Data Reconciliation and Parameter Estimation in Plant Performance Analysis

Results of a computer study are reported applying data reconciliation and interpretation techniques to a single-stage flash operation. Measured errors in the process variables are significantly reduced. The techniques are successfully extended to estimate flash efficiency. The flash efficiency is estimated using two methods: (1) a statistically rigorous simultaneous data reconciliation and efficiency estimation procedure, and (2), a faster, nonrigorous sequential procedure. It is found that the simultaneous procedure provides better measured variable and flash efficiency estimates.

R. J. MacDonald, C. S. Howat
Kurata Thermodynamics Laboratory
Department of Chemical and Petroleum Engineering
University of Kansas
Lawrence, KS 66045

Introduction

The understanding of chemical plant performance is important because operators and designers frequently must make decisions based on current plant operations, and because automated process control systems must be relied upon to make decisions concerning current operation without human intervention. In either case, decisions are reached by analyzing the data in accordance with some mathematical model of the process and adjusting the model accordingly. The model then becomes a predictor for design and operation.

For proper interpretation, the plant data must satisfy the material and energy balance constraints for the process. They frequently do not because of plant measurement errors. There are two types of errors, random and systematic. Both must be recognized and handled in order to properly reconcile the data to the constraints and interpret the plant performance.

Therefore, plant performance analysis can be described by five steps prior to using the model as a predictor:

1. Development of theoretical model
2. Plant measurement of necessary data
3. Analysis of data for systematic errors
4. Reconciliation of data to the constraints
5. Interpretation of the reconciled data and adjustment of the model

Development of the theoretical model is covered in the simulation literature and needs no further comment. Historically, steps 2 through 5 have been accomplished by individual engineers relying on experience. This presents the potential that two engineers could come to two different interpretations. A formal method for analyzing and interpreting plant performance is required to minimize the potential for erroneous conclusions in design, operation, and control.

Recent literature discusses the importance of handling incomplete data sets (step 2), gross (systematic) error identification (step 3), and data reconciliation (step 4). Heretofore, formal interpretation (step 5) has not been discussed.

Plant performance analysis was first considered by Kuehn and Davidson (1961). Britt and Leucke (1973) and Knepper and Gorman (1980) provided an algorithm that can be used to adjust plant data to meet the constraints. In the last ten years, considerable effort has been expended to develop procedures to handle very large flowsheets. See reviews by Hlaváček (1977) and Mah (1981).

Crowe (1986) describes a complete method for reconciling data and recognizing gross errors. Rosenberg et al. (1987) test several gross error detection methods. Successful industrial application and testing has begun to be implemented (Stanley and Mah, 1977; Ham et al., 1979; Beckman, 1982; Serth and Heenan, 1986). These applications focused on detecting gross errors and adjusting the process data to satisfy the material and energy balance constraints.

These statistical procedures can be extended to analyze unit operations to obtain performance parameter estimates, for example, tray efficiencies for distillation, heat transfer coefficients (Stephenson and Shewchuk, 1986), and reaction rate constants. Hlaváček (1977) suggested that parameter estimation can be done sequentially after reconciliation or simultaneously with it.

The purpose of this work is to extend data reconciliation techniques to estimate process parameters. Two developments are presented. The first is a sequential, "decoupled" procedure that reconciles the data to satisfy the material and energy balances,

Correspondence concerning this paper should be addressed to C. S. Howat.

and then estimates the process parameters using maximum-likelihood estimation. The second is a "coupled" procedure that simultaneously reconciles the data to satisfy the constraints and estimate the process parameters. The decoupled procedure is computationally faster and is more easily adapted to existing reconciliation algorithms, but is not statistically rigorous. The coupled procedure is statistically rigorous.

This article discusses the results of a computer study of the two procedures applied to flash operations. The purpose of the study was to determine the relative merits of both procedures with respect to reconciliation and efficiency estimation. The principal restriction was that there were no uncertainties in the physical properties. Both methods adjust the data such that deviations from the true values are significantly reduced compared to the simulated plant measurements, with the coupled procedure marginally better. Both methods accurately predict the flash efficiency, but the coupled procedure gives a better estimate for its precision.

Mathematical Procedures

Flash development

A single-stage flash is shown schematically in Figure 1. A feed stream F containing a mixture of components is separated into a light vapor phase V and a heavy liquid phase L. There are $3C + 8$ process variables, where C is the number of components. These variables are placed in an $[n \cdot 1]$ vector Z Eq. 1:

$$Z = \begin{bmatrix} T_F \\ P_F \\ T \\ P \\ F \\ L \\ V \\ Q \\ Z_{itoc} \\ X_{itoc} \\ Y_{itoc} \end{bmatrix} \tag{1}$$

The constraints are shown in Table 1. These are written as a constraint vector $F(Z)$ that is equal to zero when satisfied. Initially, the measured data, Z_o, will not satisfy the constraints. The data are adjusted until a best set $\hat{Z}$ is found that satisfies $F(\hat{Z}) = 0$.

In practice, flashes are nonequilibrium, and separation is less than that achieved in an equilibrium flash. Consequently, it is convenient to define an efficiency parameter that reflects this difference. The efficiency was defined as the ratio of the actual separation to the equilibrium separation for each component:

$$\theta_i = (y_i - x_i)/(y_i^* - x_i^*) \tag{2}$$

The equilibrium flash is calculated at the same feed conditions and requires two other independent specifications. Initially, these were chosen to be T and P; they were later changed to T and V/F. A limitation of using efficiencies is that only $C - 1$ θ_i's are independent (King, 1980, p. 133).

Several key assumptions were made for the purposes of this study:

1. No gross errors exist in the measurements
2. The random errors are independent and normally distributed
3. All process variables are measured with known uncertainty
4. Only one set of measurements is taken
5. The flash is operating at steady state
6. All components have the same efficiency
7. No uncertainty exists in the data base

The mathematical developments to follow are not inherently limited by these restrictions, with the exception of the normality specification. If component efficiencies are assumed to be unequal, multiple data sets are required.

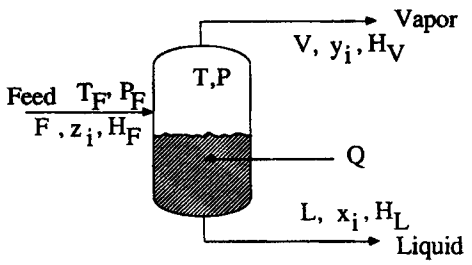

Figure 1. Single-stage flash operation.

Table 1. Material and Energy Balance and Summation Equation Constraints

$Fz_i - Lx_i - Vy_i = 0$	C
$H_F F + Q - H_L L - H_V V = 0$	1
$\Sigma z_i - 1 = 0$	1
$\Sigma x_i - 1 = 0$	1
$\Sigma y_i - 1 = 0$	1
	Total = $C + 4$ constraints

Statistical development

The statistical development follows that of Britt and Leucke (1973). For one set of measurements, the probability density function for the measured data is:

$$f(Z_o) = \frac{1}{(2\Pi)^{n/2}|R|^{1/2}} \exp\left[\frac{(Z_o - \tilde{Z})^T R^{-1}(Z_o - \tilde{Z})}{2}\right] \tag{3}$$

where

R is the variance-covariance matrix for the measured data

$|R|$ is the determinant of R

Z_o represents the measured data values

$\tilde{Z}$ represents the true data values

Maximum likelihood estimation reduces the problem to the following:

Minimize: $(Z_o - \hat{Z})^T R^{-1}(Z_o - \hat{Z})$ (4)

Subject to the constraints: $F(\hat{Z}) = 0$

The quantity $(Z_o - \hat{Z})^T R^{-1}(Z_o - \hat{Z})$ reduces to:

$$\sum_{i=1}^{n} \frac{(Z_0 - \hat{Z})_i^2}{(\sigma_i)^2} \tag{5}$$

when there is no covariance among the measured data. Thus it can be seen that the statistical "best" adjustment to the data is that which satisfies the constraints while minimizing the weighted sum of squares adjustment to the measured data.

Decoupled procedure

The decoupled procedure is a two-step calculation. First, Eq. 4 is solved to satisfy the constraints. Second, the adjusted data, $\hat{Z}$, are used to calculate the efficiency $\hat{\theta}$. The procedure of Britt and Leucke (1973) is applicable for step one. They use the method of Lagrange multipliers with a Taylor expansion about the nonlinear constraints. Their iterative algorithm is:

$$\hat{Z}_{k+1} - \hat{Z}_k = Z_o - \hat{Z}_k - RF_z^T(F_z R F_z^T)^{-1}[F(\hat{Z}_k) + F_z(Z_o - \hat{Z}_k)] \tag{6}$$

where F_z is the $[m \cdot n]$ Jacobian matrix resulting from differentiating the constraint equations F, Table 1, with respect to $\hat{Z}_k$. Note that all the derivatives are analytically simple except for the energy balance derivatives, which are calculated by finite difference. Updating techniques to approximate these derivatives, such as that of Lucia and Macchietto (1983), should not be used since the correct solution depends on accurate derivatives.

At each iteration, the Jacobian is recalculated at the new estimate, and Eq. 6 is recalculated until the absolute values of the constraint discrepancies are less than some tolerance, that is, $|F(\hat{Z}_{k+1})| < \epsilon$.

Britt and Leucke (1973) do not derive the variance-covariance matrix for $\hat{Z}$ that is required for the efficiency calculation. However, it is given by Knepper and Gorman (1980):

$$\mathcal{E}v[(\hat{Z} - \tilde{Z})(\hat{Z} - \tilde{Z})^T] = R - RF_z^T(F_z R F_z^T)^{-1}F_z R \tag{7}$$

Once $\hat{Z}$ are obtained, the component efficiencies are estimated. An equilibrium flash is done at $\hat{T}$, $\hat{V}/\hat{F}$, and $\hat{z}$. The equilibrium compositions are used to determine $\hat{\theta}_i$ (which may be unequal) by Eq. 2. Maximum likelihood is then used to estimate an overall efficiency $\hat{\theta}$. V, the variance-covariance matrix for $\hat{\theta}_i$, is calculated from the estimated variance-covariance matrix for $\hat{Z}$, Eq. 7, and propagation of error in Eq. 2 as follows:

$$v_{ii} = \frac{1}{(y_i^* - x_i^*)^2}[\mathrm{var}(\hat{x}_i) + \mathrm{var}(\hat{y}_i) - 2\,\mathrm{cov}(\hat{x}_i, \hat{y}_i)] \tag{8}$$

$$v_{ij} = \frac{1}{(y_i^* - x_i^*)(y_j^* - x_j^*)} \cdot [\mathrm{cov}(\hat{x}_i, \hat{x}_j) + \mathrm{cov}(\hat{y}_i, \hat{y}_j) - \mathrm{cov}(\hat{x}_i, \hat{y}_j) - \mathrm{cov}(\hat{x}_j, \hat{y}_i)] \tag{9}$$

Although there is no covariance in the measured values x_i and y_i, covariance in the adjusted values $\hat{x}$ and $\hat{y}$ is introduced as a result of the constraints. The maximum likelihood estimate for $\hat{\theta}$ is then:

$$\text{Minimize:} \quad (\theta - \hat{\theta})^T V^{-1}(\theta - \hat{\theta}) \tag{10}$$

where

$$\theta = \begin{bmatrix} \hat{\theta}_1 \\ \overline{\hat{\theta}_3} \end{bmatrix} \qquad \hat{\theta} = \begin{bmatrix} \hat{\theta} \\ \overline{\hat{\theta}} \end{bmatrix}$$

V = variance-covariance matrix for θ_i

$\hat{\theta}_i$ = estimated component efficiency, Eq. 2

$\hat{\theta}$ = overall estimated efficiency

The solution of Eq. 10 is:

$$\hat{\theta} = (\beta^T V^{-1}\beta)^{-1}\beta^T V^{-1}\theta \tag{11}$$

$$\mathrm{var}(\hat{\theta}) = (\beta^T V^{-1}\beta)^{-1} \tag{12}$$

where

$$\beta = \begin{bmatrix} 1 \\ \overline{1} \end{bmatrix} \text{dimensions: } [C - 1, 1]$$

Since only $C - 1$ of the θ_i are independent, V is singular for the C components. However, testing showed that the equivalent results are obtained using any combination of $C - 1$ components. Thus, in this calculation only, Eqs. 11–12, V and θ have dimensions $[C - 1 \cdot C - 1]$ and $[C - 1 \cdot 1]$, respectively.

Coupled procedure

The method just described for the decoupled procedure is a simplified version of the developments of Britt and Leucke (1973). The more general problem is one where a number of parameters θ are actually part of the constraint equations, that is, find the "best" estimates $\hat{Z}$ and $\hat{\theta}$ simultaneously such that $F(\hat{Z}, \hat{\theta}) = 0$. Including the efficiency definition, Eq. 2, as a constraint allows this more rigorous approach to be used. This forces the data to be adjusted to satisfy the material and energy balances, the summation equations, and equal component efficiencies. Thus the coupled procedure calculates $\hat{Z}$ and $\hat{\theta}$ to satisfy the constraints in Table 1 with the following additional $C - 1$ constraints:

$$\frac{y_i - x_i}{y_i^* - x_i^*} - \hat{\theta} = 0 \tag{13}$$

For the coupled procedure, there are $(C + 4) + (C - 1) = 2C + 3$ constraints.

From maximum likelihood, the problem now becomes:

$$\begin{aligned} &\text{Minimize:} && (Z_o - \hat{Z})^T R^{-1}(Z_o - \hat{Z}) \\ &\text{Subject to the constraints:} && F(\hat{Z}, \hat{\theta}) = 0 \end{aligned} \tag{14}$$

Following the development of Britt and Leucke (1973), the iteration equations are:

$$\hat{\theta}_{k+1} - \hat{\theta}_k = -[F_\theta^T(F_z R F_z^T)^{-1}F_\theta]^{-1}F_\theta^T(F_z R F_z^T)^{-1} \cdot [F(\hat{Z}_k, \theta_k) + F_z(Z_o - \hat{Z}_k)] \tag{15}$$

$$\hat{Z}_{k+1} - Z_k = Z_o - \hat{Z}_k - RF_z^T(F_z R F_z^T)^{-1}[F(\hat{Z}_k, \hat{\theta}_k) + F_\theta(\hat{\theta}_{k+1} - \hat{\theta}_k) + F_z(Z_o - \hat{Z}_k)] \tag{16}$$

where F_z now includes the partial derivatives for Eq. 13, and F_θ is an $[m \cdot 1]$ Jacobian matrix obtained by differentiating F with respect to $\hat{\theta}$.

The estimated variance-covariance matrix for $\hat{Z}$ is:

$$\mathcal{E}v[(\hat{Z} - \tilde{Z})(\hat{Z} - \tilde{Z})^T] = R - RF_z^T(F_zRF_z^T)^{-1}F_zR + RF_z^T(F_zRF_z^T)^{-1}F_\theta[F_\theta^T(F_zRF_z^T)^{-1}F_\theta]^{-1}F_\theta^T(F_zRF_z^T)^{-1}F_zR \quad (17)$$

The variance-covariance matrix for $\hat{\theta}$ (in this case, a $1 \cdot 1$ matrix) is:

$$\mathcal{E}v = [(\hat{\theta} - \tilde{\theta})(\hat{\theta} - \tilde{\theta})^T] = [F_\theta^T(F_zRF_z^T)^{-1}F_\theta]^{-1} \quad (18)$$

Finite-difference methods are used to calculate derivatives in the nonequilibrium constraints, Eq. 13, since x^* and y^* are complex functions of several variables. This is the primary cause of the longer execution time since the equilibrium flash must be recalculated several times per iteration.

Several authors have derived Eqs. 15–18. The method of Britt and Leucke using Lagrange multipliers has been criticized for requiring good starting values. Anderson et al. (1978) use a Newton-Raphson approach, although their method is limited to only two constraints. This limitation is lifted by Ricker (1984), who compares the two methods and finds that while the method of Anderson et al. is less likely to exhibit convergence problems, it requires substantially more computer storage and execution time. The work of Britt and Leucke was adequate for this work: No convergence difficulties were encountered.

The rigorous and constant-direction methods of Knepper and Gorman (1980) were also tested. The rigorous method is identical to that of Britt and Leucke. The constant-direction method required more computing time. This is not a general result, however, and more complex operations might show the advantage of the constant-direction method, perhaps for difficult convergence situations.

Method of Testing

Monte Carlo testing with 100 simulations was used to test the coupled and decoupled procedures. The testing procedure is shown schematically in Figure 2. First, a nonequilibrium flash was made giving canonical or "true" values for $Z(\tilde{Z})$. Then, random normal errors were added to each of the measurements, thus simulating Z_o. The values Z_o were adjusted by the coupled and decoupled procedures, the results of which are $\hat{Z}$ and $\hat{\theta}$. The adjusted values, $\hat{Z}$, were compared to the canonical values to determine if the statistical procedures provided better estimates for $\tilde{Z}$ than Z_o.

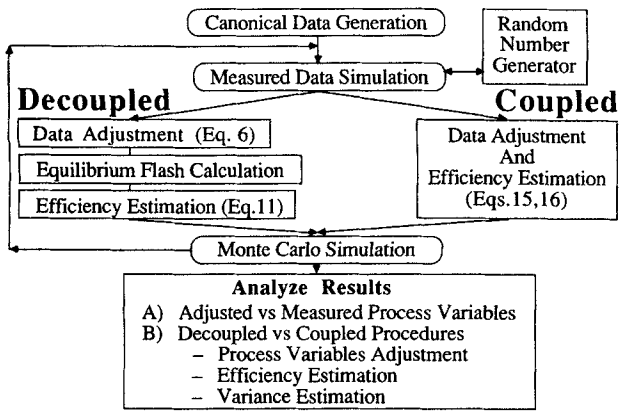

Figure 2. Monte Carlo simulation test procedure.

Both coupled and decoupled procedures estimate the variances in $\hat{Z}$, which can be compared with the actual scatter in $\hat{Z}$ from the Monte Carlo simulation. Thus, a check is provided for the statistical procedures. The same comparison can be made for the efficiency variance. The estimated variances in $\hat{Z}$ and $\hat{\theta}$ provide valuable information, giving the operator or designer the opportunity to place a confidence interval on any decision made.

The nonequilibrium flash calculation requires two steps. First an equilibrium flash is solved specifying feed, P, Q (Henley and Seader, 1981). Then, the equilibrium equations ($y_i^* = K_i x_i^*$) are replaced by the nonequilibrium equations $(y_i - x_i) - \theta(y_i^* - x_i^*) = 0$, and the system is resolved. The simultaneous "hybrid" Newton method of Lucia and Macchietto (1983) is used as the solution method. Because only $C - 1$ of the nonequilibrium equations are independent, one degree of freedom is lost and the nonequilibrium flash is underspecified. The additional specification is that the nonequilibrium flash temperature is equal to the resultant equilibrium flash temperature. Thus, the nonequilibrium flash is specified at the same flash temperature, pressure, and heat input. An alternative is to specify flash temperature, pressure, and vapor-to-feed ratio.

Z_o was used as the starting estimate of $\hat{Z}$. The decoupled procedure was run first and its converged results were used as the initial estimates for the coupled procedure to speed coupled convergence.

Results

The Monte Carlo simulations were used to determine for both coupled and decoupled procedures:

- The relative accuracy of Z_o and $\hat{Z}$ compared to the canonical values $\tilde{Z}$
- The accuracy and precision of the efficiency estimate
- The accuracy of the estimated variances in $\hat{Z}$ and θ

Several flashes having a variety of characteristics were simulated. The following parameters were varied:

1. Efficiency
2. Feed boiling range
3. Vapor-to-feed ratio
4. Feed composition
5. Measurement uncertainty

This paper presents the results of one typical flash. The results are then generalized to the other cases. Table 2 shows the canonical data, $\tilde{Z}$, and Table 3 lists the specified measurement uncertainty for each variable.

Adjustments to process variables

Table 4 lists the statistical information concerning the adjustment of the measured data for the coupled and decoupled procedures.

Column two (% runs with error reduction) shows the improvement in accuracy of the variables due to the adjustment proce-

Table 2. Wide-Boiling Flash at 75% Efficiency

	Feed		Vapor		Liquid	
Component	Z	kmol/h	Y	kmol/h	X	kmol/h
Propane	0.60000	600.000	0.80515	280.776	0.49015	319.224
n-Butane	0.20000	200.000	0.12689	44.251	0.23915	155.749
n-Pentane	0.20000	200.000	0.06796	23.699	0.27070	176.301
Total	1.00000	1,000.000	1.0000	348.725	1.0000	651.274
T, K		300.00		259.87		259.87
P, kPa		590.40		150.00		150.00
H, kJ/kgmol		−4,899.84		9,090.70		−10,855.63

Specified heat duty = 1,000,000 KJ/HR
Flash efficiency = 75.00%

dure. The value shown is the percentage of runs in which the absolute error was reduced by the adjustment.

The flash temperature T is seen to be improved only 42–44% of the time, which is worse than purely random. However, there is systematic improvement to the flow rates F, L, V and heat Q. The best improvement is seen for the mole fractions z, x, y, where the only poor adjustment is for y_2 and y_3. The coupled procedure resulted in equivalent or better improvements than those resulting from the decoupled procedure.

Column three (coupled more accurate, %) compares the accuracy of the coupled and decoupled procedures. A value above 50% shows that the coupled proceduure was more accurate in the majority of runs. Stanley and Mah (1977) hypothesized that simultaneous estimation of parameters and reconciliation of the measured data might degrade the adjustment to the data. There is little difference in the accuracy of the methods, with the coupled procedure providing marginally better compositions.

Columns four and five compare the measured and adjusted root mean square error (RMSE) for the measured and adjusted values, Eqs. 19 and 20.

$$\text{Measured RMSE for } Z^{(K)} = [\Sigma(\mathbf{Z}_o^{(K)} - \tilde{Z}^{(K)})^2/s]^{1/2} \tag{19}$$

$$\text{Adjusted RMSE for } Z^{(K)} = [\Sigma(\hat{Z}^{(K)} - \tilde{Z}^{(K)})^2/s]^{1/2} \tag{20}$$

where the superscript (K) indicates that this is done for each variable, and the summation is over s simulations.

The measured RMSE should be close to the specified uncertainty in Table 3. When the adjusted RMSE is lower than the measured, the adjustment procedures are successful in reducing measured error. For instance, the specified measurement uncertainty in the feed temperature T_F is 2 K. The measured RMSE in the sample was 2.1 K, but after adjustment, the RMSE was reduced to 1.5 K for both procedures.

Table 3. Process Variable Measurement Uncertainties

Process Var.	1 Std. Dev.
Temperature	2 K
Pressure	5%
Flow rates	5%
Steam	15%
Mole fractions	10%

There is no adjustment to the feed pressure, and insignificant adjustment to the flash pressure. This is because the constraints are nearly independent of the pressures.

There is error reduction to the temperatures, flow rates, and heat. The greatest improvement is for the mole fractions. For example, the RMSE in z_1 is reduced from 9.4% to only 2.1–2.2% after adjustment.

The last column (estimated standard deviation) shows how well the procedures estimate the uncertainty in the adjusted process variables. Ideally, the estimated standard deviation should be equal to the adjusted RMSE. In the cases tested, the bias was usually insignificant, thus the RMSE and estimated standard deviations are directly comparable. It can be seen that the estimated standard deviations are good estimators. Overall then, both adjustment procedures reduce measured error and successfully estimate the uncertainty of the adjusted process variables.

The results in Table 4 are typical of the various flashes tested.

Table 4. Error Reduction to Process Variables

Process Var.	% Runs w/Error Reduc.		Coup. More Acc. %	RMSE*			Est. Std. Dev.	
					Adj.			
	Dec.	Coup.		Meas.	Dec.	Coup.	Dec.	Coup.
T_F	62	62	37	2.1	1.5	1.5	1.6	1.6
P_F	No adjustment to feed stream pressure							
T	44	42	48	1.9	1.8	1.8	1.8	1.8
P	60	60	51	5.1	5.1	5.1	5.0	5.0
F	68	68	49	4.6	3.0	3.0	3.0	3.0
L	63	62	47	5.2	3.7	3.7	3.9	3.9
V	67	68	48	5.2	4.2	4.2	4.1	4.1
Q	68	66	44	14.7	13.0	13.0	14.3	14.3
Z_1	86	85	43	9.4	2.1	2.2	2.6	2.6
Z_2	74	77	56	9.3	5.3	5.0	5.9	5.4
Z_3	67	66	62	10.8	6.6	6.4	6.1	6.1
X_1	73	72	43	8.5	3.8	4.1	4.8	4.5
X_2	67	78	64	9.9	6.5	4.9	7.4	5.3
X_3	63	64	62	10.2	7.6	7.2	6.9	6.8
Y_1	83	86	50	10.0	1.8	1.6	1.7	1.5
Y_2	44	56	65	9.1	9.1	6.2	9.7	6.5
Y_3	48	56	59	11.4	11.4	11.2	9.9	9.5

*T_F and T in Kelvins, all other variables in %

Table 5. Efficiency Estimate for Coupled and Decoupled Procedures

	Flash Eff., 75%	
	Dec.	Coup.
Overall Eff., %		
Avg. value	75.0	75.0
Range	28.2	28.6
Overall Statistical Population Tests		
Actual RMSE, %	4.8	4.8
Est. Std. Dev., %		
Avg.	6.5	4.3
% Runs within x Est. Std. Dev.		
±1 std. dev.	87	71
±2 std. dev.	98	93
±3 std. dev.	100	98
% Runs with Coup.		
Error < Dec. Error		51

It should be noted that the adjustment to each variable is dependent on the flash characteristics. For example, a large flow rate is adjusted more than a small flow rate. This follows from using uncertainty on a percentage basis. Other error specifications would result in different adjustments.

Scatter plots (MacDonald, 1986) comparing the measured and adjusted errors show that the procedures are most successful correcting the worst measurements (>±2 standard deviations). Adjustments to satisfy the constraints tend to be spread among several variables. Consequently, some variables that are measured with a very low error are adjusted, making their adjusted values slightly worse than the measured value.

The coupled procedure may provide better estimates for the process variables because it contains the constraint that the efficiencies are equal. Since this accurately reflects the foundation of the canonical data, there is a bias in variable adjustment in favor of the coupled procedure. Further testing is under way to eliminate this bias in the comparison.

Efficiency estimation

Table 5 lists the efficiency estimation results for the Table 2 flash. First, the average efficiency and range are shown. The average value is unbiased.

Second, the root mean square error (RMSE) in the efficiency for both procedures is 4.8%. This estimate results from the population of Monte Carlo simulations. These values should be compared to the estimated standard deviations. The decoupled procedure overestimates the standard deviation with a value of 6.5%, while the coupled procedure slightly underestimates it at 4.3%. This trend prevailed for all the other flashes tested, with the coupled procedure providing more accurate estimates for the efficiency standard deviation.

Next the percentage of runs within 1, 2, and 3 estimated standard deviations are shown. For the coupled procedure, 71% of the runs were within one estimated standard deviation—that is, 75% ± 4.3%—and 93% were within two standard deviations. This shows that the estimates are effectively normally distributed.

The final statistic shows that the coupled and decoupled procedures were essentially equivalent in accurately predicting the efficiency. The coupled procedure provided better estimates in 51% of the simulations.

The example flash was fairly typical. In the other flashes tested, there was usually very little difference between coupled and decoupled procedures, except for a small fraction of the cases when the coupled procedure was slightly better. The decoupled procedure consistently and sometimes significantly overestimated the efficiency standard deviation. The coupled procedure was more accurate, although consistently underestimating it slightly.

The efficiency using the measured data without any adjustment was also calculated to give an idea of the improvement due to the adjustment procedures. Mole fractions were merely normalized. The results were much worse, with a RMSE in efficiency of 10%, compared to 4.8% for the example case. The range of estimates was 71%, compared to 28–29% using the adjusted data.

The results for the variety of other flash conditions that were studied by MacDonald (1986) are generalized below.

The lower the efficiency, the greater the uncertainty is in the estimate. This follows from the efficiency definition, Eq. 2. The lower the efficiency, the lower the actual separation $y_i - x_i$. But the uncertainty in y and x is not a function of the efficiency, so on a percentage basis, the uncertainty in $y_i - x_i$ is greater at low efficiencies. For instance, the same feed flashed at 25% efficiency resulted in a RMSE of 8.0% for both procedures.

For a narrow-boiling flash, the separation $y_i - x_i$ is near zero,

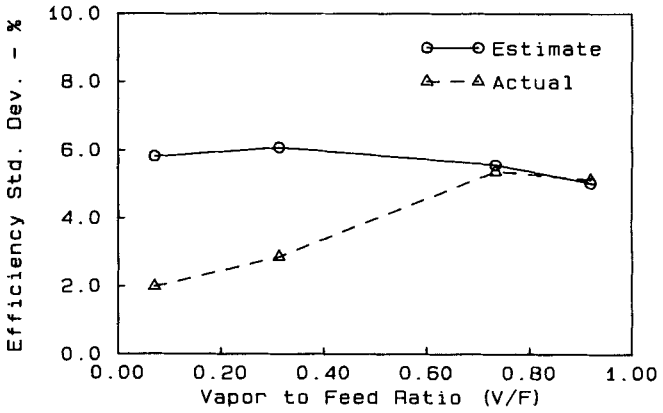

Figure 3. Effect of vapor-to-feed ratio on efficiency estimate standard deviation using decoupled procedure.

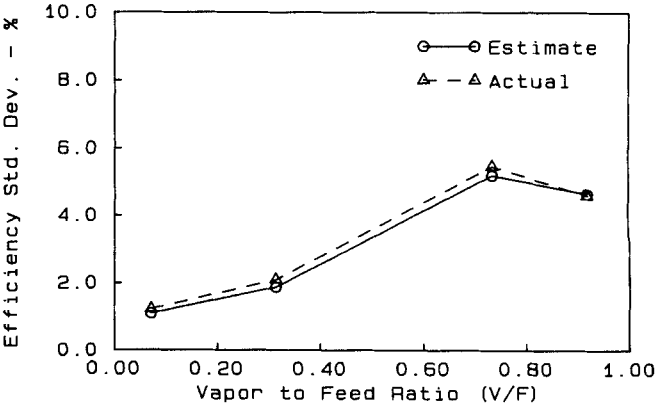

Figure 4. Effect of vapor-to-feed ratio on efficiency estimate standard deviation using coupled procedure.

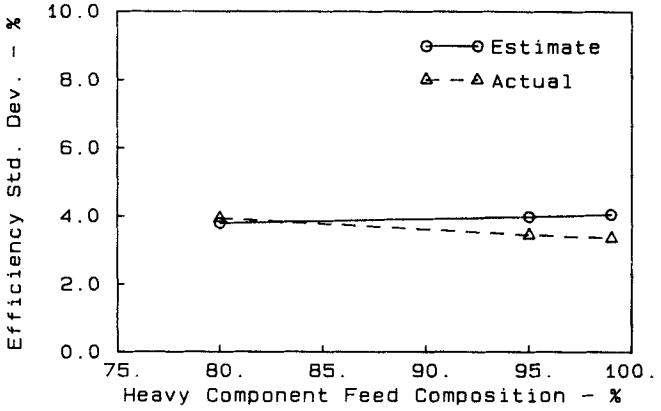

Figure 5. Effect of feed composition on efficiency estimate standard deviation using decoupled procedure.

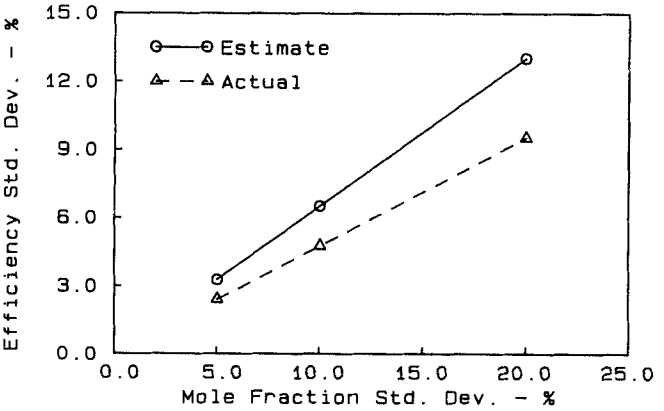

Figure 7. Effect of composition standard deviation on efficiency estimate standard deviation using decoupled procedure.

so the same problem occurs. Unfortunately, the uncertainty becomes so high that the efficiency estimate is useless. For one narrow-boiling flash tested ($\alpha \approx 1.2$), uncertainties in the efficiency were greater than 50%. It is some consolation that narrow flashes are impractical by themselves, and thus would not be encountered in a plant performance analysis. Also, the procedures do predict the high uncertainty, and would indicate when the results are poor. The two procedures could still be used to adjust the process variables to satisfy the constraints.

The vapor-to-feed ratio has a strong effect on the standard deviation of the efficiency estimate. Figures 3 and 4 compare the actual and estimated standard deviations for both the decoupled and coupled procedures, respectively, as a function of vapor-to-feed ratio. Note that the decoupled procedure seriously overestimates the efficiency standard deviation at low V/F. Further note that the coupled procedure provides a lower standard deviation across the entire range, indicating that the coupled procedure provides an efficiency estimate in which more confidence can be placed.

The impact of feed composition on efficiency standard deviation was determined by running a series of flashes where two lighter components with low feed concentration were flashed from a principal heavy component. Figures 5 and 6 present the results. Note that both procedures accurately estimate the standard deviation and provide essentially equivalent estimates.

As process variable measurement uncertainty increases, so does the uncertainty in the efficiency estimate. Figures 7 and 8 show the efficiency standard deviation as a function of the mole fraction measurement uncertainty. The coupled procedure estimated uncertainty is again a better estimator of the actual value. Furthermore, note that the decoupled procedure is a poorer estimator at higher measurement uncertainties, which is a reflection of its lack of statistical rigor.

The adjustments to the process variables and determination of the efficiency estimate are dependent upon the definitions and values of the measurement uncertainties. Accurate measurement uncertainties are required in order to obtain accurate interpretation of the plant performance.

Conclusions

Both coupled and decoupled procedures were successful in significantly reducing measured error in the process variables. The adjusted data can be expected to be more accurate than the measured values, with the exception of the feed and flash pressures, which are not adjusted. The coupled procedure adjusts the

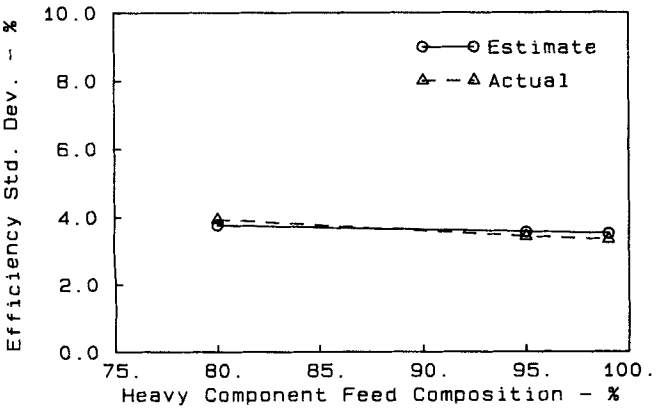

Figure 6. Effect of feed composition on efficiency estimate standard deviation using coupled procedure.

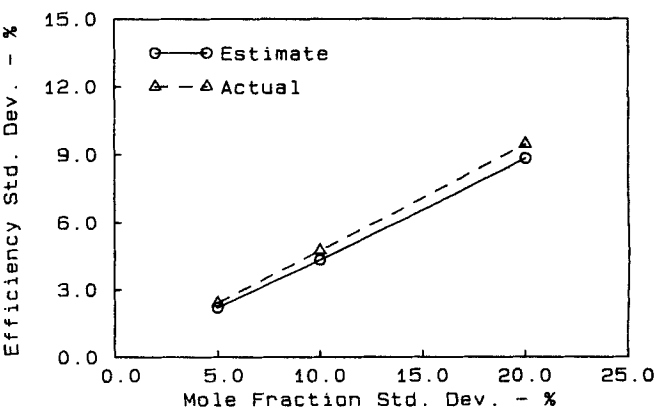

Figure 8. Effect of composition standard deviation on efficiency estimate standard deviation using coupled procedure.

process variables better than the decoupled procedure. Both procedures tend to slightly overestimate the uncertainty in the adjusted process variables.

Both procedures result in more accurate and precise efficiency estimates than if the data are not adjusted. In most cases, there was no significant difference in the efficiency estimate calculated using the coupled and decoupled procedures.

The coupled procedure provided more accurate estimates of the flash efficiency precision. The differences between the two become larger as the measurement uncertainty increases.

The coupled procedure requires substantially more computation time. It provides only marginal improvement over the decoupled procedure, with the exception of efficiency standard deviation prediction. Therefore, the decoupled procedure may be adequate, especially when measurement uncertainties are known to be low.

Acknowledgment

Partial support of this work by the Henry L. Doherty Educational Foundation is gratefully acknowledged.

Notation

C = number of components
F = feed flow rate, kmol/h
$\boldsymbol{F}$ = m vector of constraint equations
$\boldsymbol{F}_z$ = $m \cdot n$ Jacobian matrix obtained by differentiating $\boldsymbol{F}$ with respect to $\hat{\boldsymbol{Z}}$
$\boldsymbol{F}_\theta$ = $m \cdot 1$ Jacobian matrix obtained by differentiating $\boldsymbol{F}$ with respect to $\hat{\boldsymbol{\theta}}$
H = enthalpy, kJ/kmol
K = K value
L = liquid flow rate, kmol/h
m = number of constraint equations: decoupled, $C + 4$; coupled, $2C + 3$
n = number of process variables $3C + 8$
P = pressure, kPA
Q = heat requirement, kJ/h
$\boldsymbol{R}$ = variance-covariance matrix for the measured data
RMSE = Root mean square error
s = number of simulations
T = temperature, K
V = vapor flow rate, kmol/h
$\boldsymbol{V}$ = variance-covariance matrix for θ_i
v_{ij} = row i, column j entry in $\boldsymbol{V}$
x = liquid phase mole fractions
y = vapor phase mole fractions
$\boldsymbol{Z}$ = m vector process variables
Z = general process variable
z = feed mole fractions

Greek letters

$\boldsymbol{\beta}$ = $[C - 1, 1]$ unit matrix
α = relative volatility
ϵ = constraint equation convergence tolerance
$\boldsymbol{\theta}$ = θ_i vector
θ = specified (true) flash efficiency
θ_i = component flash efficiency, Eq. 2
σ^2 = variance; σ = standard deviation, units vary

Subscripts

F = feed
i = component
j = component
k = kth iteration
l = process variable
L = liquid
v = vapor
o = initial/measured values

Superscripts

$\hat{}$ = estimated or adjusted
$\tilde{}$ = true value
$Z^{(K)}$ = Kth element of vector $\boldsymbol{Z}$
T = transpose of matrix
$*$ = equilibrium
-1 = inverse of matrix

Statistical functions

$\mathcal{E}v$ = expected value
cov = covariance
var = variance
f = probability density function

Literature Cited

Anderson, T. F., D. S. Abrams, and E. A. Grens II, "Evaluation of Parameters for Nonlinear Thermodynamic Models," *AIChE J.*, **24**(1), 20 (1978).

Beckman, J. R., "Data Adjustment for Non-Reactive Batch or Steady-State Processes," *Chem. Eng. Commun.*, **15**, 357 (1982).

Britt, H. I., and R. H. Luecke, "The Estimation of Parameters in Nonlinear, Implicit Models," *Technometrics*, **15**(2), 233 (1973).

Crowe, C. M., "Reconciliation of Process Flow Rates by Matrix Projection. II: Nonlinear Case," *AIChE J.*, **32**(4), 616 (1986).

Ham, P. G., G. W. Cleaves, and J. K. Lawlor, "Operation Data Reconciliation: An Aid to Improved Plant Performance," *Proc. 10th World Petroleum Congr.*, Bucharest, **4**, 281 (1979).

Henley, E. J., and J. D. Seader, *Equilibrium-Stage Separation Operations in Chemical Engineering*, Wiley, New York, 270 (1981).

Hlaváček, V., "Analysis of a Complex Plant--Steady State and Transient Behavior," *Comput. Chem. Eng.*, 1, 75 (1977).

King, C. J., *Separation Processes*, McGraw-Hill, New York, 133 (1980).

Knepper, J. C., and J. W. Gorman, "Statistical Analysis of Constrained Data Sets," *AIChE J.*, **26**(2), 260 (1980).

Kuehn, D. R., and H. Davidson, "Computer Control. II: Mathematics of Control," *Chem. Eng. Prog.*, **57**(6), 44 (1961).

Lucia, A., and S. Macchietto, "New Approach to Approximation of Quantities Involving Physical Properties Derivatives in Equation-Oriented Process Design," *AIChE J.*, **29**(5), 705 (1983).

MacDonald, R., "Analyzing Plant Performance Data for a Flash Unit Operation Using Data Reconciliation Techniques to Estimate the Efficiency," M.S. Thesis, Univ. Kansas (1986).

Mah, R. S. H., "Design and Analysis of Process Performance Monitoring Systems," *Chem. Proc. Control II*, Proc. Eng. Found. Conf., Sea Island, GA, AIChE, 525 (1981).

Ricker, N. L., "Comparison of Methods for Nonlinear Parameter Estimation," *Ind. Eng. Chem. Process Des. Dev.*, **23**, 283 (1984).

Rosenberg, J., R. S. H. Mah, and C. Iordache, "Evaluation of Schemes for Detecting and Identifying Gross Errors in Process Data," *Ind. Eng. Chem. Res.*, **26**(3), 555 (1987).

Serth, R. W., and W. A. Heenan, "Gross Error Detection and Data Reconciliation in Steam-Metering Systems," *AIChE J.*, **32**(5), 733 (1986).

Stanley, G. M., and R. S. H. Mah, "Estimation of Flows and Temperatures in Process Networks," *AIChE J.*, **23**(5), 642 (1977).

Stephenson, G. R., and C. F. Shewchuk, "Reconciliation of Process Data with Process Simulation," *AIChE J.*, **32**(2), 247 (1986).

Manuscript received Mar. 6, 1987, and revision received June 29, 1987.