### AN INTEGRAL APPROACH TO LINEAR DYNAMIC DATA RECONCILIATION

Hsiao-Ping Huang
Department of Chemical Engineering, National
Taiwan University, Taipei, Taiwan, R.O.C
<a href="mailto:huanghpc@ntu.edu.tw">huanghpc@ntu.edu.tw</a>

Kuo-Yuan Luo Department of Chemical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C d91524010@ntu.edu.tw

#### **Abstract**

:Steady-state data reconciliation theories have been well developed and are easy to implement. In the contrast, dynamic data reconciliation techniques are not so easy to cope with. Here, we propose a new dynamic data reconciliation method by integration the differential-algebraic equations, which transforms the differential-algebraic equations into algebraic constraints, combined with filtering the measurement signals beforehand. It is easy to cope with compared with the other proposed methods.

Keywords: data processing, integral equation formulation

#### 1. INTRODUCTION

Through the use of data reconciliation techniques, the corruption of process variables due to measurement noise can be reduced and the reconciliation provides the more correct process data information which is useful to improve the understanding of the process and the control performance etc. The general methodology can be divided into three main steps: 1. Classification of process variables and problem decomposition. 2. Detection, identification, and estimation of gross errors. 3. Measurement adjustment and estimation of the unmeasured process variables. Therefore some issues are associated with a general reconciliation problem including process classification, gross error detection, identification, and estimation etc. Formally, data reconciliation can be defined as the estimation of measurement process data variables to reduce measurement error through the use of the temporal and functional redundancies. Mathematically, data reconciliation is the optimal estimation to a constrained least-squares or maximum likelihood objective function. Reviews of these methods and strategies have been proposed and are introduced in the published books (Mah, 1990; Madron, 1992; Romagnoli and Sanchez, 2000; Narasimhan and Jordache, 2000). Dynamic data

reconciliation have been attracted more attentions due to the facts that in the real processes there are always variations of the process variables. Theories of steady-state data reconciliation have been well developed, whereas theories of dynamic systems are fairly nascent and still needs to evolve. The formulation of the dynamic data reconciliation can be written as the following equations:

As mentioned above, the general data reconciliation problem must satisfy minimizing an object function, i.e. Eq. (1), where  $\hat{y}_{(t)}$  are estimations, y are measurements,  $\sigma$  is measurement noise standard deviation.

$$\min \varphi[y, \hat{y}; \sigma] \tag{1}$$

The objective function subjects to Eq. (2), Eq. (3), and Eq. (4) where  $g_1$  is the differential equation constraint,  $g_2$  is the algebraic equality constraint, and  $g_3$  is the inequality constraint.

$$g_1 \left[ \frac{d\hat{\mathbf{y}}(t)}{dt}, \hat{\mathbf{y}}(t) \right] = 0 \tag{2}$$

$$\mathbf{g}_{2}\left[\hat{\mathbf{y}}(t)\right]=0\tag{3}$$

$$g_{3}[\hat{y}(t)] \ge 0 \tag{4}$$

Methods like Kalman filter estimation, nonlinear programming, integral approach..., etc. are adopted to cope with the dynamic problems.

The Kalman filter estimation techniques are under the broad umbrella of dynamic data reconciliation. Under the processing of Kalman filter techniques, a discrete dynamic system model is needed. The drawbacks of the Kalman filter techniques are that it cannot deal with the inequality constraints and a precise process model is needed. It will result in bad performances due to an inaccurate process model. In dealing with the nonlinear dynamic systems linearization which leads to inevitable model errors is needed. Dynamic data reconciliation by the use of nonlinear programming techniques is another widely used estimation method. The general formulation of the nonlinear dynamic data reconciliation problem comes with prices. The computation burden is the main drawback of this kind of methods, that's why these techniques have not been applied to industrial processes. An integral approach method has been proposed (Bagajewicz and Jiang, 1997). By the use of polynomial approximation of the measurement signals, it can obtain pretty smooth results compared with most of the dynamic data reconciliation methods. But it still is limited to linear system dynamics and the approximating results may be not so good if the measurement signals are more disturbed.

Although it may be improved the performance by increasing the polynomial order, but determining the order of the polynomial still needs to try out and with increasing order there will be a lot parameters and huge matrices for reconciliation whereas increasing the order doesn't assure increasing the performances. Here, we propose an integral approach combined with filtering for dynamic systems. By integrating the differential-algebraic equations, which actually means material balances during the integrating interval, we transform the DAE to algebraic constraints. Then, the dynamic data reconciliation problem can be solved readily by steady-state data reconciliation theories. Due to the well de-noising ability of the wavelet transformation, we treat the measurement signals by the discrete wavelet transformation before the data reconciliation is underway.

#### 2. WAVELET FILTERING

The discrete wavelet transformation (DWT) is adopted to de-noise the measurement signals before the reconciliation. The DWT can eliminate the abnormal data in measurements and utilize the temporally redundant information of measurements to reduce the measurement errors. DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low2004 程序系統工程研討會論文集 pass and high-pass filters respectively, perform like A filters. time-scale (frequency) representation of a signal is obtained using digital filtering techniques combined with up-sampling and down-sampling (sub-sampling) operations. By the Fourier transformation analysis of the wavelet function digital filter, it has a band-pass like spectrum as shown in Fig.1. Due to the design of the wavelet functions, the highest amplitude of the filter is equal to  $\sqrt{2}$ . Signal frequencies in the range of amplitude equal to  $\sqrt{2}$  all pass and those in the range less than  $\sqrt{2}$  pass partly through the filter. It is customarily setting frequency at the 0.707 of the highest amplitude as cut-off frequency, i.e. f<sub>cw</sub>, of the high-pass filter. Different band-pass range comes from the sub-sampling operations and it can filter at the different frequencies desired. In Fig. 1, the cutoff frequencies of the high-pass filter at different levels of the wavelet transformation are represented by  $f_{cw1}$ ,  $f_{cw2}$ ,  $f_{cw3}$ ,..., etc,  $f_s$  means the sampling frequency. The exact value of fcw must be determined by sampling frequency  $f_{\text{s}}$  and the relations of each  $f_{\text{cw}}$ and f<sub>s</sub> are listed in Table 1. In the other words, we adopt DWT filtering the measurement signals at different frequency levels to de-noise measurement signals. Then we need to determine which level of wavelet decomposition shall we adopt.

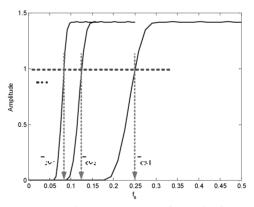


Fig. 1. Band-pass spectrum of wavelet function.

Table 1 Cut-off frequencies of different wavelet decomposition level

Wavelet decomposition level	Cut-off frequency $(f_{cw})$
1	0.25*f <sub>s</sub>
2	$0.125*f_{s}$
3	$0.0625*f_{s}$
:	:
n	$0.25*f_s*(0.5)^{n-1}$

We determine the level of wavelet decomposition by the use of the dynamic characteristics of the dynamic system. It is known that the process itself is like a filter and high frequency signals can be more or less vanished. The main of the filter behaviour comes from the dynamic characteristic time constant, τ. From the bode plot we can know how a signal of

different frequencies can effect the process. According to the bode plot of a system with time constant equal to  $\tau$  shown in Fig. 2, the corner frequency is equal to  $1/\tau$  if it is an FOPDT dynamic system. Then we set the cut-off frequency equal to c times of the corner frequency. The value of c can be determined according to the existed frequencies in the dynamic system. The cut-off frequency, f<sub>c</sub>, can be calculated by Eq. (5). The signals in the processes are mostly low frequent, then we can regard the frequencies exceeding the cut-off frequency as high frequency noises in the measurement signals. After obtaining the cut-off frequency, we can determine which level of wavelet decomposition must be reached and filter the measurement signals before the reconciliation. We will illustrate the filtering examples in the last part in this paper.

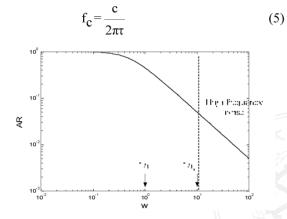


Fig. 2. Bode plot of a system with time constant  $\tau$ .

### 3. INTEGRAL-APPROACH RECONCILIATION

The dynamics of a material balance in a process plant can be rep-resented by the following differential-algebraic equations (DAE):

$$\frac{dh}{dt} = Af \tag{6}$$

$$Cf=0 (7)$$

In Eq. (6) and Eq. (7), h is the collection of variables associated with the derivative term, f is the collection of variables of the non-derivative term and A, C are constant matrices from the algebraic part of the DAE associated with the dynamic characteristic. Integrating Eq. (6) and Eq. (7) between some time  $t_1$  and  $t_2$ , we can get the following algebraic Eq. (8) and Eq. (9).

$$\int_{t_0}^{t_n} \frac{dh}{dt} dt = A \int_{t_0}^{t_n} f dt$$
(8)

$$\begin{array}{c}
t_n \\
C \int_{t_0}^{t_n} f dt = 0
\end{array} \tag{9}$$

Let  $Z_1 = \int_{t_0}^{t_n} \frac{dh}{dt} dt$ ,  $Z_2 = \int_{t_0}^{t_n} f dt$ , and we can rearrange the integrations and get the following matrix form in Eq. (10).

$$\begin{bmatrix} A & -I \\ C & 0 \end{bmatrix} \begin{bmatrix} Z_2 \\ Z_1 \end{bmatrix} = 0 \tag{10}$$

We will see that the result is an algebraic constraint and the following reconciliation procedure is to deal with the integrating by Simpson's rule.

The Simpson's n+1 (n is even) points rule is shown in Eq. (11).

$$\int_{0}^{t_{n}} f(t)dt = \frac{s}{3} \left( f_{0} + 4f_{1} + 2f_{2} + 4f_{3} + \dots + 4f_{n-1} + f_{n} \right), s = \frac{t_{n} - t_{0}}{n}$$
 (11)

Define new variable H and F which represent the collections of all measurements of all instruments during the integrating time interval  $t_0$  to  $t_n$ . Assuming there are k variables (k instruments) in h and m variables (m instruments) in f. H and F is shown in Eq. (12).

$$H = \begin{bmatrix} h_{1,t_0} \\ \vdots \\ h_{1,t_n} \\ \vdots \\ h_{k,t_0} \\ \vdots \\ h_{k,t_n} \end{bmatrix}, F = \begin{bmatrix} f_{1,t_0} \\ \vdots \\ f_{1,t_n} \\ \vdots \\ f_{m,t_0} \\ \vdots \\ f_{m,t_n} \end{bmatrix}$$

$$(12)$$

Then, the integration of Eq. (8) and Eq. (9) can be represented as

Eq. (13) and Eq. (14).

$$Z_1 = Q_1 * F$$
 (13)

$$Z_2 = Q_2 * H$$
 (14)

 $Q_1$  and  $Q_2$  are matrices shown as Eq. (15) and Eq. (16) and s is the sampling time interval.

$$Q_{1} = \frac{s}{3} \begin{bmatrix} & \frac{n+1}{1 & 4 & 2 & \cdots & 4 & 1} \\ & & & & \ddots & & \\ & & & & & 1 & 4 & 2 & \cdots & 4 & 1 \end{bmatrix}$$
 (15)

Finally, we can obtain the algebraic constraint represented by H and F.

$$\begin{bmatrix} A & -I \\ C & 0 \end{bmatrix} \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} \begin{bmatrix} F \\ H \end{bmatrix} = 0$$
 (17)

Let's recall the general formulation of the steady-state data reconciliation problem. There is usually a model to describe the measurement variables and usually the distribution of  $\varepsilon$  is normal distribution with a process specific standard deviation. The measurement model is shown in Eq. (18).

$$y=x+\varepsilon$$
 (18)

y are measurement variables and x are true value of measurement variables.

And there must be constraints to construct the reconciliation problem. Assuming the constraint matrix is A. It will satisfy Eq. (19).

$$Ax=0$$
 (119)

Then, the data reconciliation problem is to estimate x which satisfy the constraints and maximize the likelihood estimation problem or minimize the following objective function in Eq. (20).

$$\min_{X} (y-x)^{T} \Sigma^{-1} (y-x)$$
 (20)

 $\Sigma$  is covariance matrix of the measurement variables.

The solution of the maximum likelihood estimation problem can be obtained using the method of Lagrange multipliers and is shown in Eq. (21).

$$\hat{\mathbf{x}} = \mathbf{y} - \sum \mathbf{A}^{\mathrm{T}} (\mathbf{A} \sum \mathbf{A}^{\mathrm{T}})^{-1} \mathbf{A} \mathbf{y}$$
 (21)

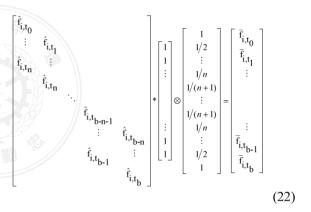
With Eq. (21), we can get the reconciled values of the measurements at each time interval if we have known the measurement signals y, constraint in Eq. (17) and  $\Sigma$ .

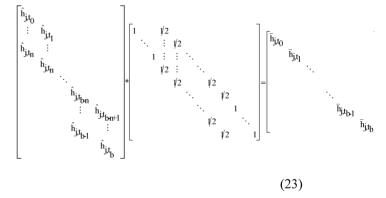
In the theory, we use a batch-like wavelet filtering and a moving- window integral data reconciliation in each batch collection. The length of the batch measurement numbers must be long enough to adopt the valuable temporal redundancies. The length of the integrating doesn't need so long to obtain good erformance.

Throughout this procedure, it can adopt the temporal and spatial redundancies of the measurement signals. As the moving-windows goes on, we will get repeated reconciled variables at any time instant. We average the repeated reconciled variables as the final reconciled variables at every time instant. The average procedure can be illustrated as follows:

The length of the moving window is equal to the length of integration points. Assuming there are n+1 integration sampling points (from  $t_0$  to  $t_n$ ) and b sampling points in a batch of wavelet filtering (b>n+1).

We save the variables after each reconciliation as the moving window moves on. It will execute b-n reconciliation procedures in each batch. Reconciled values will be repeated after the repeated reconciliation procedures. Finally, we must average the repeated reconciled values. The averaging must be executed in two different ways according to F and H. For F, the averaging can be done by Eq. (22). The repeated reconciled values are stored like the first matrix from the left in Eq. (22). After multiplying a column vector of all elements equal to 1 it changes to a column vector equal to summation of all repeated reconciled variables at each time instant. Finally, multiplying the third matrix from the left gets the average values of the repeated reconciled variables. For H, the average is performed in Eq. (23) whereas the repeated reconciled values occur at the first and final points.





# 3. GROSS-ERROR DETECTION

Gross errors are briefly classified in three categories: 1. true outliers, 2. process leaks, 3. biased instrumentation. Outliers have been defined as measurement values depart from the expected distribution interval of the values. And it is usually adopted the normal distributions for the measurement values. Leaks, which are referred to some

unexpected leaks in process equipments, lead to a unbalanced mass conservation in the process. Biased instrumentation is typically a constant value added to the measurement values due to miscalibration or a constant drift in the measurement values.

Here, we adopt the gross-error strategy proposed by Bagajewicz and Jiang in 1997. Consider the measurement adjustments for instrument i at time j shown in Eq. (24).

$$W_{ij} = z_{ij} - \hat{z}_{ij} \tag{24}$$

 $z_{ij} \ is \ the \ measurement \ for \ instrument \ i \ at \ time \ j, \ \hat{z}_{ij} \ is \\ the \ corresponding \ reconciled \ value. \ W_{ij} \ should \\ follow \ a \ normal \ distribution \ with \ zero \ mean.$ 

$$E\left(W_{ij}\right) = 0 \tag{25}$$

The sample deviation for  $W_{ij}$  can be calculated by Eq. (26)

$$S = \sqrt{\frac{1}{n} \sum_{j=0}^{n} \left(W_{ij} - \bar{W}_{j}\right)^{2}}$$
 (26)

 $\overline{W}_j$  is the mean of all  $W_{ij}$ . Thus, the following variable,  $R_k$ , follows a t-student distribution.

$$R_k = \frac{\bar{W}_k}{S/\sqrt{n+1}} \tag{27}$$

Variable j will be suspected of containing a gross error if the following Eq. (28) holds.

$$R_k > t_{1-(\eta/2)}$$
 (28)

 $\eta$  is usually selected as 0.05 (95% confidence level). The critical value of  $t_{1-(\eta/2)}$  at  $\eta$  equal to 0.05 is 2.01.

## 4. Example

Here, we illustrate a four-tank system show in Fig. as an example.

There are two main flows  $f_5$ ,  $f_6$  split into two branches apiece. The four branches,  $f_1$ ,  $f_2$ ,  $f_3$ ,  $f_4$ , flow into four tanks respectively. Each tank has flow out of it. The flow out of tank 3 is fed into tank 1 and the one out of tank 4 is fed into tank 2.

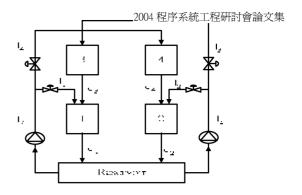


Fig. 3. A four tanks system

The DAE of this example is showed in Eq. (29). Parameters of the process are listed in Table 2.

$$A_{1} \frac{dh_{1}}{dt} = -a_{1} \sqrt{2gh_{1}} + a_{3} \sqrt{2gh_{3}} + f_{1}$$

$$A_{2} \frac{dh_{2}}{dt} = -a_{2} \sqrt{2gh_{2}} + a_{4} \sqrt{2gh_{4}} + f_{2}$$

$$A_{3} \frac{dh_{3}}{dt} = -a_{3} \sqrt{2gh_{3}} + f_{3}$$

$$A_{4} \frac{dh_{4}}{dt} = -a_{4} \sqrt{2gh_{4}} + f_{4}$$
(29)

By linearization at nominal values we can obtain the approximated time constant values of each tank. The results are showed in Table 2. According to the known time constants and setting c equal to 10, we can obtain the cut-off frequencies by Eq. (5), and the results are listed in Table 3.

Table 2 Tank parameters

Symbol	State/Parameters	Value	Dimension
$h_0$	Nominal levels	[20.4; 20.4; 11.5; 11.5]	cm
$a_i$	Area of the drain	[3; 3; 2;2]	cm <sup>2</sup>
$A_i$	Areas of the tanks	1000	$cm^2$
$f_i$	flow into the tank	[0.3; 0.3; 0.3;0.3]	cm <sup>3</sup> /sec
$T_i$	Time constants	[68; 68; 76.5; 76.5]	sec
g	Gravitation constant	1981	cm/sec <sup>2</sup>
$\sigma_{ m f}$	Standard deviation of flow	0.015	cm <sup>3</sup> /sec
$\sigma_{\rm h}$	Standard deviation of level	0.6	cm

Assuming we measure all flows and levels, then the DAE becomes a linear system problem shown in Eq. (30). The standard deviations of the measurement variables are listed in Table 2.

$$A_{1} \frac{dh_{1}}{dt} = -q_{1} + q_{3} + f_{1}$$

$$A_{2} \frac{dh_{2}}{dt} = -q_{2} + q_{4} + f_{2}$$

$$A_{3} \frac{dh_{3}}{dt} = -q_{3} + f_{3}$$

$$A_{4} \frac{dh_{4}}{dt} = -q_{4} + f_{4}$$
(30)

Table 3 Cut-off frequencies from the time constants

Dynamics	Time	Cut-off
	constant	frequency(Hz)
Tank 1	68	0.0234
Tank 2	68	0.0234
Tank 3	76.5	0.0208
Tank 4	76.5	0.0208

<u>Table 4 Cut-off frequencies of different wavelete</u>
<u>decomposition level of sampling time equal</u>
<u>to 2 second</u>

Wavelet	Cut-off frequency	
decomposition level	$(f_{cw})$ (Hz)	
1	0.125	
2	0.0625	
3	0.03125	
4	0.015625	
5	0.0078125	

Since the cut-off frequencies are known, we must filter the measurement signals frequencies higher than the values listed in Table 3. From Table 4, we can know that the level of the wavelet filtering must be chosen to 4-th level in order to filter frequencies exceeding 0.0208 Hz or 0.0234 Hz.

# 4.1 Example 1

Assuming there are not any gross-errors, and all the flows and levels are measured. The batch time interval of the wavelet filtering is chosen to 105 points and the moving window time interval is chosen to 21 points. We compare this method with the other reconciliation methods, Kalman filter estimation and an integral approach proposed by Bagajewicz and Jiang in 1997. The minimum square errors of these methods listed in Table 5 are calculated in order to compare the reconciliation results.

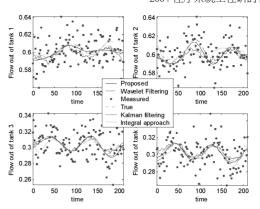


Fig. 4. Reconciled results of flows out of tanks

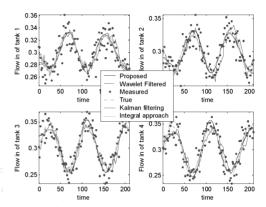


Fig. 5. Reconciled results of flows into the tanks

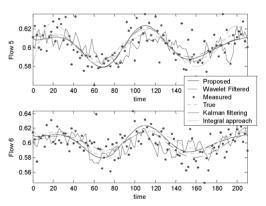


Fig. 6. Reconciled results of flow 5 and flow 6

Table 5 Minimum square errors of three methods

M.S.E. of	Proposed	Kalman	Integral
measurement		filter	approach
$f_1$	0.00234	0.00474	0.00222
$f_2$	0.00273	0.00842	0.00180
$f_3$	0.00505	0.01199	0.00217
$f_4$	0.00625	0.01211	0.00326
$f_5$	0.00198	0.00859	0.00320
$f_6$	0.00154	0.01229	0.00229
$q_1$	0.00127	0.00113	0.00094
$q_2$	0.00149	0.00142	0.00175
$q_3$	0.00107	0.00042	0.00196
$q_4$	0.00054	0.00036	0.00178
$h_1$	0.912	5.231	1.824
$h_2$	1.159	6.517	2.632
$h_3$	1.685	2.482	2.811
$h_4$	0.660	2.146	1.293

Table 7 Gross error detection with process leak in tank 3

From Table 5, we can see the minimum square errors of the measurement signals. Most of our reconciliation results of the measurement signals are superior to the others. The order of the polynomial is chosen to 8 in the integral approach method proposed by Bagajewicz and Jiang. The operation of the Kalman filter estimation is under the condition without model error and algorithms are introduced briefly in the Appendix.

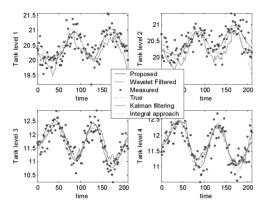


Fig. 4. Reconciled results of tank levels

## 4.2 Example 2

In example 2, we illustrate the gross error detection under the condition of measurement bias and process leak. Assuming there is a constant bias of 3 times value of the standard deviation in  $f_1$  at time equal to 140. Using the strategy mentioned in section 3, the gross error is detected successfully shown in Table 6. We can notice that if there is a gross error in one of the measurements, the gross error will be smeared into the other associated measurements by the reconciliation procedure. If there is a bias in  $f_1$ , we can find that both  $f_5$  and  $q_1$ exceed the detection criteria 2.01. But we can point out that the gross error comes from the measurement which has the biggest M.S.E. value. From Table 6, we can point out that  $f_1$  is with gross error. The same results are found in the gross error detection of process leak. Giving a process leak in tank 3, f<sub>2</sub>, f<sub>3</sub>, f<sub>6</sub>, q<sub>1</sub>, q<sub>3</sub> will exceed the detection criteria whereas tank 3 is the most suspicious.

 $\frac{\text{Table 6 Gross error detection with measurement bias}}{\text{in } f_1}$ 

Measurement	$R_k$	Measurement	$R_k$
$f_1$	4.007	$q_2$	0.847
$f_2$	1.033	$q_3$	1.791
$f_3$	1.546	$q_4$	0.332
$f_4$	1.708	$h_1$	0.220
$f_5$	2.832	$h_2$	0.006
$f_6$	0.038	$h_3$	0.236
$q_1$	2.912	$h_4$	0.262

Measurement  $R_k$ Measurement  $R_k$  $f_1$ 1.149 0.701  $q_2$  $f_2$ 2.023 3.606  $q_3$  $f_3$ 3.442 0.468  $q_4$  $f_4$ 0.500  $h_1$ 0.138  $f_5$ 1.199 0.222  $h_2$  $f_6$ 2.772 0.108  $h_3$ 2.6968 0.121  $h_4$  $q_1$ 

#### 5. CONCLUSIONS

This article has presented an integral approach to cope with dynamic data reconciliation. Before the reconciliation, a filtering procedure based on DWT is involved to de-noise the measurement signals. The integration is expanded by Simpson's rule and the repeated reconciled values are averaged. Through the wavelet filtering by larger sampling points in a batch we during the procedure utilize temporal redundancies of the measurement signals and by the moving window integral reconciliation in a batch we utilize functional redundancies of the system. The results are comparable to Bagajewicz and Jiang's integral approach and are superior to the Kalman filter estimation. The gross error detection works according to the detection theory proposed by Bagajewicz and Jiang in 1997.

## **REFERENCES**

Mah, R.S.H. (1990). *Chemical process structures* and information flows, Butterworths, Stoneham, MA.

Madron, F (1992). Process plant performance-Measurement and data processing for optimization and retrofits, Ellis Horwood, Chichester, UK.

Miguel J. Bagajewicz and Qiyou Jiang (1997). Integral approach to plant dynamic reconciliation. *AIChE Journal*, **vol. 43**, No. 10, pp. 2546-2558.

Jose A. Romagnoli and Mabel Cristna Sanchez (2000). *Data processing and reconciliation for chemical process operations*, Academic Press, New York.

M. J. Liebman, T. F. Edgar and L. S. Lasdon (1992). Efficient data reconliliation for dynamic processes using nonlinear programming thehniques. *Computes Chem. Engng.*, vol. 16, No. 10/11, pp. 963-986.

Shankar Narasimhan and Cornelius Jordache (2000).

Data reconciliation & gross error detection-An intelligent use of process data, Gulf Publishing Company, Houston, Texas