Korean J. of Chem. Eng., **13**(2), 159-164 (1996)

STAGE EFFICIENCY ESTIMATION BY MODIFIED MIMT USING NLP

Mun Sik Kang*, In-Won Kim† and Sunwon Park**

*Doosan Technical Center, Yongin, Korea
Department of Chemical Engineering, Kon Kuk University, Seoul, Korea
**Department of Chemical Engineering, KAIST, Taejon, Korea
(Received 16 September 1995 • accepted 3 January 1996)

Abstract—Unmeasured process variables or parameters caused by cost consideration or technical infeasibility can be mostly estimated using data reconciliation techniques. Since, however, the gross errors possibly present in the process measurements deteriorate the data reconciliation results, the reconciled estimates may be biased solutions that are different from the true values. In this paper, the enhanced data reconciliation and gross error detection method, modified MIMT using NLP, was applied to a flash distillation system. It calculated the reconciled values of the measurements as well as the optimal estimates of stage efficiencies which were not measured. These techniques using NLP showed the robustness when compared to the conventional algorithms using linearization techniques.

Key words: Data Reconciliation, Gross Error Detection

INTRODUCTION

Accurate process models are required for the optimization and control in chemical plants and petrochemical manufacturing facilities. Theses models involve various equipment parameters, such as stage efficiencies in distillation columns, the values of which must be determined by fitting the models to process data. But process measurements contain random errors or possibly gross errors. Random errors are normally distributed with zero means and known covariance matrix and gross errors may result from sources such as unsuspected leak, miss calibration of the measurement device, and malfunctioning sensors. Since these inconsistent data therefore do not satisfy the physical constraints of the process, such as material and energy balances, the reliability of the data is greatly reduced. In the case of estimating the stage efficiency and parameters of measurement using these inconsistent measured data, the computed estimates are biased which are different from true values. The problem thus involves parameter estimation coupled with gross error detection and data reconciliation.

Until now the above problems are solved via the method based on linearization techniques to compute the optimal estimates of unmeasured variables and parameters in processes. The performance of these linearization techniques is considerably reduced as the nonlinearity of model and the number of gross error in the measurement data are increased. But when the enhanced data reconciliation and gross error detection by modified MIMT using NLP was applied to a CSTR system, the performance of the enhanced algorithm was superior to the method using linearization techniques [Kim et al., 1995].

Thus, in this work, the enhanced data reconciliation and gross error detection algorithm using NLP is applied to estimate the stage efficiency of a flash distillation system which was considered by Serth et al. [1993], and the performance of the proposed algorithm is compared to that of the conventional method.

†To whom all correspondences should be addressed.

PROBLEM STATEMENTS

In the past 30 years, the data reconciliation and gross error detection of steady state processes has received considerable attention in chemical engineering literature [Terry and Himmelblau, 1993]. A number of methods for detecting and identifying gross errors in linearly constrained data have been developed, most of which involve the use of statistical tests based on the assumption that the random errors in the data are normally distributed. In one of the simplest methods, the set of residuals from the least-squares procedure is tested for outliers, and any measurement for which the corresponding residual fails the test is considered to contain a gross error. This data reconciliation and gross error detection algorithm has been advocated by several investigators including Ripps [1965], Hogg and Tanis [1977], Knepper and Gorman [1980], Iordache et al. [1985], Tamhane and Mah [1985], Crowe [1986], Serth and Heenan [1986], Rosenberg et al. [1987], and Kao et al. [1990], and its performances have been studied on a number of problems by Iordache et al. [1985] and Serth and Heenan [1986].

In the case of not all the variables are measured, the objective function for data reconciliation can be written as

$$\operatorname*{Min}_{z,\theta} \frac{1}{2}(\mathbf{z}_m-\mathbf{z})^T\mathbf{R}^{-1}(\mathbf{z}_m-\mathbf{z}) \tag{1}$$
$$\text{s.t. } \mathbf{f}(\mathbf{z},\ \theta)=0$$

where θ represents the estimates of unmeasured variables or parameters. The solutions of this objective function give the reconciled data, **z**, satisfying the process model and the optimal estimates of unmeasured variables, θ. The above data reconciliation problem can be solved by linearization techniques and nonlinear programming techniques. The solution technique using NLP has the advantage that it explicitly handles nonlinear constraints and specifies the upper and lower bounds on the optimal solution. Therefore this technique can compute the robust optimal solution regardless of the nonlinearity of the process and the number of gross errors in the measurement [Kim et al., 1995].

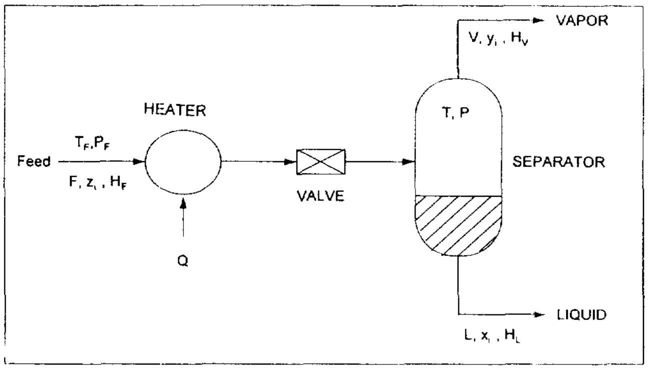

Fig. 1. Flash distillation system.

In the data reconciliation described above, it is assumed that no gross errors were present in the measurements. When the gross errors are present in the measurements, the gross error detection must be preceded before the data reconciliation and the suspect measurement variables are identified and corrected. In the MIMT, the gross error detection algorithm proposed by Serth and Heenan [1986], at each stage the residuals are tested for outliers, and the measurement corresponding to the most significant residuals is deleted from the set of measured variables and estimated. The iterations are terminated when all remaining residuals satisfy the test for outliers. In a comparative study of a number of gross error detection algorithms, this method was found to represent the best combination of robust and effectiveness. In this work the modified MIMT algorithm using NLP, proposed by Kim et al. [1995], was used for the gross error detection.

In order to estimate the suspected variable including the gross error, we can formulate another objective function similar to Eq. (1).

$$\mathbf{Min}_{z,w,\theta}\frac{1}{2}(\mathbf{z}_m'-\mathbf{z}')^T\mathbf{R}'^{-1}(\mathbf{z}_m'-\mathbf{z}') \quad (2)$$
$$\text{s.t. } \mathbf{f}(\mathbf{z}', \mathbf{w}, \theta)=0$$

where $\mathbf{z}_m'$ are the measurement set in which a suspected measurement is deleted, and $\mathbf{w}$ is the estimate of the suspected variable in the measurements. The objective function given in Eq. (2) can be solved by the methods based on linearization techniques and nonlinear programming techniques, which are same as the data reconciliation. When we incorporated the nonlinear programming techniques into the gross error detection algorithm and applied to a CSTR system, this enhanced algorithm could give the robust solution regardless of the nonlinearity of model and the number of gross errors in the measurement [Kim et al., 1995].

SIMULATION EXAMPLE

1. Flash Distillation System

A non-adiabatic, non-equilibrium single-stage flash system considered by Serth et al. [1993] is shown in Fig. 1. For a feed containing C components, the mesh equations for the system are listed in Table 1 in terms of the vaporization efficiency. Though many efficiency equations can be written using other definitions of stage efficiency, two fundamental models, Vaporization efficiency and Modified Murphree efficiency (based on mole fraction),

Table 1. Mesh equation for a flash distillation using vaporization efficiency

Type	Equation	Number
Material Balance	$Fz_i-Lx_i-Vy_i=0$	C
Efficiency	$y_i-\theta_i^V K_i x_i=0$	C
Sum of Mole Fractions	$\Sigma_i x_i-1.0=0$	3
	$\Sigma_i y_i-1.0=0$	
	$\Sigma_i z_i-1.0=0$	
Enthalpy Balance	$F+Q-H_L-H_vV=0$	1
Total		2C+4

Table 2. Specification for example problem

Variable	Value
F	0.454 kmol/s
z_1	0.15
z_2	0.35
z_3	0.30
T_F	316.7 K
P_F	3,447.4 kPa
P	1,723.7 kPa
Q	2,108.4 kJ/s

are considered and compared in this work.

$$\text{Vaporization Efficiency}: \theta_i^V=\mathbf{y}_i/\mathbf{K}_i\mathbf{x}_i \quad (4)$$

$$\text{Modified Murphree}: \theta_i^{MM}=\frac{y_i-x_i}{K_ix_i-x_i} \quad (5)$$

For simulation purposes, a feed containing four components [(1) ethane, (2) propane, (3) propylene, and (4) isobutene] was selected. Specifications for the example problem are given in Table 2. For given values of the component efficiencies, θ_i, the values of the remaining process variables [T, L, V, x_i, y_i (i=1,2,3,4) and z_4] were determined implicitly by the mesh equations. The thermodynamic relations given by Holland [1981] were used for enthalpies and K values. It should be noted that these K values are independent of composition.

2. Simulation Procedure

The performance of data reconciliation and gross error detection algorithms were tested via 100 computer simulation runs. For each simulation run, a measurement vector was constructed as

$$\mathbf{z}_m=\mathbf{x}+\varepsilon+\delta \quad (6)$$

where $\mathbf{x}$ is the original value, ε is the vector of random measurement errors, and δ is the vector of systematic errors. The true values of the process variables were obtained by solving the mesh equations subject to the constraints given in Table 2 and specified values of component efficiencies. For simplicity, all component efficiencies were assumed equal, so that a single-state efficiency characterized the flash. The calculations were performed for an efficiency of 75% only. A Gaussian pseudo-random number generator was first used to generate ε. For the purpose of these experiments, the relative standard deviation of temperature was taken to be 0.4% for temperatures, and 2.5% for flow rates, mole fractions and heat flow. Random errors were assumed to be statistically independent so that all covariance terms were zero. After generation of the random error vector, a uniform pseudo random number generator was used to define the number, position, magnitude and algebraic sign of the non-zero systematic errors. The number of non-zero systematic errors was allowed to vary between one and three, while the range of systematic errors magnitudes (as a percentage of true values) used was 2% to 10% for temperatures and 10% to 100% for other variables. Pressure measurements were assumed to be exact since pressure effects are small in this system [Serth et al., 1993].

For the purpose of application of the MIMT algorithm, the critical test value z_c was computed as follow. For $\alpha=0.05$ ($z_{\alpha/2}=1.96$), we have $\beta=0.0028$, and $z_{1-\beta/2}=2.98$. For the gross error detection algorithm, the lower bounds on the variables are set to 0.05 times the true value for stage efficiency and zero for other variables. The upper bounds on the variables were set at 3.0 times the true value for stage efficiency, and 4.0 times the other parameter values. For nonlinear programming, the lower and upper bounds of the optimal solution were set at 65% to 85% for the unmeasured stage efficiency and 0.8 and 1.2 times the corresponding true value for the other variables.

3. Performance Evaluation

The performance of each algorithm was tested by the percentage reduction in total rms error in the data computed as follows [Serth et al., 1987];

$$\%\ \text{Total Error Reduction}=\frac{E_1-E_2}{E_1}\times 100 \tag{7}$$

$$E_1=\sqrt{\sum_{j=1}^{18}(z_{mj}-x_j)_s^2}$$

$$E_2=\sqrt{\sum_{j=1}^{18}(\hat{x}_j^*-x_j)_s^2}$$

In these equations, E_1 and E_2 are initial and final rms (root mean square) errors; $\mathbf{z}_m$, $\mathbf{x}$, and $\hat{\mathbf{x}}^*$ are the vectors of measured values, true values, and final reconciled values, respectively; and the subscript 's' indicates that scaled values of the variables are employed.

RESULTS AND DISCUSSION

For the tests, the random and gross errors were added to 18 variables of the flash system by Eq. (6), and the 100 measurement data sets were obtained by simulation. The same seed was used in the random number generator in order to give the same measurement set of test cases. The true value of the stage efficiency was set at 75%.

In the discussion, the abbreviated names of the algorithms are used: data reconciliation techniques using the linearization technique (DR/SL) and the nonlinear programming technique (DR/NLP); gross error detection algorithms using the linearization technique (GED/SL) and using the nonlinear programming technique (GED/NLP).

The results for the four methods are summarized in Table 3 for The modified Murphree efficiency model as a function of the number of gross errors. The data reconciliation results of both DR/SL and DR/NLP, shown in the lower part of Table 3, are similar when the measurements were not corrupted by gross errors. However the performance differences of the two data reconciliation techniques become wider as the number of gross errors was increased. When the data were corrupted by three gross errors, the average value of stage efficiencies, estimated by DR/SL, is 70.14%, which is considerably different from the true value. The maximum and minimum values of stage efficiency were 121.1 % and 0% respectively, which is inconsistent with real processes. For DR/NLP, however, the maximum and minimum values of the stage efficiency were 85% and 65% respectively, which are the specified inequality constraints in NLP.

Table 3. Performance results for modified Murphree efficiency model with one-sided systematic errors

	Gross Error Detection							
Number of	GED/SL				GED/NLP			
gross errors	0	1	2	3	0	1	2	3
Error reduction %	38.3	71.8	74.9	61.7	38.2	76.1	72.7	77.4
Efficiency %								
Mean	75.1	74.8	74.3	72.9	75.1	74.9	75.2	75.0
STD	2.3	2.8	6.5	11.6	2.3	2.6	2.9	3.7
Max	82.7	82.7	83.9	113.5	82.7	82.7	83.4	84.2
Min	70.1	61.2	34.3	0	70.1	65.6	66.1	65.0

	Data Reconciliation							
Number of	DR/SL				DR/NLP			
gross errors	0	1	2	3	0	1	2	3
Efficiency %								
Mean	75.0	74.9	74.6	70.1	75.0	74.8	75.2	74.3
STD	2.2	9.2	14.5	20.9	2.2	5.4	7.0	7.7
Max	79.8	100.3	109.1	121.1	79.8	85.0	85.0	85.0
Min	70.1	43.6	28.9	0	70.1	65.0	65.0	65.0

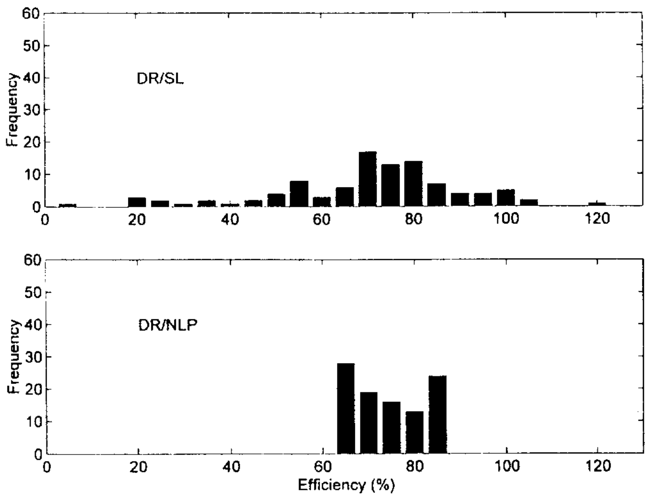

Fig. 2. Histograms of the stage efficiency estimates by DR/SL and DR/NLP for modified Murphree efficiency model.

The histograms for 100 stage efficiency estimates by the two data reconciliation methods with three gross errors are shown in Fig. 2. The estimates by DR/SL are scattered widely and there exists several biased estimates. While the estimates by DR/NLP are scattered between the specified bounds of 85% and 65%, most estimates are located near the bounds. Therefore the optimal estimates of the unmeasured variable (stage efficiency) cannot be computed correctly by the data reconciliation techniques when the data are corrupted with many gross errors. Hence the gross error detection step must be performed before data reconciliation, and the suspected measurement variables with gross errors should be identified and corrected.

The results from the gross error detection algorithms are presented in the upper rows of Table 3. Similar to the data reconcil-

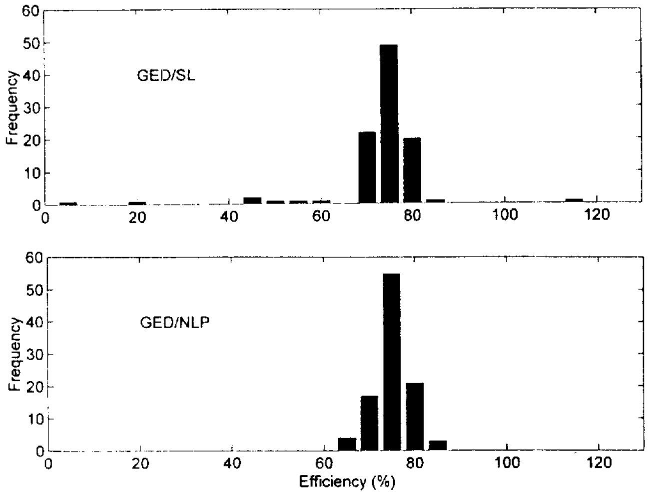

Fig. 3. Histograms of the stage efficiency estimates by GED/SL and GED/NLP for modified Murphree efficiency model.

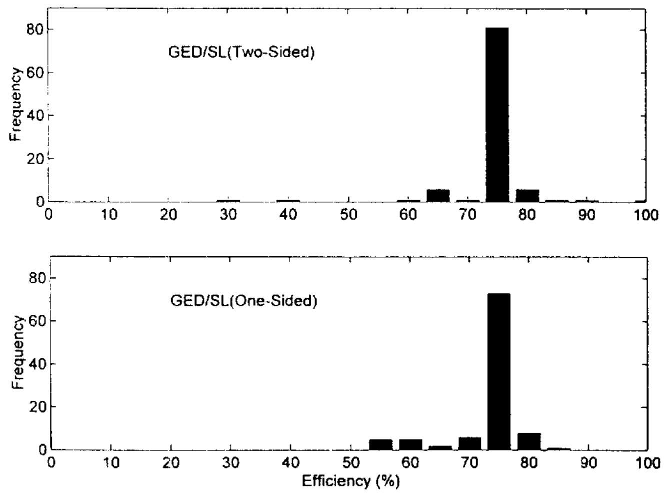

Fig. 4. Histograms of the stage efficiency estimates by GED/SL for different systematic error types.

Table 4. Performance results for vaporization efficiency model with two-sided systematic errors

	Gross Error Detection							
Number of	GED/SL				GED/NLP			
gross errors	0	1	2	3	0	1	2	3
Error reduction %	39.6	76.9	77.5	67.7	39.6	79.0	83.4	81.0
Efficiency %								
Mean	75.1	75.3	74.7	74.2	75.1	75.2	74.9	75.0
STD	1.2	1.9	2.9	7.1	1.2	1.6	2.2	2.9
Max	77.7	85.8	83.6	99.2	77.7	82.4	84.5	83.4
Min	70.8	70.8	53.3	31.2	70.8	70.8	67.2	65.4

	Data Reconciliation							
Number of	DR/SL				DR/NLP			
gross errors	0	1	2	3	0	1	2	3
Efficiency %								
Mean	75.0	75.9	76.2	75.3	75.0	75.3	75.1	75.3
STD	1.2	7.7	9.6	8.8	1.2	3.6	4.9	5.1
Max	77.7	118.1	118.6	117.8	77.7	85.0	85.0	85.0
Min	70.8	55.9	55.8	43.4	70.8	65.0	65.0	65.0

Table 5. Performance results for vaporization efficiency model with one-sided systematic errors

	Gross Error Detection							
Number of	GED/SL				GED/NLP			
gross errors	0	1	2	3	0	1	2	3
Error reduction %	39.6	74.9	73.0	58.6	39.6	75.1	77.2	77.5
Efficiency %								
Mean	75.1	75.3	74.7	73.3	75.1	75.1	75.2	74.8
STD	1.2	1.3	4.3	5.9	1.2	1.3	2.5	3.6
Max	77.7	78.4	90.6	82.9	77.7	78.4	85.0	84.5
Min	70.8	70.8	42.6	54.6	70.8	70.8	65.0	65.0

	Data Reconciliation							
Number of	DR/SL				DR/NLP			
gross errors	0	1	2	3	0	1	2	3
Efficiency %								
Mean	75.0	74.4	73.8	70.6	75.0	74.8	74.5	72.4
STD	1.2	4.3	6.7	8.0	1.2	3.1	4.6	5.6
Max	77.7	81.3	89.7	86.6	77.7	81.4	85.0	85.0
Min	70.8	52.4	52.9	51.0	70.8	65.0	65.0	65.0

iation results, the performances of GED/SL, such as power, average error reduction percentage, and maximum and minimum efficiencies, deteriorate as the number of gross errors is increased from one to three. In the case of GED/NLP, however, even though the measurements were corrupted by three gross errors, the power of the algorithm is still 0.92. The value of average stage efficiency is 74.97%, which is close to the true value of stage efficiency, and the standard deviation of the efficiency distribution is 3.7%, which is narrower than that for GED/SL (11.6%).

The histograms of the stage efficiencies computed by the gross error detection algorithms for three gross errors are shown in Fig. 3. The outliers shown in the histogram of GED/SL indicates that the gross errors in the measurements cannot be correctly identified and estimated by GED/SL. For GED/NLP, however, most of the suspected variables are corrected, resulting in an optimal estimate of stage efficiency which is close to the true value.

In order to investigate the robustness of the data reconciliation and gross error detection algorithms to systematic errors, the data were corrupted by one-sided and two-sided systematic errors in all variables. The gross errors were only subtracted from the random data for the one-sided data, and were subtracted and added randomly for the two-sided data. The results for the two-sided systematic errors are presented in Table 4 and those for the one-sided systematic errors are summarized in Table 5. In general, the results for the two-sided errors are better than those with the one-sided errors, based on the averaged values shown in the tables. In the case of GED/NLP, the estimates with one-sided errors are not much worse than those with two-sided errors. The histograms of the estimates by GED/SL for the different types of data are different as shown in Fig. 4. In contrast the histograms obtained by GED/NLP, shown in Fig. 5, have similar distributions, with most estimates around the true value of stage efficiency, 75%.

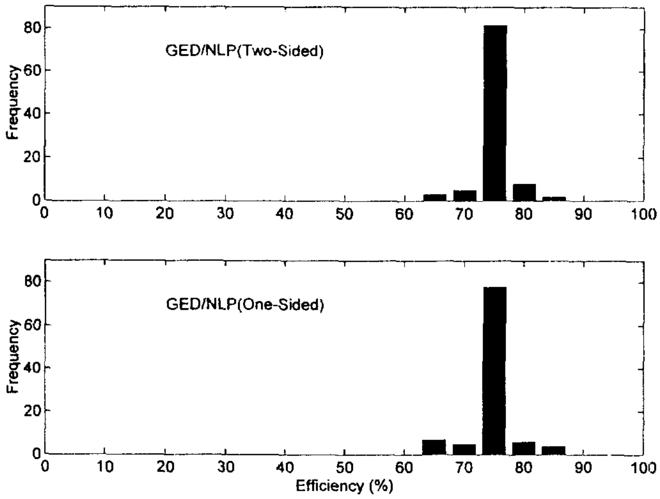

Fig. 5. Histograms of the stage efficiency estimates by GED/NLP for different systematic error types.

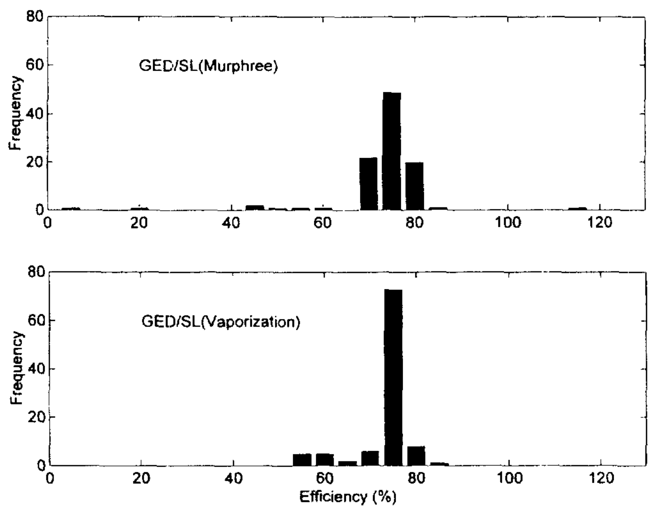

Fig. 6. Histograms of the stage efficiency estimates by GED/SL for different efficiency models.

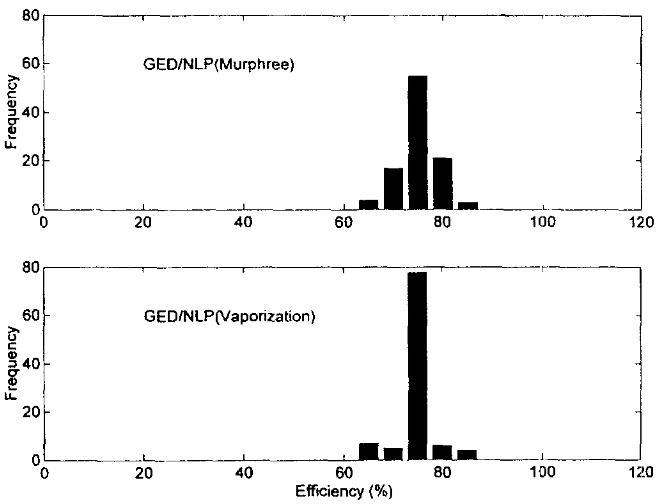

Fig. 7. Histograms of the stage efficiency estimates by GED/NLP for different efficiency models.

In order to assess the robustness of the methods to the model forms, two different efficiency models (Vaporization and Modified Murphree model) were tested and the results are summarized in Tables 3 and 5. The GED/SL shows quite different performance with different stage efficiency models. The most striking differences in the results of GED/SL and DR/SL are that the values of maximum and minimum stage efficiency are over 100% for Vaporization model, which are not realistic values for a real process. In the case of GED/NLP, however, little distinction is noted between the results of two models. In GED/SL, the histograms in Fig. 6 show that there are more outliers in the estimates with the Murphree model. However, the results of GED/NLP in Fig. 7 show both distributions in the vicinity of the true stage efficiency, even though the efficiency models are different.

CONCLUSION

Two data reconciliation and gross error detection methods were tested to estimate the unmeasured stage efficiency in a flash distillation column. Compared to the conventional methods, the modified MIMT using NAP showed consistent performance regardless of the number of gross errors, the type of systematic errors and the stage efficiency models.

We believe optimal estimates of measured variables and unmeasured variable or parameters can be computed using the enhanced data reconciliation and gross error detection algorithm using NLP for a variety of chemical engineering processes.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support from the Automation Research Center at POSTEC and the KOSEF Grant 94-1400-01-01-3.

NOMENCLATURE

C : number of components
F : feed flow rate [kmol/s]
H : enthalpy [kJ/kmol]
K : K value
L : liquid flow rate [kmol/s]
P : pressure [kPa]
Q : heat rate [kJ/s]
$\mathbf{R}$: covariance matrix of measurement errors
T : temperature [K]
V : vapor flow rate [kmol/s]
$\mathbf{w}$: estimate of suspected variables
x : liquid mole fraction
y : vapor mole fraction
$\mathbf{x}$: vector of true measurement
z : reconciled data
$\mathbf{z}_m'$: compressed measurement data set deleted from suspected variable

Superscripts

MM : modified Murphree efficiency
V : vaporization efficiency

Subscripts

F : feed stream
i : component index

j : general variable index
L : liquid stream

REFERENCES

Crowe, C., "Reconciliation of Processes Flow Rates by Matrix Projection II. The Nonlinear Case", *AIChE J.*, **32**, 616 (1986).

Hogg, R. V. and Tanes, E. A., "Probability and Statistical Inference", MacMillan Publishing Co., New York, 1977.

Holland, C. D., "Fundamentals of Multicomponent Distillation", Mcgraw-Hill, New York (1981).

Iordache, C., Mah, R. and Tamhane, A., "Performance Studies of the Measurement Test for Detection of Gross Errors in Process Data", *AIChE J.*, **31**, 1187 (1985).

Kao, C. S., Tamhane, A. C. and Mah, R. S. H., "Gross Error Detection in Serially Correlated Process Data", *Ind. Eng. Chem. Res.*, **29**, 1004 (1990).

Knepper, J. and Gorman, J., "Statistical Analysis of Constrained Data Sets", *AIChE J.*, **26**, 260 (1980).

Kim, I. W., Kang, M. S., Park, S. W. and Edgar, T. F., "Robust Data Reconciliation and Gross Error Detection: The Modified MIMT Using NLP", submitted to *Comp. Chem. Eng.* (1995).

Ripps, D. L., "Adjustment of Experimental Data", *Chem. Eng. Prog. Symp. Ser.*, **61**, 8 (1965).

Rogenberg, J., Mah, R. S. H. and Iordache, C., "Evaluation of Schemes for Detecting and Identifying Gross Errors in Process Data", *Ind. Eng. Chem. Res.*, **26**, 555 (1987).

Serth, R. and Heenan, W., "Gross Error Detection and Data Reconciliation in Steam-Metering Systems", *AIChE J.*, **32**, 733 (1986).

Serth, R. W., Srikanth, B. and Maronga, S. J., "Gross Error Detection and Stage Efficiency Estimation in a Separation Process", *AIChE J.*, **39**(10), 1726 (1993).

Serth, R., Valero, C. and Heenan, W., "Detection of Gross Errors in Nonlinearly Constrained Data: A Case Study", *Chem. Eng. Comm.*, **51**, 89 (1987).

Tamhane, A. C. and Mah, R. S. H., "Data Reconciliation and Gross Error Detection in Chemical Process Networks", *Technometrics*, **27**, 409 (1985).

Terry, P. A. and Himmelblau, D. M., "Data Rectification and Gross Error Detection in a Steady-State Process via Artificial Neural Networks", *Ind. Eng. Chem. Res.*, **32**, 3020 (1993).