Chinese J. Chem. Eng., **14**(5) 592—596 (2006)

An NT-MT Combined Method for Gross Error Detection and Data Reconciliation*

MEI Congli(梅从立)[a,b], **SU Hongye**(苏宏业)[a,**] **and CHU Jian**(褚健)[a]
[a] National Laboratory of Industry Technology, Institute of Advanced Process Control, Zhejiang University, Hangzhou 310027, China
[b] Jiangsu University, Zhenjiang 212013, China

Abstract An NT-MT combined method based on nodal test (NT) and measurement test (MT) is developed for gross error detection and data reconciliation for industrial application. The NT-MT combined method makes use of both NT and MT tests and this combination helps to overcome the defects in the respective methods. It also avoids any artificial manipulation and eliminates the huge combinatorial problem that is created in the combined method based on the nodal test in the case of more than one gross error for a large process system. Serial compensation strategy is also used to avoid the decrease of the coefficient matrix rank during the computation of the proposed method. Simulation results show that the proposed method is very effective and possesses good performance.

Keywords data reconciliation, gross error detection, measurement test, nodal test

1 INTRODUCTION

In industrial process, instrument readings do not satisfy the laws of conversation and one has to perform data reconciliation to obtain variable estimates. Unfortunately, measured process variables often systematically deviate from their true values. Miscalibrated and malfunctioning instruments are two reasons for biased measurements which are called gross errors. If the measurements are adjusted to satisfy the laws of conversation in the presence of gross error, then all the adjustments are greatly affected by such biases and would not generally be reliable indicators of the state for the process. So gross errors must be detected and either rectified or discarded before data reconciliation.

Statistical test is a useful method to detect gross errors. The most widely used methods are the global test (GT)[1], the measurement test (MT)[2], the nodal test (NT)[1,3], the generalized likelihood ratio (GLR)[4], the principal component test (PCT)[5], and the maximum power test (MP)[6]; among them, three kinds of strategies have been developed to identify and rectify multiple gross errors serial elimination, serial compensation, simultaneous or collective compensation.

To improve the efficiency of gross errors detection in industrial process, a combined method containing several statistical test methods is a novel strategy. A MT-NT combined method was proposed to identify multiple gross errors in industrial process[7]. This method combined the MT and NT together and let them compensate each other. Wang[8] improved the method by using serial compensation method to avoid the decrease of the coefficient matrix rank when estimating the variables with gross errors. But the "MT location and NT check" method still has two drawbacks: first, if there are two gross errors of the approximate magnitude linked to the same node, the NT method cannot identify the gross errors correctly. So the NT method cannot be used as a criterion to check the presence of gross errors. Second, there is no provision to prevent the gross error from affecting the whole data in the least square procedure. Since the result of least square estimation is used in r_{Mj}, the relative adjustment $I_{Mj}=\left|r_{Mj}/X_j\right|$[7] used in the MT-NT method doesnot indicate the reality of biased measurements and hence cannot be used as a criterion to identify the stream with gross error correctly.

The results of this paper are based on the idea of the combined method. To overcome the drawbacks of the MT-NT method, a novel combined method, the NT-MT combined method, is presented. The NT method is designed to search the set of imbalance nodes and the MT method is used to search the set of suspicious streams with gross errors. Then an equal weighted least square procedure is used to identify the stream with gross error correctly. The serial compensation method is also used to avoid the decrease of the coefficient matrix in estimating the instrument readings with gross errors.

2 PRINCIPLES OF DATA RECONCILIATION

2.1 Problem statement

In the absence of gross errors, data reconciliation is the procedure of optimally adjusting measurements such that the adjusted values to satisfy the laws of

---

Received 2005-09-20, accepted 2006-01-13.
* Supported by the National Creative Research Groups Science Foundation of China (No.60421002) and the National "Tenth Five-Year" Science and Technology Research Program of China (2004BA204B08).
** To whom correspondence should be addressed. E-mail: hysu@iipc.zju.edu.cn

conversation and other constraints.

The basic model of process measurements data reconciliation is

$$\boldsymbol{X} = \boldsymbol{X}^{(0)} + \boldsymbol{\varepsilon} \quad (1)$$

$$\min[(\boldsymbol{X}^{(1)} - \boldsymbol{X})^{\mathrm{T}} \boldsymbol{Q}^{-1} (\boldsymbol{X}^{(1)} - \boldsymbol{X})] \quad (2)$$

$$\text{s.t. } \boldsymbol{A}\boldsymbol{X}^{(1)} + \boldsymbol{B}\boldsymbol{U} + \boldsymbol{C} = 0 \quad (3)$$

where $\boldsymbol{X}$ and $\boldsymbol{X}^{(0)}$ are vectors of process measurements and true values respectively. The vector $\boldsymbol{\varepsilon}$ represents random error. $\boldsymbol{X}^{(1)}$ is reconciled value vector and $\boldsymbol{U}$ is vector with unmeasured variables or measurements deleted from the set of measured variables. $\boldsymbol{Q}$ is a diagonal covariance matrix which is assumed to be known or estimated. $\boldsymbol{A}$ and $\boldsymbol{B}$ are coefficient matrixes of the balance equations respectively. $\boldsymbol{C}$ is a vector of constants. $\boldsymbol{X}^{(1)}$ and $\boldsymbol{U}$ can be solved from Eqs.(4) and (5)

$$\boldsymbol{U} = -[\boldsymbol{B}^{\mathrm{T}} (\boldsymbol{AQA}^{\mathrm{T}}) \boldsymbol{B}]^{-1} \boldsymbol{B}^{\mathrm{T}} (\boldsymbol{AQA})^{-1} (\boldsymbol{AX} + \boldsymbol{C}) \quad (4)$$

$$\boldsymbol{X}^{(1)} = \boldsymbol{X} - \boldsymbol{QA}^{\mathrm{T}} (\boldsymbol{AQA})^{-1} (\boldsymbol{AX} + \boldsymbol{BU} + \boldsymbol{C}) \quad (5)$$

where, Eqs.(4) and (5) are obtained on the assumption that $\boldsymbol{AQA}^{\mathrm{T}}$ and $\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{AQA}^{\mathrm{T}})\boldsymbol{B}$ are reversible. This condition can always be satisfied by matrix projection[9] , whether $\boldsymbol{AQA}^{\mathrm{T}}$ and $\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{AQA}^{\mathrm{T}})\boldsymbol{B}$ are reversible or not.

2.2 Measurement test (MT)

The measurement test method suggests a statistic criterion $\boldsymbol{Z}_{\mathrm{M}j}$ based on $\boldsymbol{r}_{\mathrm{M}}$ and $\boldsymbol{W}$,

$$Z_{\mathrm{M}j} = \frac{r_{\mathrm{M}j}}{\sqrt{W_{jj}}} \quad (6)$$

where

$$\boldsymbol{r}_{\mathrm{M}} = \boldsymbol{X}^{(1)} - \boldsymbol{X} = \boldsymbol{QA}^{\mathrm{T}} (\boldsymbol{AQA}^{\mathrm{T}})^{-1} \boldsymbol{AX} \quad (7)$$

$$\boldsymbol{W} = \boldsymbol{QA}^{\mathrm{T}} (\boldsymbol{AQA}^{\mathrm{T}})^{-1} \boldsymbol{AQ} \quad (8)$$

and $\boldsymbol{r}_{\mathrm{M}}$ is the vector of residual errors, $\boldsymbol{W}$ is the covariance matrix of $\boldsymbol{r}_{\mathrm{M}}$.

Assuming that the errors are of normal distribution, the measurements with gross errors can be detected by comparing $Z_{\mathrm{M}j}$ with the critical value Z_{Mc}. This method has two drawbacks: first, the least squares procedure tends to spread the gross errors over the data. It tends to incur type Ⅰ error. Second, there is no provision to prevent unrealistic result from being computed, permitting negative flow rates to be generated.

2.3 Nodal test (NT)

The nodal test investigates the residuals of balance equations; statistical criterion $Z_{\mathrm{N}i}$ based on residuals are suggested. The residuals are

$$\boldsymbol{r}_{\mathrm{N}} = \boldsymbol{AX} + \boldsymbol{C} \quad (9)$$

The covariance of $\boldsymbol{r}_{\mathrm{N}}$ is

$$\boldsymbol{J} = \mathrm{Cov}(\boldsymbol{r}_{\mathrm{N}}) = \boldsymbol{AQA}^{\mathrm{T}} \quad (10)$$

The statistic criterion at node i is

$$Z_{\mathrm{N}i} = \frac{r_{\mathrm{N}i}}{\sqrt{J_{ii}}} \quad (11)$$

At 95% confidence level, α=0.05[10,11], there is a critical value Z_{Nc}. If $|Z_{\mathrm{N}i}| > |Z_{\mathrm{Nc}}|$, there can be some errors. The drawbacks of this method are: first, if there are two gross errors of the approximate magnitude, they could be offset by each other and cause the detection to fail; second, it does not specify the stream data possessing the gross error. It results in type Ⅱ error.

2.4 MT-NT combined method

The MT-NT combined method was presented by Yang and Teng[7]. The idea of this method is to: (1) find out the streams with potential gross errors by MT; (2) check the nodes to detect the “bad” stream by NT; (3) remove the bad stream from the measurements and treat it as an unmeasured variable, and start the iteration until there are no gross errors. This “MT location and NT check” method uses the advantages of both methods and avoids the huge combinatorial search problem.

When the method is applied, the column of the coefficient matrix is changed as the measurements with gross errors are removed, while the row is not changed correspondingly. As the column is decreased, the rank of the coefficient matrix may decrease. It will result in the interruption of the computation and the test may be stopped. To continue the computation, nodes must be combined to eliminate the useless nodes. This will increase operation and some useful information will be lost.

To avoid the decrease of the rank of coefficient matrix, the serial compensation strategy was introduced[8]. In this improved method, when the stream with gross error is identified, it is replaced by estimation. So the coefficient matrix is unchanged.

The improved method does not change the MT-NT combined method in nature. There are two main drawbacks: first, a stream is identified as a bad stream if and only if $Z_{\mathrm{N}i}$ of two nodes linked to the bad stream are both larger than the critical value. If there are two gross errors of the approximate magnitude linked to the same node, then the NT method cannot identify the gross errors correctly. So the NT method cannot be used as a criterion to check gross errors. Second, there is no provision to prevent the data from being affected by spread gross error in the least square procedure. The relative adjustment $I_{\mathrm{M}j}=|r_{\mathrm{M}j} / X_j|$ is used as a criterion to identify the stream with gross error, which has the same drawback of MT; as a result the least square estimation is used to compute $r_{\mathrm{M}j}$.

3 NT-MT COMBINED METHOD

Generally speaking, the NT method cannot be a perfect criterion to check the absence of the stream with gross error linked to a node, but it can be a criterion to judge the presence of gross error. And the MT method cannot identify the streams with gross error, as the gross error spreads to the whole data in the least square procedure. However, if Z_{Mj} of a stream is less than the critical value, the stream can be measured correctly.

To combine the advantages of the NT method and the MT method, let us change their position in the MT-NT combined method. Let the NT method locate the "bad nodes" linked the streams with gross error and the MT method check the streams linked to the "bad nodes" and detect the "bad streams". To avoid spreading of gross errors over correct data, an equal weighted least square procedure is applied. The serial compensation method is also used to avoid the decrease of the coefficient matrix rank in iteration procedure.

The detailed algorithm can be described as follows.

Step 1. Compute Z_{Mj} of all streams and Z_{Nj} of all nodes, let $\boldsymbol{L}$=$\boldsymbol{Q}$;

Step 2. Compare Z_{Nj} with the critical value Z_{Nc}, if $|Z_{Nj}|>|Z_{Nc}|$ denotes the node as a "bad node" and computed into the set $\boldsymbol{T}$; compare Z_{Mj} with the critical value Z_{Mc}, if $|Z_{Mj}|>|Z_{Mc}|$, denotes the stream as a "bad stream" and computed into the set $\boldsymbol{S}$; if $\boldsymbol{T}$ is empty, turn to step 6;

Step 3. Let $\boldsymbol{Q}$=$\boldsymbol{E}$ (identity matrix), compute Z_{Nj} of all nodes in $\boldsymbol{T}$, and denote the nodes with the largest Z_{Ni} as Z_{Nmax}; compute Z_{Mj} of all streams in $\boldsymbol{S}$ linked to the node Z_{Nmax}; find out the stream with the largest Z_{Mj} linked to Z_{Nmax} and denote it as Z_{Mmax};

Step 4. Estimate the Z_{Mmax} using Eq.(4) and replace the measurement of Z_{Mmax} with the estimated value; compute the measurements in $\boldsymbol{P}$ (a set used to place rectified measurements and $\boldsymbol{P}=\varnothing$ in the first iteration procedure), respectively, using Eq.(4) and replace the measurements with corresponding estimations;

Step 5. Place the stream Z_{Mmax} into P; let $\boldsymbol{Q}$=$\boldsymbol{L}$; turn to step 1;

Step 6. Reconciliate the data using Eq.(5); output the result.

4 CASE STUDY

4.1 Case 1

A schematic diagram of recycle process network is described as Fig.1[4,12]. This case shows the drawbacks of the MT-NT method and the high performance of the presented NT-MT method.

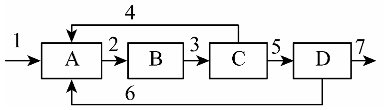

Figure 1 Recycle process network

In Table 1, streams 1 and 2 contain the approximate gross error. After using NT for the first time, only Z_{N2} (Z_{N2}=3.51) is larger than the critical value (Z_{Nc}=2.49). However, two adjacent imbalance nodes are needed to identify a stream with gross error by the MT-NT method; hence this cannot identify any stream with gross error and would result in dead cycle seriously. The column $\boldsymbol{X}_1^{(1)}$ and relative error show that no result can be obtained by the MT-NT method. By using the NT-MT combined method, node 2 (Z_{N2}=3.51) is first found as a bad node after using NT first. Then stream 2 is identified as a bad stream, for which estimate is equal to 15.12. During the iteration procedure, node 1 (Z_{N1}=3.19) is found as a bad node. Then stream 1 is found as a bad stream and the value of its estimate is equal to 5.09. The second estimate of stream 2 is equal to 15.02. The column $\boldsymbol{X}_2^{(1)}$ and relative error of Table 1 give the results of data reconciliation by the NT-MT method.

In Table 2, by using the MT-NT combined method, stream 5 (Z_{M5}=4.37) is first judged as a bad stream and the value of its estimate equals 11.44. During the second iteration, stream 2 is judged as a bad stream and the value of its estimate is equal to 16.67. In the third iteration, no stream is judged as bad stream and the result of reconciliation is according to the column $\boldsymbol{X}_1^{(1)}$ of Table 2. By using the NT-MT combined method, stream 3 is first judged as bad stream and the value of its estimate is equal to 15.45. In the second iteration, stream 6 is judged as a bad stream and the value of its estimate is equal to 5.23. The second estimate of stream 3 is equal to 15.24. In the third iterative computation, no stream is judged as a bad stream and the result of data reconciliation is shown in Table 2. Obviously, Table 2 shows that using $I_{Mj}=|r_{Mj}/X_j|$ cannot be a correct criterion to identify

Table 1 Result of data reconciliation when streams 1 and 2 with gross error

No.	$\boldsymbol{X}^{(0)}$	$\boldsymbol{X}$	Relative error, %	$\boldsymbol{X}_1^{(1)}$	Relative error, %	$\boldsymbol{X}_2^{(1)}$	Relative error, %
1	5	6.32	26.40	—	—	5.10	2.00
2	15	16.71	11.43	—	—	15.18	1.20
3	15	14.85	−1.00	—	—	15.18	1.20
4	5	5.09	1.72	—	—	5.03	0.60
5	10	10.20	2.04	—	—	10.15	1.50
6	5	5.09	1.78	—	—	5.05	1.00
7	5	5.16	3.23	—	—	5.10	2.00

Table 2 Result of data reconciliation when streams 1 and 6 with gross error

No.	$X^{(0)}$	X	Relative error, %	$X_1^{(1)}$	Relative error, %	$X_2^{(1)}$	Relative error, %
1	5	4.82	−3.60	4.98	−0.31	4.98	−0.40
2	15	15.21	1.43	16.67	11.15	15.28	1.87
3	15	16.85	12.33	16.67	11.15	15.28	1.87
4	5	5.09	1.72	5.11	2.19	5.08	1.60
5	10	10.20	2.04	11.56	15.63	10.20	2.00
6	5	6.59	31.78	6.58	31.57	5.22	4.40
7	5	5.16	3.23	4.98	−0.31	4.98	−0.40

Table 3 Comparison of the performance

No.	$X^{(0)}$	X	Relative error, %	$X_1^{(1)}$	Relative error, %	$X_2^{(1)}$	Relative error, %
1	296.6	289.94	−2.25	294.32	−0.77	293.61	−1.01
2	50.6	85.41	68.79	51.93	2.63	51.73	2.24
3	117.2	113.86	−2.85	118.33	0.96	118.36	0.99
4	115.3	120.81	4.78	116.25	0.82	116.47	1.01
5	246	241.81	−1.7	242.4	−1.46	241.87	−1.68
6	244.8	248.29	1.47	248.15	1.37	247.67	1.17
7	243	245.53	1.04	246.39	1.4	245.88	1.19
8	1.8	1.79	−0.56	1.79	−0.56	1.79	−0.56
9	4.7	4.55	−3.19	4.55	−3.19	4.55	−3.19
10	1.9	1.89	−0.53	1.89	−0.53	1.89	−0.53
11	84.9	83.09	−2.13	83.14	−2.07	83.46	−1.70
12	70.2	71.3	1.57	70.99	1.13	71.07	1.24
13	132	67.9	−48.56	132.28	0.21	130.42	−1.20
14	202.1	203.55	0.72	202.36	0.13	202.80	0.35
15	200	202.19	1.1	199.38	−0.31	199.93	−0.08
16	188	90.07	−52.09	187.06	−0.5	187.59	−0.22
17	12	12.31	2.58	12.32	2.67	12.34	2.83
18	102.6	104.31	1.67	103.64	1.01	102.00	−0.58
19	35.7	35.52	−0.5	35.48	−0.62	35.29	−1.15
20	5.7	5.48	−3.86	5.48	−3.86	5.47	−4.04
21	264.6	274.32	3.67	267.81	1.21	267.70	1.17
22	37.1	37.49	1.05	37.32	0.59	37.33	0.62
23	307.4	298.31	−2.96	310.61	1.04	310.50	1.01
24	8.7	8.33	−4.25	8.33	−4.25	8.33	−4.25
25	192	193.42	0.74	192.19	0.1	192.42	0.22

the stream with gross error.

4.2 Case 2

This is a steam power network diagram of methanol synthesis set (Fig.2)[7]. The column $X_1^{(1)}$ and the column $X_2^{(1)}$ of Table 3 are obtained by the MT-NT method and NT-MT method respectively. Table 3 shows the efficiency of the presented NT-MT method.

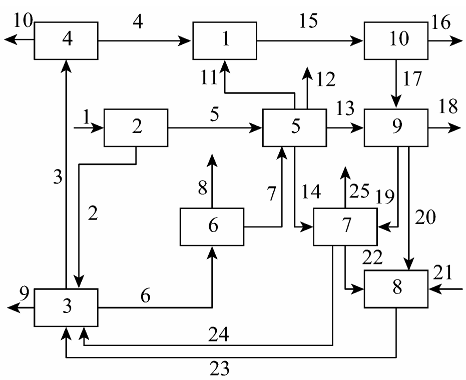

Figure 2 Steam metering processing system

5 CONCLUSIONS

An NT-MT combined method is presented for multiple gross error detection and data reconciliation in this paper. The strategy of this method is to detect gross error and reconcile it *via* successive iteration procedure. This method can make the NT and MT tests compensate each other. In order to avoid gross error spreading over measurements and affecting the identification of bad streams, the equal weighed least square procedure is introduced. The results of simulation show the effectiveness and reliability of the proposed methods.

NOMENCLATURE

$\boldsymbol{A}$, $\boldsymbol{B}$	coefficient matrix
$\boldsymbol{C}$	vector of constants
$\mathrm{Cov}(\boldsymbol{r}_{\mathrm{N}})$	covariance matrix of $\boldsymbol{r}_{\mathrm{N}}$
$\boldsymbol{E}$	identity matrix
I	relative adjustment
$\boldsymbol{L}$, $\boldsymbol{Q}$	diagonal covariance matrices
$\boldsymbol{P}$, $\boldsymbol{S}$, $\boldsymbol{T}$	sets
$\boldsymbol{r}$	residual error
$\boldsymbol{U}$	unmeasured variables
$\boldsymbol{W}$	covariance matrix
$\boldsymbol{X}$	process measurements
$\boldsymbol{X}_1$	results of MT-NT method
$\boldsymbol{X}_2$	results of NT–MT method
z	statistic criterion
$\boldsymbol{\alpha}$	probability of Type Ⅰ error
$\boldsymbol{\varepsilon}$	random error

Superscripts

(0)	true value
(1)	reconciled value

Subscripts

c	critical value
i	number of nodes
j	number of measurements
M	measurement test
N	nodal test

REFERENCES

1 Relly, P., Carpani, R., "Application if statistical theory of adjustments to material balances", In: Proceedings of the 13th Canadian Chemical Engineering Conference, Montreal Qebec (1963).

2 Mah, R.S.H., Tamhane, A.C., "Detection of gross errors in process data" , *AIChE J*., **28**, 828—830(1982).

3 Mah, R.S.H., Staneley, G., Downing, D., "Reconciliation and rectification of process flow and inventory data", *Ind*. *Eng*. *Chem*. *Process Des*. *Dev*., **15**, 175—183(1976).

4 Narasimhan, S., Mah, R.S.H., "Generalized likelihood ratio method for gross error identification", *AIChE J*., **33**, 1514—1521(1987).

5 Tong, H., Crowe, C. M., "Detection of gross errors in reconciliation by principal component analysis", *AIChE J*., **41**(7), 1712—1722(1995).

6 Crowe, C. M., "The maximum-power test for gross errors in original constraints in data reconciliation", *Can*. *J*. *Chem*. *Eng*., **70**(10), 1030—1036(1992).

7 Yang, Y., Teng, R., "Gross error detection and data reconciliation in process industrial", *J*. *Chem*. *Ind*. *Eng*., **47**(2), 248—253(1996). (in Chinese)

8 Wang, F., Jia, X., Zheng, S., Yue, J., "An improved MT-NT method for gross error detection and data reconciliation.", *Comput*. *Chem*. *Eng*., **28**, 2189—2192(2004).

9 Crowe, C.M, Garcia Campos, Y.A., Hrymak. A., "Reconciliation of process flow rates by matrix projection (Ⅰ) linear case", *AIChE J*., **29**(6), 881—888(1983).

10 Serth, R., Hennan, W., "Gross error detection and data reconciliation in steam metering system", *AIChE J*., **32**, 733—742(1986).

11 Kim, I., Kang, M., Park, S., Edgar, T., "Robust data reconciliation and gross error detection: the modified MIMT using NLP", *Comput*. *Chem*. *Eng*., **21**(7), 775—782(1997).

12 Sanchez, M., Romagnoli, J., Jiang, Q., Bagajewicz, M., "Simultaneous estimation of biases and leak in process plants ", *Comput*. *Chem*. *Eng*., **23**, 841—857(1999).