross error with a certain value is introduced and then a gross error is assumed and simulated. The dynamic integral measurement test (DIMT) (Bagajewicz and Jiang, 1997) with the above-presented extension to hold-up biases is used to identify variables suspected of having gross errors. The results are presented in Table 1.

In run 1, no gross error is introduced and data reconciliation with gross error detection is conducted by the proposed model without any biases or leaks. The result for run 1 shows no variables flagged by the DIMT or DIVMT. That means there is no any gross error in the original case.

In run 2, a constant bias in $S_4$ is introduced and no gross error is assumed in the model. After data reconciliation, $S_4$ $S_7$ $S_9$ and $V_4$ are flagged. For runs 3–5,

a constant bias in $S_4$ is also introduced. The last three columns indicate that the same results have been reached whether the gross error is assumed to be a constant bias in $S_4$, an order one bias in $V_4$ or a constant leak in $T_4$. The result of run 6 shows that the effect of a bias in $S_4$ can be eliminated by assuming a bias in $S_7$. The results from run 7–9 show that a constant bias in holdup cannot be detected. From runs 10–12, an order one bias in $V_4$ is introduced. The results show the equivalence for gross error detection of a constant bias in $S_4$, a constant leak in $T_4$ and an order one bias in $V_4$. This is also proved by the results of run 13–15.

Although the model can handle any bias order, we will restrict this paper to zero-order biases and leaks.

### Equivalency of gross errors

The model for data reconciliation with simultaneous gross error estimation cannot be solved for an arbitrary set of candidates, as certain combinations of candidates will lead to singular matrices. For example, in the case of DDR1, one cannot simulate a bias in $S_4$ and a leak in $T_4$ simultaneously, or a bias in $S_4$ and a holdup bias in $V_4$ simultaneously. The fact that singular matrices arise when two such equivalent gross errors are introduced is the manifestation of another phenomenon: the equivalency of those gross errors. Table 1 shows in runs 3–6 that after introducing a bias is $S_4$, the simulation of an order zero bias in $S_4$ or $S_7$, or a leak in $V_4$, or a first-order bias in $V_4$, have the same effect in the reconciliation. That is, the objective function has the same value, although the reconciled values of the variables corresponding to the measurements in question could vary. It is thus clear that, when the effect of two gross errors is the same, one cannot include both errors in the model

Table 1. Simulation results of Problems DDR1

| Run No. | Gross error introduced | Size of gross error introduced | Gross error simulated | Size of gross error estimated | Objective function value | Variables flagged by the DIMT/DIVMT |
|---|---|---|---|---|---|---|
| 1 | None | | None | | 353.8325 | None |
| 2 | Constant bias is $S_4$ | 1 | None | | 778.0468 | $S_4, S_7, S_9, V_4$ |
| 3 | Constant bias in $S_4$ | 1 | Constant bias in $S_4$ | 1.0943 | 349.6305 | None |
| 4 | Constant bias in $S_4$ | 1 | Order one bias in $V_4$ | $-1.0943t$ | 349.6305 | None |
| 5 | Constant bias in $S_4$ | 1 | Constnat leak in $T_4$ | 1.0943 | 349.6305 | None |
| 6 | Constant bias in $S_4$ | 1 | Constant bias in $S_7$ | $-1.0943$ | 349.6305 | None |
| 7 | Constant bias in $V_4$ | 10 | None | | 353.8325 | None |
| 8 | Constant bias in $V_4$ | 10 | Constant bias in $S_4$ | 0.0943 | 349.6305 | None |
| 9 | Constant bias in $V_4$ | 10 | Constant leak in $T_4$ | 0.0943 | 349.6305 | None |
| 10 | Order one bias in $V_4$ | $1*t$ | Order one bias in $V_4$ | $0.9057t$ | 349.6305 | None |
| 11 | Order one bias in $V_4$ | $1*t$ | Constant bias in $S_4$ | $-0.9057$ | 349.6305 | None |
| 12 | Order one bias in $V_4$ | $1*t$ | Constant leak in $T_4$ | $-0.9057$ | 349.6305 | None |
| 13 | Constant leak in $T_4$ | 1 | Constant leak in $R_4$ | 1.0943 | 349.6305 | None |
| 14 | Constant leak in $T_4$ | 1 | Constant bias in $S_4$ | 1.0943 | 349.6305 | None |
| 15 | Constant leak in $T_4$ | 1 | Order one bias in $V_4$ | $-1.0943t$ | 349.6305 | None |

because any linear combination of these gross errors will be acceptable, thus leading to singular matrices in the constraints of the problem. This leads us to the notion of equivalent sets of gross errors.

**Definition** (*Equivalent sets*). Two sets of gross errors are equivalent when they have the same effect in data reconciliation, that is, when simulating either one, leads to the same objective function value of data reconciliation.

We will first confine our attention to one tank and its connected flowrates and possible leaks. Assume also that only material balances are involved. The following equivalence holds for gross errors:

- *In an isolated tank, a constant (zero-order) flowrate bias is equivalent to a time-wise linear (first-order) holdup bias.*

This is illustrated in Fig. 4. Assume the measurements for flowrates $S_1, S_2$ and holdup $V_1$ given in Table 2. The same data reconciliation result can be reached by either assuming a constant bias in $S_1$ (case 1) or an order-one bias in $V_1$ (case 2).

Moreover, one can consider biases in $S_1$ and $S_2$. both being equivalent to the holdup bias simultaneously.

- *A constant tank leak is equivalent to a time-wise linear (first-order) holdup bias in the same tank. This is valid for isolated or non-isolated tanks.*

Consider the process in Fig. 5. Assume the measurements of $S_1, S_2$ and $V_1$ given in Table 3. The equivalencies are shown in this table.

- *All constant (zero order) bias in holdup are equivalent, as they do not affect data reconciliation, i.e. they cannot be identified.*

This is self-evident since the derivative of any constant is zero and thus a constant bias in holdup will not have any effect on constraints. This is also clear from the fact that the initial holdup of a tank is only present in the objective function and as such is not restricted to change.

- *In any isolated unit a flowrate bias in an inlet stream is equivalent to an exactly opposite bias in any output stream.*
- *In any isolated unit a flowrate bias in an output stream is equivalent to the same bias in any other output stream.*
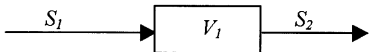
The last two are self-evident.

Fig. 4. Isolated tank.

Table 2. Isolated tank example

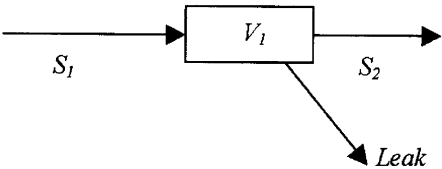|  |  | $S_1$ | $S_2$ | $V_1$ |
|---|---|---|---|---|
| Measurement |  | 4 | 4 | $1*t$ |
| Case 1 | Reconciled data | 5 | 4 | $1*t$ |
|  | Estimated bias | $-1$ |  |  |
| Case 2 | Recociled data | 4 | 4 | 0 |
|  | Estimated bias |  |  | $1*t$ |

Fig. 5. Isolated tank with a leak.

Table 3. Isolated tank with a leak

|  |  | $S_1$ | $S_2$ | $V_1$ | Leak |
|---|---|---|---|---|---|
| Measurement |  | 4 | 4 | $1*t$ |  |
| Case 1 | Reconciled data | 4 | 4 | $1*t$ |  |
|  | Estimated bias/leak |  |  |  | 1 |
| Case 2 | Recociled data | 4 | 4 | 0 |  |
|  | Estimated bias/leak |  |  | $1*t$ |  |

These equivalencies can be generalized as follows:

| $k$ order flowrate bias | $\leftrightarrow$ | $k+1$ order holdup bias (isolated tanks) |
|---|---|---|
| $k$ order flowrate bias in inlet stream | $\leftrightarrow$ | $k$ order flowrate of opposite sign in any outlet stream (isolated units) |
| $k$ order flowrate bias in outlet stream | $\leftrightarrow$ | k order flowrate of same size in any other outlet stream (isolated units) |
| $k$ order tank leak | $\leftrightarrow$ | $k+1$ order holdup bias |

Let us concentrate on the equivalency between leaks and holdup biases. In the absence of any additional information, any leak regardless of its order can be modeled by a tank measurement bias and a true leak in any linear combination. Thus, none of them can be included in a model simultaneously.

*Equivalencies of gross errors in multiple units systems*

We concentrate now on systems with more than one unit and the equivalencies of biases and leaks. We

first define the concept of gross error cardinality of a set

**Definition** (*gross error cardinality*). A set of variables has gross error cardinality $\Gamma = t$ if $t$ is the minimum number of gross errors that are required to represent all possible sets of gross errors in the variables.

We now define the sets that qualify to be equivalent to every possible set in the system.

**Definition** (*Basic subset*). A set of variables constitutes a basic subset of a system, when every set of gross errors is equivalent to a set of gross errors in the basic set.

As a corollary of this definition of basic subsets, we must note that (a) for every system, more than one basic subset may exist, and (b) all basic subsets are equivalent to each other, provided the right size of gross errors is used.

*Illustration.* Note first that the sets composed of all streams in Fig. 4 or Fig. 5 have gross error cardinality $\Gamma = 1$, as one can always represent any combination of gross errors with just one gross error. Consider now the process of Fig. 6 and assume that all streams are measured. The set $\Lambda = \{S_3, S_6\}$ has gross error cardinality $\Gamma(\Lambda) = 1$ as clearly a gross error in one of them can be alternatively placed in the other without change in the result. We now illustrate the fact that the set $\Lambda = \{S_2, S_4, S_5\}$ has gross error cardinality $\Gamma(\Lambda) = 2$. As shown in Table 4, a bias of $(+1)$ in $S_2$, a bias of $(-1)$ in $S_4$ and a bias of $(+2)$ in $S_5$ can be represented by three alternative sets of two gross errors (Cases 2–4).

In certain cases a set of two or even three gross errors can be represented by one gross error. This is illustrated in Table 5. Consider an error of $(+1)$ in $S_2$ and a set of two gross errors of $(-1)$ in $S_4$ and $S_5$. Case 2 shows that the system can be represented by two gross errors in $S_4$ and $S_5$ and Case 3 shows that it
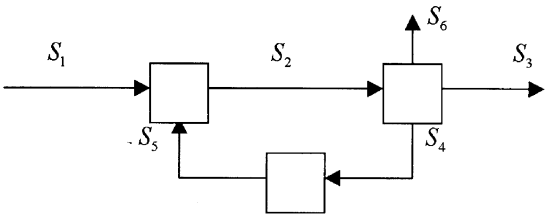


Fig. 6. Illustration of cardinality

can be represented by one gross error in $S_2$. This fact should not lead to conclude anything about the gross error cardinality of the set, which continues to be $\Gamma(S_2, S_4, S_5) = 2$.

We now relate the concept of cardinality of a set to properties of the incidence matrix $D$. Iordache *et al.* (1985) proved in a theorem that the measurement test used for steady-state data reconciliation is equal for all variables for which the columns of the incidence matrix are proportional to one column. We will now prove a similar theorem for gross errors, extending the notion to linear combination of columns.

*Gross error cardinality theorem*: Let $m$ columns $[d_1 \; d_2 \; \ldots \; d_m]$ of the system matrix $D$ correspond to a set of flowrate variables $\Lambda$. The set has gross error cardinality $\Gamma = t$ if rank $[d_1 \; d_2 \; \ldots \; d_m] = t$.

*Proof*: Let $[d_1 \; d_2 \; \ldots \; d_m]$ be the set of columns of $D$ corresponding to the subset of variables $\Lambda$. Then we want to prove that if a basis for $\Lambda$ has $t$ elements, i.e. only $t$ vectors are *l.i.*, then all gross errors can be expressed by a set of $t$ gross errors.

First note that the linear dependence of the columns in $D$ holds for the corresponding columns of

$$\begin{bmatrix} A_R \\ C_R \end{bmatrix}.$$

Table 4. Illustration of gross error cardinality in $\{S_2, S_4, S_5\}$

|  |  | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ |
|---|---|---|---|---|---|---|---|
|  | Measurement | 12 | 18 | 10 | 4 | 7 | 2 |
| Case 1 | Reconciled data | 12 | 17 | 10 | 5 | 5 | 2 |
| (Bias in $S_2, S_4, S_5$) | Estimated biases |  | 1 |  | −1 | 2 |  |
| Case 2 | Reconciled data | 12 | 18 | 10 | 6 | 6 | 2 |
| (Bias in $S_4, S_5$) | Estimated biases |  |  |  | −2 | 1 |  |
| Case 3 | Reconciled data | 12 | 19 | 10 | 7 | 7 | 2 |
| (Bias in $S_2, S_4$) | Estimated biases |  | −1 |  | −3 |  |  |
| Caxe 4 | Reconciled data | 12 | 16 | 10 | 4 | 4 | 2 |
| (Bias in $S_2, S_5$) | Estimated biases |  | 2 |  |  | 3 |  |

Table 5. Illustration of gross error cardinality in $\{S_2, S_4, S_5\}$

| | | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ |
|---|---|---|---|---|---|---|---|
| Measurement | | 12 | 18 | 10 | 4 | 4 | 2 |
| Case 1 | Reconciled data | 12 | 17 | 10 | 5 | 5 | 2 |
| (Bias in $S_2, S_4, S_5$) | Estimated biases | | 1 | | $-1$ | $-1$ | |
| Case 2 | Reconciled data | 12 | 18 | 10 | 6 | 6 | 2 |
| (Bias in $S_4, S_5$) | Estimated biases | | | | $-2$ | $-2$ | |
| Case 3 | Reconciled data | 12 | 16 | 10 | 4 | 4 | 2 |
| (Bias in $S_2$,) | Estimated biases | | 2 | | | | |

Before we start, let us notice that the following holds:

$$\left\{ \begin{array}{l} \text{Min}(f + \delta - f^+)^T R_F^{-1}(f + \delta - f^+) \\ \quad + (v_{ii} - v^+)^T R_V^{-1}(v_{ii} - v^+) \\ B_R \dfrac{dv}{dt} = A_R f \\ C_R f = 0 \end{array} \right\}$$

$$\Leftrightarrow \left\{ \begin{array}{l} \text{Min}(\hat{f} - f^+)^T R_F^{-1}(\hat{f} - f^+) \\ \quad + (v_{ii} - v^+)^T R_V^{-1}(v_{ii} - v^+) \\ B_R \dfrac{dv}{dt} = A_R(\hat{f} - \delta) \\ C_R(\hat{f} - \delta) = 0 \end{array} \right\} \quad (68)$$

where the new variable $\hat{f} = f + \delta$ was introduced. Let

$$O = \begin{bmatrix} A_R \\ C_R \end{bmatrix} = [o_1 \quad o_2 \quad \cdots \quad o_m] \quad (69)$$

and assume that $t$ vectors form a basis $\Lambda^{li}$. When $t = m$ the theorem holds trivially. When $t < m$, any element $d_r$ not belonging to $\Lambda^{li}$ can be represented as a linear combination of the elements in $\Lambda^{li}$. We recognize that $\Lambda^{li}$ is one basic subset of $\Lambda$. In addition, as anticipated, the following holds:

$$\left\{ d_r = \sum_{j \in \Lambda^{li}} \xi_{rj} d_j \right\} \Leftrightarrow \left\{ o_r = \sum_{j \in \Lambda^{ij}} \xi_{rj} o_j \right\} \quad (70)$$

by the fact that one gets the canonical form of $D$ by column form of $D$ by column rearrangement and row linear combination. Thus, the following holds:

$$\sum_{j \in \Lambda} o_j \delta_j = \sum_{j \in \Lambda^{li}} o_j \left( \delta_j + \sum_{r \in \Lambda^{id}} \xi_{rj} \delta_r \right) \quad (71)$$

which proves the theorem. Q.E.D.

**Corollary 1.** *A set with gross error cardinality $\Gamma = t$ contains subsets with gross error cardinality $k = 1, \ldots, t - 1$.*

*Proof*: Sets of smaller cardinality can be constructed by eliminating an element from $\Lambda^{li}$ and all the vectors that cannot be expressed as linear combinations of the elements left in $\Lambda^{li}$. Q.E.D.

**Corollary 2.** A l.d. set with gross error cardinality $\Gamma = t$ always contains more than one basic subsets with gross error cardinality $\Gamma = t$ and these basic subsets are equivalent to each other.

*Proof*: This is self evident from equation (71).
Q.E.D.

**Corollary 3.** *When the data reconciliation with gross error estimation is applied to a system with gross error cardinality $\Gamma = t$, only the gross errors corresponding to a basic subset of $\Lambda$ can be introduced simultaneously.*

*Proof*: Given the linear dependency, the values of gross errors are undetermined and thus the problem will become singular. Q.E.D.

**Corollary 4.** In a l.d. set of gross error cardinality $\Gamma = t$, a single gross error $\delta_r = \Delta$ in one stream not belonging to a basic subset can be represented by gross errors of size $\xi_{rj}\Delta$ in the basic subset and vice versa.

*Proof*: The corollary follows from the proof of the theorem. When $\delta_j = \xi_{rj}\Delta$ and $\delta_r = 0$, then (71) is unaltered if $\delta_j = 0$ and $\delta_r = \Delta$. Q.E.D.

*Illustration of the gross error cardinality Theorem.* Consider the process in Fig. 6. The whole process $\Lambda = \{S_1, S_2, S_3, S_4, S_5, S_6\}$ is a set of gross error cardinality $\Gamma = 3$. Matrix $D$ is

$$D = \begin{bmatrix} 1 & -1 & & & 1 & \\ & 1 & -1 & -1 & & -1 \\ & & & 1 & -1 & \end{bmatrix} \quad (72)$$

and rank $D = 3$. Consider the presence of a gross error of size $(+2)$ in $S_3$, a gross error of size $(+2)$ in $S_5$ and a gross error of size $(-1)$ in $S_6$. As illustrated

Table 6. Illustration of a set of gross error cardinality $\Gamma = 3$

| | | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ |
|---|---|---|---|---|---|---|---|
| Measurement | | 12 | 17 | 12 | 5 | 7 | 1 |
| Case 1 | Reconciled data | 12 | 17 | 10 | 5 | 5 | 2 |
| (Candidates: $S_3, S_5, S_6$) | Estimated biases | | | 2 | | 2 | $-1$ |
| Case 2 | Reconciled data | 13 | 20 | 12 | 7 | 7 | 1 |
| (Candidates: $S_1, S_2, S_4$) | Estimated biases | $-1$ | $-3$ | | $-2$ | | |
| Case 3 | Reconciled data | 12 | 17 | 11 | 5 | 5 | 1 |
| (Candidates $S_3, S_5$) | Estimated biases | | | 1 | | 2 | |
| Case 4 | Reconciled data | 12 | 17 | 12 | 5 | 5 | 0 |
| (Candidates $S_5, S_6$) | Estimated biases | | | | | 2 | 1 |

in Case 1 of Table 6, these three gross errors can be modeled by gross errors in another three *l.i.* Streams (Case 2). In Case 3 it is shown that the system can be reduced to two gross errors. This is due to the fact that $S_3$ and $S_6$ form a set of gross error cardinality $\Gamma = 1$. Case 4 shows this fact used to express errors in Case 1 by a set of errors in $S_5$ and $S_6$.

*Illustration of Corollary* 4. Consider now the problem DDR2. Consider the case of a gross error in stream $S_3$. This stream belongs to two *l.d.* sets of gross error cardinality $\Gamma = 2$: $\{S_2, S_3, S_6\}$ and $\{S_3, S_4, S_7\}$. Thus the following relations can be written:

$$o_3 = -o_2 - o_6, \quad (73)$$

$$o_3 = -o_4 - o_7. \quad (74)$$

Thus a gross error in $S_3$ can be represented by a set of gross errors of exactly opposite size in $S_2$ and $S_6$, or by gross errors of exactly opposite size in $S_4$ and $S_7$. In addition, due to the fact that $S_3$ belongs to these two sets, one can write

$$o_3 = -\alpha(o_2 + o_6) - (1-\alpha)(o_4 + o_7) \quad (75)$$

which also illustrates that a single gross error can be equivalent to gross errors in two or more basic subsets.

*Illustration of alternative basic subsets.* In the case of Fig. 6, the set $\{S_2, S_4, S_5\}$ has gross error cardinality $\Gamma = 2$, as it can be verified from the fact that rank $[d_2\ d_4\ d_5] = 2$, i.e. the three vectors are l.d. and only any two of them are l.i. Assume a true bias in $S_2(+1)$ and in $S_4(-1)$. As shown in Table 7, the effect of these biases can be equivalently expressed by assuming biases in $S_2$ and $S_5$, or in $S_4$ and $S_5$.

*Inclusion of the environmental node.* First note that some linearly dependent set of variables in $D$ corresponds to a closed loop in the system. If the environmental node is included, then all l.d. set of variables correspond to a closed loop, some of them including the environmental node.

As an illustration, consider the case of a single tank as in Fig. 7. The incidence matrix is $D = [1 \ -1]$ and the set of all flowrates is a set of gross error cardinality $\Gamma = 1$, i.e. a bias in the inlet stream is equivalent to an opposite bias in the outlet stream.

Note also that the addition of the environmental node does not change the cardinality of the set. Indeed when the environmental node is added the columns of the augmented matrix

$$D^A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (76)$$

are still l.d. Consider now the example of Fig. 6. The augmented matrix for this system is

$$D^A = \begin{bmatrix} 1 & -1 & & & 1 & \\ & 1 & -1 & -1 & & -1 \\ & & & 1 & -1 & \\ 1 & & 1 & & & 1 \end{bmatrix}. \quad (77)$$

First note that the row corresponding to the environmental node has nonzero elements in those streams that are inlet or outlet to the system. If one inlet stream or outlet stream is expressed in terms of l.i. set, this l.i. set contains one and only one outlet stream. If it contains more than one, the system would not be l.i. By introducing the environmental node the same linear dependency holds.

Thus, by adding the environmental node we can merge both cases: sets of the same cardinality consisting of streams in a loop and sets consisting of streams in a path from a feed/product stream to any feed/product stream into a single category of sets in a loop.

**Definition** (*Augmented graph*). The graph consisting of the original graph representing the flowsheet with the addition of the environmental node is called Augmented graph.

Table 7. Alternative basic subsets

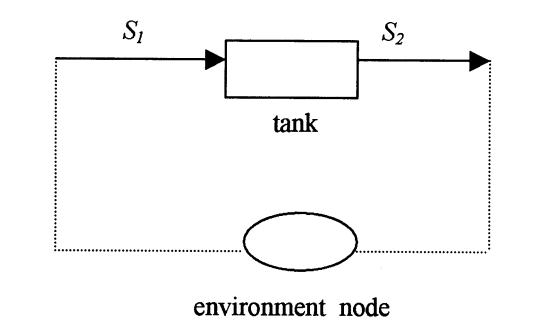| | | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ |
|---|---|---|---|---|---|---|---|
| Measurement | | 12 | 18 | 10 | 4 | 5 | 2 |
| Case 1 | Reconciled data | 12 | 17 | 10 | 5 | 5 | 2 |
| (Biases in $S_2, S_4$) | Estimated biases | | 1 | | − 1 | | |
| Case 2 | Reconciled data | 12 | 16 | 10 | 4 | 4 | 2 |
| (Biases in $S_4, S_5$) | Estimated biases | | 2 | | | 1 | |
| Case 3 | Reconciled data | 12 | 18 | 10 | 6 | 6 | 1 |
| (Biases in $S_4, S_5$) | Estimated biases | | | | − 2 | − 1 | |



Fig. 7. Augmented graph for a single tank.

We now link our notion of cardinality to graph theory.

**Loop Theorem.** *Every loop of m streams is a set with Gross Error Cardinality $\Gamma = m - 1$.*

*Proof*: Consider a loop consisting of $m$ streams $\{S_1, S_2, \ldots, S_{m-1}, S_m\}$. Then the incidence matrix of such a loop has the following form:

$$D = \begin{bmatrix} -1 & & & & & 1 \\ 1 & -1 & & & & \\ & 1 & & & & \\ & & \ddots & & & \\ & & & -1 & \\ & & & 1 & -1 \end{bmatrix}. \quad (78)$$

Clearly, the system is l.d. and rank $D = m - 1$.

Q.E.D

**Augmented graph theorem.** *The gross error cardinality $\Gamma$ of an augmented graph corresponding to an open system is equal to the number of process units (i.e. excluding the environmental node).*

*Proof*: We first recall that an open system is one that has streams connected to the environmental node, whereas a closed system is one that has no inlet or outlet streams, as in closed refrigeration systems. A path between two vertices (units in our case) of a graph is an alternating sequence of vertices (units) and edges (streams), where two consecutive edges are inlet and outlet of the vertex in between them. In addition a connected graph is a graph where there exists a path between any two vertices.

The theorem proof follows from the fact that the incidence matrix of an augmented graph has fewer rows than columns and that the number of rows is equal to the number of units. This limits the number of l.i. vectors to the number of rows. We now prove that, if the graph is connected, then the rank of the incidence matrix is equal to the number of rows. If the graph is connected, then a path from one unit to another can be represented by a matrix that has the following form:

$$D = \begin{bmatrix} -1 & & & & & \\ 1 & -1 & & & & \\ & 1 & & & & \\ & & \ddots & & & \\ & & & -1 & \\ & & & 1 & -1 \end{bmatrix} \quad (79)$$

which has rank equal to the number of rows. In particular, a path that includes all units exists, which proves the theorem. If the graph is not connected, the theorem applies for all open connected systems.

Q.E.D

**Corollary 1.** *The maximum number of gross errors that can be modeled in an open system is equal to the number of process units.*

**Corollary 2.** *The maximum number of gross errors that can be modeled in a closed connected system is equal to the number of process units minus one.*

*Proof*: In a closed connected system a path that contains all units is a loop, and therefore the loop theorem applies. Q.E.D.

*Extension of the gross error cardinality theorem to leaks*

We will now extend the conclusions of the cardinality theorem to leaks. As leaks are part of the

constraints, the proof of the theorem hold when matrix $H$ is augmented to include leak streams. As leaks of order $k$ are equivalent to the corresponding tank holdup bias of order $k + 1$, by including leaks we are actually considering all possibilities. Since zero-order bias in holdup are undetectable, and in the absence of any other criterion holdup biases can be ignored from the analysis if leaks are included.

*Illustration.* Figure 8 shows the augmented graph for problem DDR1. Table 8 shows all the loops that have been identified.

We now illustrate the equivalent gross error behavior of Problem DDR2 in the runs summarized in Table 9.

### Determination of equivalent basic sets

Assume that a set of gross errors $\Psi$ has been identified. Assume now that the set has $t$ elements. The following is true:

(1) If the number of gross errors $t$ is equal to the gross error cardinality of the whole flowsheet, then the set $\Psi$ is a basic subset of the whole flowsheet.
(2) If the number of gross errors $t$ is smaller to the gross error cardinality of the whole flowsheet, then the set $\Psi$ is a basic subset of a proper subset of the flowsheet. Moreover, this subset has gross error cardinality $t$.

Thus, in the second case, we face the task to identify:

(a) The proper subset $\Lambda$ of which $\Psi$ is a Basic Subset.
(b) All other equivalent basic subsets. This information is important as the true location of all gross errors may not be in the set $\Psi$, but rather in some of the equivalent basic subsets of the unknown set $\Lambda$.

We construct $\Lambda$ by simply adding to the set $\Psi$ all variables that will not increase the gross error cardinality of the resulting set. Consider a candidate variable $S_i$. If $S_i \cup \Psi$ form a loop in the augmented graph, then, according to the loop theorem, the resulting gross error cardinality of the set is that of $\Psi$. Conversely, if $S_i \cup \Psi$ do not form a loop, then the gross
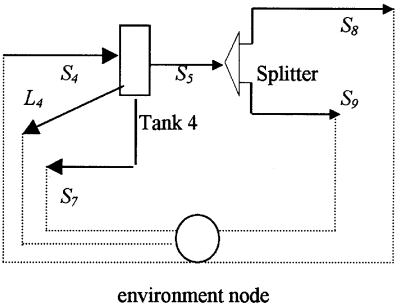


Fig. 8. Augmented graph for problem DDR1.

Table 8. Loops in DDR1

| Loop | Streams | Gross error cardinality |
|---|---|---|
| 1 | $S_4, S_7$ | 1 |
| 2 | $S_4, L_4$ | 1 |
| 3 | $L_4, S_7$ | 1 |
| 4 | $S_8, S_9$ | 1 |
| 5 | $L_4, S_5, S_8$ | 2 |
| 6 | $L_4, S_5, S_9$ | 2 |
| 7 | $S_4, S_5, S_8$ | 2 |
| 8 | $S_4, S_5, S_9$ | 2 |
| 9 | $S_7, S_5, S_8$ | 2 |
| 10 | $S_7, S_5, S_9$ | 2 |

error cardinality of the resulting set is larger, i.e. $\Gamma(S_i \cup \Psi) > \Gamma(\Psi)$. This can be easily proven by noticing that not forming a loop is synonym of linear independence in the augmented incidence matrix.

Therefore, $\Lambda$ can be constructed by adding all streams that are in a loop with all subsets of $\Psi$. Thus, the list of equivalent basic sets $\{\varphi_i\}$ is also easy to construct. For every subset of $n$ variable $\Omega$ in $\Psi$, identify the variables $S_i$, such that $S_i \cup \Omega$ form a loop. All subsets of $n$ variables in $S_i \cup \Omega$ can replace $\Omega$.

*Illustrations.* Consider the Problem DDRI. Assume first that $\Psi = \{S_8\}$. Clearly, from Fig. 8, one can verify that there only $S_9$ forms a loop with $S_8$. Thus,

Table 9. Illustration of equivalent sets gross error behavior in DDR2

| Loop | Streams in $\Lambda^{li}$ (basic subsets) | Gross error introduced | Gross error simulated | Magnitude of gross error estimated | | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|---|---|---|
| Loop 1 | $S_3, S_6$ | $S_3(+1)$ | $S_3, S_6$ | 0.959, | 0.959 | 794.158 | None |
| $(S_2, S_3, S_6)$ | $S_2, S_6$ | $S_6(+1)$ | $S_2, S_6$ | $-0.959,$ | $-0.001$ | 794.158 | None |
| | $S_2, S_3$ | | $S_2, S_3$ | $-0.959,$ | 0.001 | 794.158 | None |
| Loop 2 | $S_4, L_3$ | $V_3(-1*t)$ | $S_4, L_3$ | $-0.905,$ | 1.940 | 790.870 | None |
| $(S_4, L_3, L_4)$ | $L_3, L_4$ | $V_4(-1*t)$ | $L_3, L_4$ | 1.034, | 0.905 | 790.870 | None |
| | $S_4, L_4$ | | $S_4, L_4$ | 1.034, | 1.940 | 790.870 | None |
| Loop 3 | $S_8$ | $S_8(+1)$ | $S_8$ | 0.971 | | 793.054 | None |
| $(S_8 S_9)$ | $S_9$ | | $S_9$ | 0.971 | | 793.054 | None |

$\Lambda = \{S_8, S_9\}$ and the equivalent basic subset is $\varphi_1 = \{S_9\}$. In the case where $\Psi = \{S_4\}$ we can identify that $L_4$ and $S_7$ can form a loop with $S_4$. Thus, $\Lambda = \{S_4, S_7, L_4\}$ the two equivalent basic subsets are $\varphi_1 = \{S_7\}$ and $\varphi_2 = \{L_4\}$ and by eqivalency, a first order bias in $V_4$. From now on these equivalencies of leaks to holdup biases will be omitted.

Assume now that two gross errors have been found, $\Psi = \{S_4, S_5\}$. Then, $\Lambda$ corresponds to the whole flowsheet and the equivalent sets are all the pair of variables that do not form a loop: $\varphi_1 = \{S_4, S_8\}$, $\varphi_2 = \{S_4, S_9\}$, $\varphi_3 = \{S_5, S_8\}$, $\varphi_4 = \{S_5, S_9\}$, $\varphi_5 = \{S_5, S_7\}$, $\varphi_6 = \{S_5, L_4\}$, $\varphi_7 = \{S_8, S_7\}$, $\varphi_8 = \{S_8, L_4\}$, $\varphi_9 = \{S_9, S_7\}$, $\varphi_{10} = \{S_9, L_4\}$. Since DDR1 has gross error cardinality $\Gamma = 2$, any set of the above list of equivalent sets can represent any set of gross errors in the system.

Consider now problem DDR2. Assume that $\Psi = \{S_2, S_3\}$. By simple observation one can realize that $\Lambda = \{S_2, S_3, S_6\}$, and thus $\varphi_1 = \{S_2, S_6\}$, $\varphi_2 = \{S_3, S_6\}$. Finally assume that $\Psi = \{S_1, S_4, L_4\}$. By taking subsets $\Omega$ composed of one element of $\Psi$ at a time one can verify that $S_1$ and $L_1$ form a loop. By taking subsets $\Omega$ composed of two elements of $\Psi$ at a time one can verify that $S_4$, $L_4$ and $L_3$ form a loop. Finally, the union of $S_6$ and $\Psi$ form a loop. Thus $\Lambda = \{S_1, S_4, L_4, L_1, L_3, S_6\}$, and the equivalent basic sets are: $\varphi_1 = \{S_1, S_4, L_3\}$, $\varphi_2 = \{S_1, S_4, S_6\}$, $\varphi_3 = \{S_1, L_4, L_3\}$, $\varphi_4 = \{S_1, L_4, S_6\}$, $\varphi_5 = \{S_4, L_4, L_1\}$, $\varphi_6 = \{S_4, L_4, S_6\}$, $\varphi_7 = \{S_4, L_1, L_3\}$, $\varphi_8 = \{S_4, L_1, S_6\}$, $\varphi_9 = \{S_4, L_3, S_6\}$, $\varphi_{10} = \{L_4, L_1, L_3\}$, $\varphi_{11} = \{L_4, L_1, S_6\}$, $\varphi_{12} = \{L_4, L_3, S_6\}$.

### Degenerate cases

The equivalencies above are built in the assumption that the number of gross errors identified is equal to the real number of gross errors. However, there are examples where the actual number of gross errors can be larger than the number of gross errors identified. One such example has been shown in Table 5, where a set of two gross errors is equivalent to one gross error. These cases are rare, as they require the real gross errors to have equal sizes. As we shall see later, for systems with larger gross error cardinality, the restriction to certain gross error sizes exists but is not restricted to all of them being equal. In gross error detection, this poses an additional uncertainty. If a certain number of gross errors are detected, there is a possibility that the actual number of gross errors is larger. Although this has a low probability of taking place, as the sizes of the gross errors must add up properly, we will briefly illustrate this case. If one wants to find equivalent sets containing one more variable, the algorithm developed for the nondegenerate case can be modified as follows:

Construct $\Lambda$ by adding to the set $\Psi$ all pairs of variables that will increase the gross error cardinality of the resulting set by one. Therefore, $\Lambda$ can be constructed by adding all pairs of streams that are in a loop with all subsets of $\Psi$. The list of equivalent basic sets $\{\varphi_i\}$ is constructed by picking all subsets of $n + 1$ variables $\Omega$ in $\Psi$ that do not form a loop.

### Gross error exact detectability

Consider the case of one gross error present in the system. If the corresponding variable is not connected to the environmental node, i.e. it is not an inlet or outlet stream of the system, then it connects two units and, therefore, there is no other variable or leak that will form a loop, except for a recycle stream connecting the same two units. Thus, in the absence of loops between the two units connected by the stream containing the detected gross error, exact detectability is possible for one gross error, as no equivalent set exist. However, if the variable is an inlet/outlet stream, then the equivalent streams will be the other inlet/outlet streams to/from the same unit or leaks associated to the unit where the stream is connected. As the number of gross errors identified grows, the number of equivalent sets also grows. A useful practical conclusion arises:

*A set of gross errors can be exactly identified only if no subset of these variables forms a loop with the addition of another stream/leak. In other words, a set of gross errors can be exactly identified only if the corresponding set of columns of the incident matrix does not form a linearly dependent set with any other additional column.*

Degenerate cases are excluded from the above conclusion. An illustration of the above fact is the set $\Psi = \{S_2, S_4\}$ in Problem DDR2. Since $S_2$ or $S_4$ is not forming a loop with another stream/leak, there is no equivalent set to $\{S_2, S_4\}$ if we assume degeneracy does not happen.

### Strategy for gross error identification and estimation

All the key ingredients for a successful identification and estimation of gross errors have been presented. The DIMT/DIVMT allows the identification of candidates to have a flowrate and/or a holdup bias. Once a set of gross errors is proposed, then the Data Reconciliation with gross error estimation model can be run. As pointed out in Corollary 3 of the gross error cardinality theorem, these models cannot be run using a set of candidates that constitute a l.d. set with running into singularities. These singularities are a result of linearly dependent columns of the matrix $D$. Therefore, when a candidate set is used in the data reconciliation with gross error estimation model, the candidate set has to consist of a set of variables that correspond to columns of $D$ that are l.i.

Before we present our strategies for gross error detection we must note that after we obtain a list of candidates flagged by the DIMT and the DIVMT, we must face a decision of how many gross errors one desires to identify.

*Assumption. For any system, the probability of containing k gross errors is larger than the probability of containing k + 1 gross errors.*

In view of the above assumption, we now present a serial strategy to identify gross errors which has a principal objective to identify the least possible number of gross errors that after reconciliation will not flag any new candidate. The strategy is based on identifying one candidate for a gross error at a time and solving the data reconciliation model with bias detection using all suspects. Such strategy has been proposed by Keller *et al.* (1994) and consists of serial identification with collective estimation. We outline the strategy next:

### Serial identification with collective estimation strategy

1. Use the Data Reconciliation Model and the DIMT and DIVMT to detect variables in suspect of gross error. If there are $r$ variables in suspect ($r > 0$), go to step 2. Otherwise declare no gross error and stop.
2. Construct a list of candidates (LC) by including all $r$ variables that failed the DIMT and/or DIVMT. If any two members in LC form a loop, erase one of them. Create a list of confirmed gross errors (LCGE). This list is empty at this stage.
3. Run the data reconciliation with gross error estimation model simulating a gross error in all the members of the LCGE and one member of the LC at a time.
4. Determine which member of the LC leads to the smallest value of the objective function. Add that variable to the LCGE.
5. Run the DIMT and DIVMT for the run chosen in step 4. Erase all elements of the LC and place the latest flagged variables in LC. If there are any two members in LC forming a loop with any member(s) in LCGE, erase one of them. If LC is empty, go to step 6. Otherwise, go to step 3.

6. Determine all equivalent basic sets. Determine the sizes of the equivalent basic sets by using Corollary 4 of the gross error cardinality theorem.

Let us illustrate this method by using Problem DDR2. We added a constant bias $\delta = +0.5$ to $S_4$. The serial strategy is illustrated in Table 10. The gross error is exactly identified as no other equivalent basic set exists.

We tested the serial strategy in a variety of situation. In each case, the method has been successful in identifying the set of gross errors introduced. Table 11 shows the results of introducing a leak in tank 4. Table 12 shows the results of introducing a bias in $S_4$ and a leak in tank 4. Table 13 shows the effective location and estimation of two gross errors, a bias in $S_1$ and bias in $S_4$. Finally Tables 14 and 15 shows the introduction of gross errors in $S_2$ and $S_3$.

Note that in Table 13, the gross errors are successfully identified, even though $S_1$ is not initially flagged by the DIMT. In the case of Table 14, the strategy identifies one gross error, which through degeneracy is equivalent to the gross errors introduced. Table 15 also includes gross errors in $S_2$ ($+0.5$) and in $S_3$ ($+1.0$). The different size of the gross error eliminates the degeneracy.

The method was tested for the case of more than two gross errors. This is shown in Tables 16–22. Note first that the flowsheet of problem DDR2 has Gross Error Cardinality $\Gamma = 5$. Thus a maximum of 5 gross errors can be identified. If more gross errors are present they will be represented by the basic set identified.

Table 16 shows three gross errors, two of them associated to the same unit. Intermediate steps of the procedure are omitted.

Table 17 shows how an equivalent basic set of four gross errors introduced is found. The set of gross errors introduced, belongs to the set $\Lambda = \{S_1, S_2, L_1, L_2, S_4, S_8, S_9\}$ which has gross error cardinality $\Gamma = 4$. The set $\Lambda$ was obtained by adding to the set of gross errors, all the streams/leaks that will

Table 10. Serial strategy for Problem DDR2

| | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|
| Step 1 | None | | 984.827 | $S_3, S_4, S_7, S_9, L_3, L_4$ |
| Step 2 | LC: $S_3, S_4, S_7, S_9, L_3, L_4$ | LCGE: None | | |
| Step 3 | $S_3$ | $-0.347$ | 958.057 | $S_4, S_7, S_9, L_4$ |
| | $S_4$ | 0.580 | 791.465 | None |
| | $S_7$ | $-0.505$ | 855.518 | $S_3, S_4, L_3$ |
| | $S_9$ | $-0.085$ | 966.720 | $S_3, S_{4,5}, S_7, L_3, L_4$ |
| | $L_3$ | $-0.360$ | 959.234 | $S_4, S_7, S_9, L_4$ |
| | $L_4$ | 0.532 | 845.314 | $S_3, L_3$ |
| Step 4 | LCGE: $S_4(0.580)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: none | | | |

*Note*: Gross error in $S_4(+0.5)$

Table 11. Serial strategy for Problem DDR2

| | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|
| Step 1 | None | | 963.103 | $S_4, S_7, S_9, L_4$ |
| Step 2 | LC: $S_4, S_7, S_9, L_4$ | LCGE: None | | |
| Step 3 | $S_4$ | 0.459 | 842.218 | $S_3, L_3$ |
| | $S_7$ | $-0.551$ | 808.813 | None |
| | $S_9$ | $-0.093$ | 941.676 | $S_4, S_5, S_7, L_4$ |
| | $L_4$ | 0.591 | 791.095 | None |
| Step 4 | LCGE: $L_4(0.591)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: none | | | |

*Note*: Gross error in $L_4(+0.5)$

Table 12. Serial strategy for Problem DDR2

| | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|
| Step 1 | None | | 1370.352 | $S_4, S_7, S_9, L_4$ |
| Step 2 | LC: $S_4, S_7, S_9, L_4$ | LCGE: None | | |
| Step 3 | $S_4$ | 0.959 | 842.218 | $S_3, L_3$ |
| | $S_7$ | $-0.975$ | 888.068 | $S_3, S_4, L_2, L_3$ |
| | $S_9$ | $-0.149$ | 1315.040 | $S_4, S_5, S_7, L_4$ |
| | $L_4$ | 1.032 | 845.314 | $S_3, L_3$ |
| Step 4 | LCGE: $S_4$ | | | |
| Step 5 | LC: $S_3, L_3$ | | | |
| Step 3 | $S_4, S_3$ | 1.071, 0.445 | 805.458 | None |
| | $S_4, L_3$ | 1.095, 0.560 | 790.870 | None |
| Step 4 | LCGE: $S_4$, (1.095), $L_3(0.560)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: $S_4(0.534)$, $L_4(0.560)$ $L_3(-0.534)$, $L_4(1.095)$ | | | |

*Note*: Gross errors in $S_4(+0.5)$ and $L_4(+0.5)$

Table 13. Serial strategy for Problem DDR2

| | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|
| Step 1 | None | | 1253.481 | $S_2, S_4, S_6, S_7, S_9, L_1, L_4$ |
| Step 2 | LC: $S_2, S_4, S_6, S_7, S_9, L_1, L_4$ | LCGE: None | | |
| Step 3 | $S_2$ | $-0.334$ | 1190.914 | $S_3, S_4, S_6, S_7, S_9, L_3, L_4$ |
| | $S_4$ | 0.776 | 907.585 | $S_2, S_6, L_1$ |
| | $S_6$ | 0.473 | 1130.421 | $S_4, S_7, S_9, L_4$ |
| | $S_7$ | $-0.736$ | 978.764 | $S_3, S_4, S_6, L_1, L_2, L_3$ |
| | $S_9$ | $-0.117$ | 1219.392 | $S_2, S_4, S_5, S_6, S_7, L_1, L_4$ |
| | $L_1$ | 0.374 | 1179.007 | $S_3, S_4, S_6, S_7, S_9, L_3, L_4$ |
| | $L_4$ | 0.781 | 952.943 | $S_3, S_4, S_6, L_1, L_3$ |
| Step 4 | LCGE: $S_4$ | | | |
| Step 5 | LC: $S_2, S_6, L_1$ | | | |
| | $S_4, S_2$ | 0.827, $-0.433$ | 803.576 | None |
| Step 3 | $S_4, S_6$ | 0.726, 0.371 | 833.302 | $L_3$ |
| | $S_4, L_1$ | 0.827, 0.471 | 791.028 | None |
| Step 4 | LCGE: $S_4(0.827)$, $L_1(0.471)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: $S_4(0.827)$, $S_1(0.471)$ | | | |

*Note*: Gross errors in $S_1(+0.5)$ and $S_4(+0.75)$

Table 14. Serial strategy for Problem DDR2

| | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|
| Step 1 | None | | 951.242 | $S_2, S_6, L_1$ |
| Step 2 | LC: $S_2, S_6, L_1$ | LCGE: None | | |
| | $S_2$ | 0.437 | 843.841 | $S_3, S_6, L_3$ |
| | $S_6$ | $-0.534$ | 794.502 | None |
| Step 3 | $L_1$ | $-0.454$ | 842.034 | $S_3, S_6, L_3$ |
| Step 4 | LCGE: $S_6(-0.534)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: none | | | |
| | Degenerate equivalent set: $S_2(0.534), S_3(0.534)$ | | | |

*Note*: Gross error in $S_2(+0.5)$ and $S_3(+0.5)$

Table 15. Serial strategy for Problem DDR2

| | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|
| Step 1 | None | | 1054.638 | $S_2, S_3, S_6, S_7, L_1, L_3$ |
| Step 2 | LC: $S_2, S_3, S_6, S_7, L_1, L_3$ | LCGE: None | | |
| Step 3 | $S_2$ | 0.333 | 992.500 | $S_3, S_6, S_7, L_3$ |
| | $S_3$ | 0.715 | 940.997 | $S_2, S_6, L_1$ |
| | $S_6$ | $-0.630$ | 836.785 | $S_3, L_3$ |
| | $S_7$ | $-0.185$ | 1037.230 | $S_2, S_3, S_4, S_6, L_1, L_3$ |
| | $L_1$ | $-0.376$ | 979.380 | $S_3, S_6, S_7, L_3$ |
| | $L_3$ | 0.738 | 946.990 | $S_2, S_6, L_1$ |
| Step 4 | LCGE: $S_6$ | | | |
| Step 5 | LC: $S_3, L_3$ | | | |
| Step 3 | $S_6, S_3$ | $-0.542, 0.459$ | 794.158 | None |
| | $S_6, L_3$ | $-0.543, 0.443$ | 802.148 | None |
| Step 4 | LCGE: $S_6(-0.542), S_3(0.459)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: | | | |
| | $S_6(-1.001), S_2(-0.459)$ | | | |
| | $S_2(0.542), S_3(1.001)$ | | | |

*Note*: Gross errors in $S_2(+0.5)$ and $S_3(+1.0)$

Table 16. Serial strategy for Problem DDR2

| | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|
| Step 1 | None | | 1211.860 | $S_2, S_3, S_4, S_6, S_7, S_9, L_1, L_3, L_4$ |
| Step 2 | LC: $S_2, S_3, S_4, S_6, S_7, S_9, L_1, L_3, L_4$ | LCGE: None | | |
| Step 4 | LCGE: $S_6(-0.484), L_1(0.944), L_4(0.595)$ | | | |
| Step 5 | LC: None | | | |
| | Equivalent sets: | | | |
| Step 6 | $\{S_1, S_4, L_3\}, \{S_1, S_4, L_3\}, \{S_1, S_4, S_6\}, \{S_1, L_3, L_4\}, \{S_1, S_6, L_4\}, \{S_4, L_4, L_1\}$ | | | |
| | $\{S_4, L_4, S_6\}, \{S_4, L_1, L_3\}, \{S_4, L_1, S_6\}, \{S_4, L_3, S_6\}, \{L_4, L_1, L_3\}, \{L_4, L_3, S_6\}$ | | | |

*Note*: Gross errors in $S_1(+0.5), S_4(+0.5)$ and $L_3(+1)$

allow the formation of a single loop. These are indicated in dotted lines in Fig. 9. The following loops can be formed: $\{S_1, L_1\}$, $\{S_1, S_2, L_2\}$ and $\{S_8, S_9\}$.

Table 18 shows the effect of introducing 5 gross errors. The basic set found consists of five streams as the set of gross errors introduced belongs to the set $\Lambda = \{S_1, S_2, S_4, S_8, L_1, L_2, L_3, S_6, S_7, L_4, S_9\}$, which

Table 17. Serial strategy for Problem DDR2

|       | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|-------|-----------------------|-----------------------------------|--------------------------|----------------------------------|
| Step 1 | None | | 4998.367 | $S_1 \sim S_9, L_1, \sim L_4$ |
| Step 2 | LC: $S_1 \sim S_8, L_2 \sim L_4$ | LCGE: None | | |
| Step 4 | LCGE: $S_2(2.032), S_4(1.565), S_8(0.478), L_2(0.964)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: | | | |
| | $\{S_1, S_2, S_4, S_8\}, \{S_1, S_4, S_8, L_2\}, \{S_2, S_4, S_8, L_1\}, \{S_2, S_4, S_8, L_2\}, \{S_4, S_8, L_1, L_2\}$ | | | |
| | $\{S_1, S_2, S_4, S_9\}, \{S_1, S_4, S_9, L_2\}, \{S_2, S_4, S_9, L_1\}, \{S_2, S_4, S_9, L_2\}, \{S_4, S_9, L_1, L_2\}$ | | | |

*Note*: Gross errors in $S_1(-1.0), S_2(+1.0), S_4(+1.5)$ and $S_8(+0.5)$

Table 18. Serial strategy for Problem DDR2

|       | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|-------|-----------------------|-----------------------------------|--------------------------|----------------------------------|
| Step 1 | None | | 39077.517 | $S_2 \sim S_9, L_2, L_4$ |
| Step 2 | LC: $S_2 \sim S_8, L_2, L_4$ | LCGE: None | | |
| Step 4 | LCGE: $S_2(3.978), S_4(2.831), S_6(-1.063), S_8(-2.519), L_2(0.924)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: 429 sets | | | |

*Note*: Gross errors in $S_1(-2), S_2(+3), S_4(+1.75), S_5(-1)$ and $S_8(-3.5)$

Table 19. Serial strategy for Problem DDR2

|       | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|-------|-----------------------|-----------------------------------|--------------------------|----------------------------------|
| Step 1 | None | | 10,222.355 | $S_1, S_2, S_4, \sim, S_9, L_1$ |
| Step 2 | LC: $S_1, S_2, S_4, \sim, S_8,$ | LCGE: None | | |
| Step 4 | LCGE: $S_4(-0.409), S_5(-1.982), S_6(0.953)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: None | | | |

*Note*: Gross errors in $S_1(+1), S_4(+1.5), S_8(+2)$, and $L_3(+1)$

Table 20. Serial strategy for Problem DDR2

|       | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|-------|-----------------------|-----------------------------------|--------------------------|----------------------------------|
| Step 1 | None | | 37755.671 | $S_1 \sim S_9, L_1, L_3, L_4$ |
| Step 2 | LC: $S_1 \sim S_8, L_3, L_4$ | LCGE: None | | |
| Step 4 | LC: $S_4(-1.909), S_5(-3.982), S_6(0.953)$ | | | |
| Step 5 | LC: None | | | |
| Step 6 | Equivalent basic sets: None | | | |

*Note*: Gross errors in $S_1(+1), S_4(+2), S_8(+3), S_9(+1)$ and $L_3(+3)$

has gross error cardinality $\Gamma = 5$. Note the large number of equivalent sets.

### Degeneracy

We now show two examples where one has to resort to the concept of degeneracy to explain why the set of errors found is equivalent to the one introduced. Although these examples are unlikely to occur in practice they are introduced to show the intricacies of gross error detection.

Table 19 shows a degenerate case for four gross errors. The four gross errors introduced belong to the

Table 21. Serial strategy for Problem DDR2

|  | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|
| Step 1 | None |  | 37742.934 | $S_1 \sim S_9, L_1, L_2, L_4$ |
| Step 2 | LC: $S_1 \sim S_8, L_2, L_4$ | LCGE: None |  |  |
| Step 4 | LCGE: $S_2(5.040), S_7(-2.831), S_8(-2.519), L_2(0.844), L_3(0.787)$ |  |  |  |
| Step 5 | LC: None |  |  |  |
| Step 6 | Equivalent basic sets: 429 sets |  |  |  |

*Note*: Gross errors in $S_1(-2), S_2(+3), S_4(+1.75), S_5(-1), S_8(-3.5)$ and $L_3(+2.5)$

Table 22. Serial strategy for Problem DDR2

|  | Gross error simulated | Magnitude of gross error estimated | Objective function value | Variables flagged by DIMT/DIVMT |
|---|---|---|---|---|
| Step 1 | None |  | 45479.062 | $S_1 \sim S_9, L_1, L_2, L_4$ |
| Step 2 | LC: $S_1 \sim S_8, L_2, L_4$ | LCGE: None |  |  |
| Step 4 | LCGE: $S_2(5.040), S_7(-2.831), S_8(-3.019), L_2(0.844), L_3(0.732)$ |  |  |  |
| Step 5 | LC: None |  |  |  |
| Step 6 | Equivalent basic sets: 429 sets |  |  |  |

*Note*: Gross errors in $S_1(-2), S_2(+3), S_4(+1.75), S_5(-1), S_8(-3.5), S_9(-0.5)$ and $L_3(+2.5)$
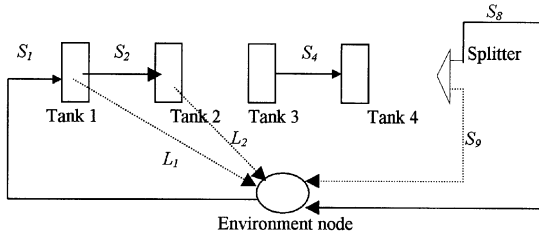


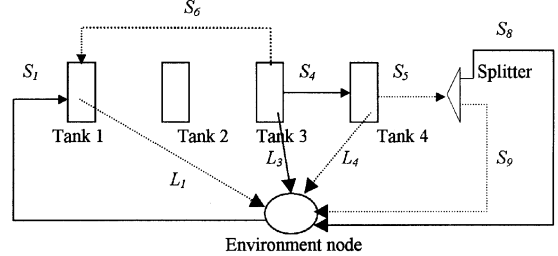Fig. 9. Loops formed by the gross errors introduced in Table 17.



Fig. 10. Loops formed by the gross errors introduced in Table 19.

following set: $\Lambda = \{S_1, L_1, S_6, S_4, L_3, L_4, S_5, S_8, S_9\}$ which has gross error cardinality $\Gamma = 4$ (four units). This is illustrated in Fig. 10. Indeed, $\{L_3, S_1, S_6\}$ form a loop, so do $\{L_3, S_4, L_4\}$, $\{L_3, S_4, S_5, S_8\}$ and $\{S_8, S_9\}$. However, the degenerate situation is obtained because:

(a) The set $\{S_5, S_4, S_8, L_3\}$ has gross error cardinality three. Thus, a $(+2)$ value in $S_8$, a $(+1.5)$ value in $S_4$ and a $(+1)$ gross error in $L_3$ is equivalent to a $(-0.5)$ bias in $S_4$, a $(-1)$ value in $L_3$ and a $(-2)$ value is $S_5$.
(b) In turn, the set $\{S_1, S_6, L_3\}$ has gross error cardinality two. Thus a $(-1)$ value in $L_3$ and a $(+1)$ gross error in $S_1$ are equivalent to a $(+1)$ bias in $S_6$.

The first is a regular equivalency of basic sets whereas the second is a degenerate case, which is only possible because of the particular value of $L_3$. This case, although highly unlikely, illustrates the effect of degen-

eracy on exact gross error detection. Words of caution must therefore be given against such claims.

Consider now the set of 5 gross errors introduced in Table 20. Three gross errors are identified by the serial strategy. Figure 11 shows the gross errors introduced and the streams that will contribute to form loops. One can immediately anticipate that a basic subset representing these errors will contain not more than four gross errors, as it is evident from the fact that $S_8$ and $S_9$, which are part of the original gross errors, form a set of gross error cardinality one, i.e. the streams with gross errors (solid lines) form a loop. Thus, the gross errors introduced are a basic subset of the system $\Lambda = \{S_1, S_4, S_6, S_8, L_1, L_3, L_4, S_5, S_9\}$. Consider now the following: the set $\{S_4, S_5, S_8, S_9, L_3\}$ has the following basic subset $\{S_4, S_5, L_3\}$. Thus the four gross errors in streams $S_4, S_8, S_9$ and $L_3$ can be represented by the following gross errors in the basic subset: $S_4(-2), S_5(-4), L_3(-1)$. Now consider the set $\{S_1, S_6, L_3\}$. The set of gross errors in this set, i.e. $S_1$
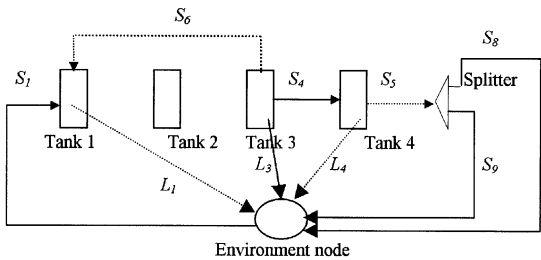
Fig. 11. Loops formed by the gross errors introduced in Table 20.

$(+1)$, $L_3$ $(-1)$, which is equivalent to the following gross error: $S_6$ $(+1)$. This was obtained using Corollary 4, that is, adding the value of the leak $L_3$, to all the other streams in the loop.

A reduction from the anticipated four gross errors to only three, is due to the fact that the original gross errors have such special values that allow the degenerate case in loop $\{S_1, S_6, L_3\}$. We again stress out that these cases of degeneracy are unlikely to show often in practice.

*Maximum cardinality.* Tables 21 and 22 illustrate the introduction of six and seven gross errors. As anticipated, not more than 5 gross errors can be identified.

## Conclusions

This paper has presented a strategy for the simultaneous identification and estimation of gross errors in process plants. Three kinds of gross errors, i.e. flowrate biases, holdup biases and tank leaks are considered and included in the integral model. The equivalencies among different gross errors are discussed analytically and a gross error detection and estimation strategy has been presented. The strategy for identification and estimation of gross errors performs with efficiency even in the presence of a high number of errors. Gross error exact detection has been discussed, and it has been established that this is possible in very few cases. Future work includes a serial strategy to detect first order biases and leaks in possible combination with zero-order biases and leaks. For this case the dynamic integral measurement test has to be modified, and a strategy to detect the order of a bias has to be developed. In addition, future work will include an assessment of the power of the technique, in the presence of gross errors of different sizes.

## Acknowledgements

## Nomenclature

| | |
|---|---|
| $A$ | system matrix, equation (1) |
| $B$ | system matrix, equation (5) |
| $C$ | system matrix, equation (2) |
| $D$ | system matrix, equation (4) |
| DIMT | Dynamic Integral Measurement Test |
| DIVMT | DIMT for tank holdup bias |
| $E$ | matrix defined by equations (12), (29) and (39) |
| $G$ | matrix defined by equations (43) and (45) |
| $f$ | vector of mass flowrates |
| $h$ | matrix defined by equations (50) and (51) |
| $I$ | identity matrix |
| $J, D_m, R_m, C_\alpha$ | matrices defined by Bagajewicz and Jiang (1997) |
| $LC$ | vector for gross error candidates |
| $LCGE$ | vector for confirmed gross errors |
| $M$ | matrix defined by equation (17) |
| $o$ | matrix defined by equation (69) |
| $O$ | matrix defined by equation (69) |
| $P$ | matrix defined by equations (15) and (32) |
| $q$ | vector defined by equations (15) and (33) |
| $s$ | polynomial order |
| $S$ | sample standard deviation |
| $S_F$ | variance matrix of flowrates |
| $SF$ | flowrate bias polynomial order |
| $SL$ | leak polynomial order |
| $S_V$ | variance matrix of holdups |
| $SV$ | tank holdup polynomial order |
| $t$ | variable |
| $U$ | matrix defined by equations (15) and (34) |
| $v$ | vector of tank holdups |
| $x$ | vector defined by equation (17) |
| $\tilde{x}$ | reconciled value without assuming any gross errors |
| $y$ | a statistics defined by equation (61) |
| $z$ | deviations of the measurements from the reconciled values |

*Greek letters*

| | |
|---|---|
| $\alpha$ | coefficients in flowrate polynomial |
| $\omega$ | coefficients in holdup polynomial |
| $\eta$ | significant level |
| $\xi$ | variable |
| $\delta$ | bias vector |
| $\hat{\delta}$ | bias vector without zero elements |
| $\rho$ | coefficients in flowrate bias polynomial |
| $\mu$ | coefficients in tank holdup bias polynomial |
| $\tau$ | vector defined by equation (34) |
| $\lambda$ | coefficients in leak polynomial |
| $\psi$ | matrix defined by equation (12) |
| $\Lambda$ | vector of gross errors |
| $\Psi$ | vector of gross errors |
| $\Omega$ | vector of gross errors |
| $\Gamma$ | gross error cardinality |
| $\Pi$ | variable defined by equation (65) |
| $\Delta$ | variable |

*Subscripts or superscripts*

| | |
|---|---|
| $F$ | flow rate related quantities |
| $k$ | variable |
| $l$ | leak related quantities |
| $R$ | redundant quantities |

| | |
|---|---|
| $V$ | holdup related quantities |
| $+$ | measured quantities |
| $*$ | optimal value |
| $0$ | at time $t = 0$ |

## References

Albuquerque, J.S. and Biegler, L.T. (1996) Data reconciliation and gross-error detection for dynamic systems. *A.I.Ch.E. J.* **42**, 2841.

Bagajewicz, M. and Jiang, Q. (1997) An integral approach to dynamic data reconciliation. *A.I.Ch.E. J.* **43**, 2546.

Crowe, C.M., Hrymak, A. and Garcia Campos, Y.A. (1983) Reconciliation of process flow rates by matrix projection. I. The linear case. *A.I.Ch.E. J.* **29**, 881.

DATACON Manual (1993) Simulation Sciences.

Dunia, R. *et al.* (1996) Identification of faulty sensors using principal component analysis. *A.I.Ch.E. J.* **42**, 2797.

Iordache, C., Mah, R.S.H., and Tamhane, A.C. (1985) Performance studies of the measurement test for detection of gross errors in Process data. *A.I.Ch.E. J.* **31**, 1187.

Keller, J.Y., Darouach, M. and Krzakala, G. (1994) Fault detection of multiple biases or process leaks in linear steady state systems. *Comput Chem. Engng* **18**(10), 1001.

Kim, I., Kang, M.S., Park, S. and Edgar, T. (1997) Robust data reconciliation and gross error detection: the modified MIMT using NLP. *Comput Chem. Engng* **21**(7), 775.

Madron, F. (1985) A new approach to the identification of gross errors in chemical engineering measurements. *Chem. Engng Sci.* **40**, 1855.

Madron, F. (1992) *Process Plant Performance. Measurement and Data Processing for Optimization and Retrofits*. Ellis Horwood Series in Chemical Engineering, Chichester, England.

Mah, R.S.H., Stanley, G.M. and Downing, D.M. (1976) Reconciliation and Rectification of Process Flow and Inventory Data. *Ind. Engng Chem. Process Des. Dev.*, **15**, 175.

Mah, R.S.H. and Tamhane, A.C. (1982) Detection of gross errors in process data. *A.I.Ch.E. J.* **28**, 828.

Narasimhan, S. and Mah, R.S.H. (1987) Generalized likelihood ratio method for gross error detection. *A.I.Ch.E. J.* **33**, 1514.

Narasimhan, S. and Mah, R.S.H. (1988) Generalized likelihood ratios for gross error identification in dynamic processes. *A.I.Ch.E. J.* **33**, 1514.

Reilly, P.M. and Carpani, R.E. (1963) Application of statistical theory of adjustment to material balances. *13th Can. Chem. Engng Conf.*, Montreal, Que.

Rollins, D.K. and Davis, J.F. (1992) Unbiased estimation of gross errors in process measurements. *A.I.Ch.E. J.* **38**, 563.

Rollins, D.K. and Devanathan, S. (1993) Unbiased estimation in dynamic data reconciliation. *A.I.Ch.E. J.* **39**, 1330.

Sanchez, M. (1996) Process Monitoring: instrumentation analysis and plant data reconciliation. Ph.D. Thesis, South University, Argentina.

Serth, R. and Heenan, W. (1986) Gross error detection and data reconciliation in steam metering systems. *A.I.Ch.E. J.* **32** 733.

Tong, H. and Crowe C.M. (1994) The application of principal component analysis to tests for gross errors in data reconciliation. *PSE '94*, Kyongju, Korea.