Gross Error Detection and Stage Efficiency Estimation in a Separation Process

R. W. Serth and B. Srikanth
Dept. of Chemical and Natural Gas Engineering, Texas A&M University-Kingsville, Kingsville, TX 78363
S. J. Maronga
Dept. of Chemical and Process Engineering, University of Dar-Es-Salaam, Dar-Es-Salaam, Tanzania

Accurate process models are required for optimization and control in chemical plants and petroleum refineries. These 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. Since the data contain random and systematic measurement errors, some of which may be large (gross errors), they must be reconciled to obtain reliable estimates of equipment parameters. The problem thus involves parameter estimation coupled with gross error detection and data reconciliation.

MacDonald and Howat (1988) studied the above problem for a single-stage flash distillation process. Their analysis was based on the definition of stage efficiency due to Hausen (King, 1980; Standart, 1965), which has some significant disadvantages in this context, as discussed below. In addition, they considered only data sets which contained no gross errors. The purpose of this article is to extend the above work by considering alternative definitions of stage efficiency and efficiency estimation in the presence of gross errors.

Flash System

The nonadiabatic, nonequilibrium single-stage flash system is shown in Figure 1. For a feed containing C components, the MESH equations for the system are listed in Table 1 in terms of the vaporization efficiency. Similar efficiency equations can be written using other definitions of stage efficiency given below.

For computational purposes, a feed containing four components [ethane (1), propane (2), propylene (3), and isobutene (4)] 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] are 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.

Stage Efficiency

Departures from equilibrium in flash systems are due to entrainment and mass-transfer effects (Hughes et al., 1953). While it is possible to design industrial separators to operate with virtually no liquid entrained in the vapor stream, some entrainment is often accepted for economic reasons. Nonequilibrium mass-transfer effects result from additional flashing that occurs due to the pressure drop in the separator. Although these two effects can be handled separately (Graham et al., 1973), they are typically combined for simplicity in the stage efficiency. The latter approach is followed here. In addition, it is assumed that the vapor and liquid streams leaving the separator are in thermal and mechanical equilibrium.

The stage efficiencies considered in this work are listed in Table 3. (Note that an asterisk denotes a quantity associated with an equilibrium flash.) The definitions are obtained by analogy with the corresponding efficiency definitions for multistage processes given in the references. These references also discuss various advantages and disadvantages of different efficiencies. The present discussion will be limited to the attributes that are relevant to the problem under consideration.

The Murphree and Hausen efficiencies share a common drawback in that for a C component system, only $C-1$ of the component efficiencies are independent (King, 1980). This result follows from the fact that both the $K_i^* x_i^*$ and the $K_i^+ x_i$ sum to unity. Thus, the number of MESH equations is reduced to $2C+3$, and one additional specification is required to define the system. For example, McDonald and Howat (1988) used either $T=T^*$ or $V=V^*$ for the additional specification. These conditions, however, are essentially arbitrary and in general will not correspond to actual process conditions.

The lack of independence among the Hausen efficiencies

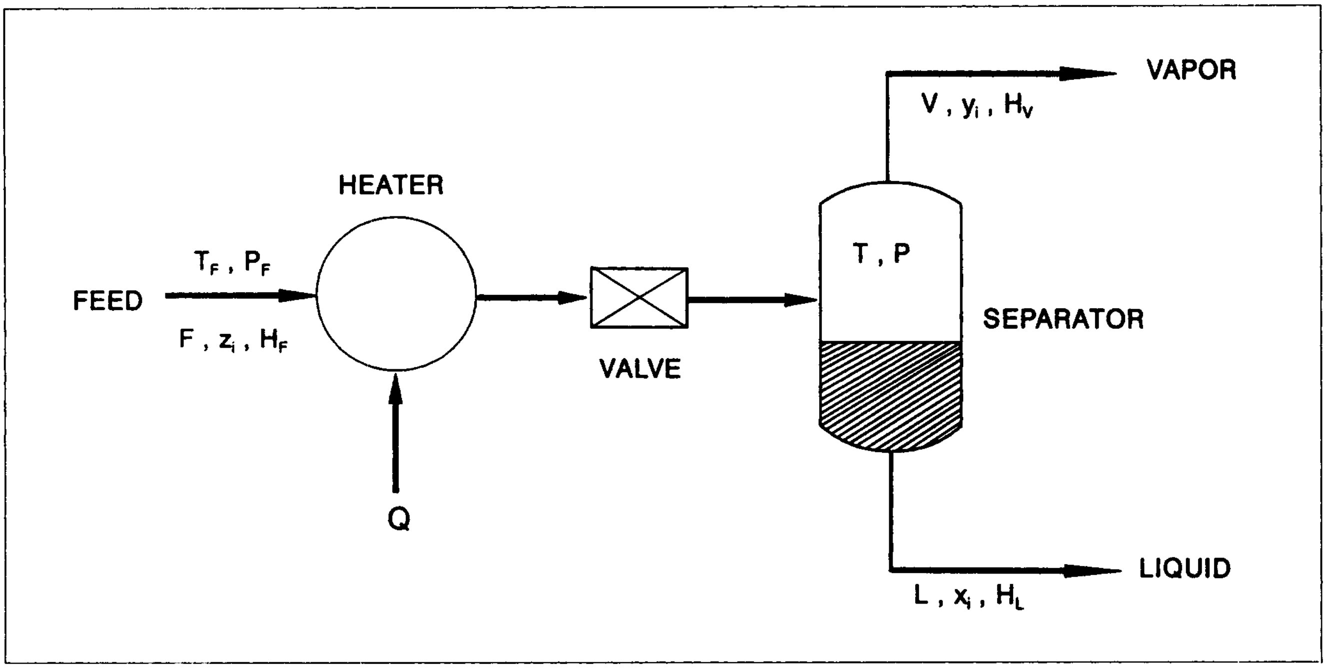

Figure 1. Flash distillation system.

can be overcome by basing them on component flow rates, rather than mole fractions (Standart, 1965). The resulting efficiencies, however, retain a second drawback associated with the Hausen efficiency, namely, the inclusion of equilibrium quantities in the definition. In the data reconciliation procedure (see below), derivatives of the process variables with respect to these equilibrium variables are required and must be obtained numerically at each iteration of the data reconciliation algorithm. Thus, a large number of equilibrium flash calculations are required. The resulting numerical difficulties have been well documented by MacDonald and Howat (1988).

Both of the above problems can be avoided by using the vaporization or modified Murphree efficiencies, with the latter based on either mole fractions or component flow rates. Although not significant from a computational standpoint, the vaporization efficiency has an advantage over the others in that the denominator can never be zero (Holland and McMahon, 1970).

The different efficiencies are, of course, interrelated, and the relationship between the vaporization and modified Murphree efficiencies is particularly straightforward:

$$\theta_i^V = \frac{1}{K_i} + \theta_i^{MM}\left(1 - \frac{1}{K_i}\right) \tag{1}$$

This is a special case of the equation given by Holland and McMahon (1970) for multistage distillation.

Table 1. MESH Equations for Single-Stage Flash Distillation Using Vaporization Efficiency

Type	Equation	No.
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	$\Sigma_i y_i - 1.0 = 0$	3
	$\Sigma_i x_i - 1.0 = 0$	
	$\Sigma_i z_i - 1.0 = 0$	
Enthalpy (H) Balance	$H_F F + Q - H_L L - H_v V = 0$	1
Total		$2C+4$

Data Reconciliation Procedure

The data reconciliation problem can be stated as follows (MacDonald and Howat, 1988):

$$\text{Min}(\hat{Z} - Z_m)^T R^{-1} (\hat{Z} - Z_m) \tag{2}$$

subject to

$$G(\hat{Z}, \hat{\chi}) = 0$$

where

Z_m = vector of measured values of process variables
$\hat{Z}$ = vector of adjusted values of process variables
$\hat{\chi}$ = vector of estimated values of process parameters
G = vector of MESH functions (Table 1)
R = covariance matrix of random measurement errors

In general, $\hat{\chi}$ may include values of unmeasured process variables as well as process parameters. The nonlinear regression problem represented by Eq. 2 can be solved for ($\hat{Z}$, $\hat{\chi}$) by iterative linear regression (Britt and Luecke, 1973), for example. If the random errors are normally distributed and systematic errors are absent, the solution will provide unbiased

Table 2. Specifications for Example Problem

Variable	Value
F	0.4536 kmol/s
z_1	0.15
z_2	0.35
z_3	0.30
T_F	316.67 K
P_F	3,447.37 kPa
P	1,723.68 kPa
Q	2,108.37 kJ/s

Table 3. Efficiency Definitions for Single-Stage Flash Distillation

Efficiency	Definition	Temp. at Which K_i is Evaluated*	References
Vaporization	$\theta_i^V = y_i / K_i x_i$	T	Holland (1981); Holland and McMahan (1970); King (1980)
Murphree	$\theta_i^M = \frac{y_i - x_i}{K_i^+ x_i - x_i}$	T^+	Holland (1981); Holland and McMahan (1970); King (1980)
Modified Murphree	$\theta_i^{MM} = \frac{y_i - x_i}{K_i x_i - x_i}$	T	Holland (1981); Holland and McMahan (1970); King (1980)
Hausen	$\theta_i^H = \frac{y_i - x_i}{K_i^* x_i^* - x_i^*}$	T^*	Standart (1965); King (1980)
Modified Murphree Based on Component Flow Rates	$\theta_i^{MC} = \frac{V y_i - L x_i}{V K_i x_i - L x_i}$	T	Standart (1965)
Hausen Based on Component Flow Rates	$\theta_i^{HC} = \frac{V y_i - L x_i}{V^* K_i^* x_i^* - L^* x_i^*}$	T^*	Standart (1965); King (1980)

*T = actual (nonequilibrium) flash temperature
T^* = equilibrium flash temperature
T^+ = hypothetical bubble-point temperature that satisfies $\Sigma_i K_i(T^+, P) x_i = 1.0$
K_i^* = $K_i(T^*, P)$
K_i^+ = $K_i(T^+, P)$

maximum likelihood estimates of the true values of process variables and parameters (MacDonald and Howat, 1988).

A gross error detection algorithm can be incorporated into the above procedure to eliminate systematic errors. The MIMT algorithm (Serth and Heenan, 1986; Serth et al., 1987) was used in this work. In this method, the regression problem (Eq. 2) is solved recursively. At each stage the residuals are tested for outliers, and the measurement corresponding to the most significant residual is deleted from the set of measured values. Iterations are continued until none of the residuals is statistically significant. Thus, in this procedure, the calculations for gross error detection, data reconciliation, and parameter estimation are all coupled.

Simulation Procedure

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-stage efficiency characterized the flash. Calculations were made for efficiencies of 75% and 100%.

Simulated measurement vectors were constructed by using pseudorandom number generators to add random and systematic errors to the true values. The relative standard deviations used for random errors were 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. The range of systematic error 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 (MacDonald and Howat, 1988). Since the simulated measurements were constructed using the same model as was used to analyze the data, no modeling error (including errors in thermodynamic data) was involved in the simulations.

Algorithm performance was determined by means of a Monte-Carlo procedure using sets of 1,000 runs per case studied. The measurement test was implemented at a nominal significance level of 95%, and all calculations were performed on an IBM 9370 computer using double-precision arithmetic.

Results and Discussion

Calculations were made using the vaporization and modified Murphree efficiencies, with the latter based both on mole fractions and component flow rates. For comparison, calculations were also made for an equilibrium flash, which does not involve parameter estimation. Results for the equilibrium flash are summarized in Table 4. It can be seen that the error detection algorithm performed well on this system, especially with respect to errors in mole fraction and flash temperature measurements. It should be noted that errors in the feed temperature, T_F, and heat rate, Q, cannot be resolved because each of these variables occurs just once in the MESH equations, in different terms of the H equation. Hence, the results for "all variables"

Table 4. Results for Equilibrium Flash

No. of Systematic Errors	0	1	2	3
Flow Rates				
Errors detected, %	—	95	89	83
Erroneous identifications, %	2.6	9.1	17	20
Mole Fractions				
Errors detected, %	—	99	99	99
Erroneous identifications, %	3.8	8.1	3.9	5.3
Flash Temperature				
Errors detected, %	—	100	100	98
Erroneous identifications, %	0.2	1.9	7.2	1.9
All Variables				
Errors detected, %	—	93	92	90
Erroneous identifications, %	10	16	12	12

Table 5. Results for Nonequilibrium Flash: 75% Vaporization Efficiency

No. of Systematic Errors	0	1	2	3
Flow Rates				
Errors detected, %	—	92	86	78
Erroneous identifications, %	1.3	9.1	14	22
Mole Fractions				
Errors detected, %	—	99	99	98
Erroneous identifications, %	4.5	8.1	5.6	7.4
Flash Temperature				
Errors detected, %	—	89	86	84
Erroneous identifications, %	0.7	12	18	23
All Variables				
Errors detected, %	—	92	91	88
Erroneous identifications, %	6.6	14	12	14
Mean Efficiency, %	75.0	75.1	75.7	75.3
Maximum Efficiency, %	79.7	86.0	126.6	145.5
Minimum Efficiency, %	68.6	63.2	46.5	46.8
Standard Deviation, %	1.3	1.7	4.3	5.9

Table 7. Results for Nonequilibrium Flash: 75% Modified Murphree Efficiency Based on Component Flow Rates

No. of Systematic Errors	0	1	2	3
Flow Rates				
Errors detected, %	—	96	90	82
Erroneous identifications, %	1.9	9.2	17	25
Mole Fractions				
Errors detected, %	—	99	98	97
Erroneous identifications, %	4.6	9.9	5.9	8.6
Flash Temperature				
Errors detected, %	—	96	96	88
Erroneous identifications, %	0.6	19	13	23
All Variables				
Errors detected, %	—	93	91	88
Erroneous identifications, %	7.7	17	13	17
Mean Efficiency, %	74.9	75.0	75.3	75.2
Maximum Efficiency, %	82.2	85.2	140.4	199.1
Minimum Efficiency, %	66.4	59.1	28.0	20.8
Standard Deviation, %	2.1	2.5	4.6	8.9

tend to understate the effectiveness of the error detection algorithm to some extent. They are nonetheless useful for comparative purposes.

Results for a 75% efficient flash are presented in Tables 5-8. Results for each efficiency definition are given in Tables 5-7, respectively, while Table 8 compares results for the three efficiencies on runs with two systematic errors. The results obtained for a flash efficiency of 100% were similar and are therefore omitted. The overall performance of the error detection algorithm on the nonequilibrium flash was comparable to, but somewhat poorer than, that on the equilibrium flash. This behavior is expected because the estimation of stage efficiency results in the loss of one degree of freedom for error detection. The most striking difference in performance among the three efficiencies was with respect to flash temperature errors, where use of θ^{MM} resulted in extremely effective error detection. Although the reason for this behavior is not apparent, the effect is very reproducible and is not an artifact of the simulation procedure.

The calculated mean efficiency was close to the true value of 75% in all cases. As expected, the sample range and estimated standard deviation of the efficiency distribution increased monotonically with number of gross errors. In all cases, use of θ^V gave the smallest standard deviation.

Table 6. Results for Nonequilibrium Flash: 75% Modified Murphree Efficiency

No. of Systematic Errors	0	1	2	3
Flow Rates				
Errors detected, %	—	94	88	79
Erroneous identifications, %	1.4	10	17	25
Mole Fractions				
Errors detected, %	—	98	97	95
Erroneous identifications, %	4.3	8.3	5.4	7.8
Flash Temperature				
Errors detected, %	—	100	100	100
Erroneous identifications, %	0.2	5.8	5.3	5.9
All Variables				
Errors detected, %	—	92	91	88
Erroneous identifications, %	6.2	16	13	14
Mean Efficiency, %	74.9	74.8	75.3	75.2
Maximum Efficiency, %	85.9	92.0	153.7	154.1
Minimum Efficiency, %	65.5	55.9	39.8	10.8
Standard Deviation, %	2.4	2.9	5.1	8.3

The increased spread of the efficiency distribution in runs with systematic errors is due to gross errors that were not eliminated by the MIMT algorithm. Occasionally, the algorithm failed to identify any of the gross errors in a data set. In these cases, large adjustments were incorrectly made in other measurements to compensate for the gross errors. This resulted in a reconciled data set that was worse than the original data. The maximum and minimum efficiencies listed in Tables 5-7 indicate the worst-case performance of the algorithm in this respect. Performance can be controlled to some extent, because the MIMT algorithm allows the user to specify lower and upper bounds on all variables, including stage efficiency (Serth and Heenan, 1986). In the present study, very loose bounds (0.05 and three times the true value for efficiency; zero and four times the true value for other variables) were specified to observe the algorithm's performance without undue forcing. In actual practice, tighter bounds would normally be employed or, at the very least, unusually high or low values of efficiency would not be accepted without further analysis, which would include rerunning the algorithm with tighter bounds.

For runs without systematic errors, the results in Tables 5-7 are similar to those obtained by MacDonald and Howat (1988). Somewhat smaller ranges and standard deviations were obtained for the estimated flash efficiency in this work, but these differences are due primarily to different assumptions of measurement precision in the two studies.

It is difficult to compare reconciled vs. true values of individual process variables in a concise format for simulations involving gross errors. However, some general observations can be made. When the MIMT Algorithm correctly eliminated all gross errors from a data set, the reconciled data, including the estimated values for the abandoned measurements, were in excellent agreement with the true values. Results for these cases were comparable to those for runs without gross errors. When the algorithm failed to eliminate all gross errors from a data set, significant discrepancies between reconciled and true values sometimes occurred, with the extent of the discrepancies depending on the number, size and location of un-

Table 8. Comparison of Results for Runs with Two Gross Errors Using Different Efficiencies

Efficiency Type	Vaporization	Modified Murphree (Mole Fractions)	Modified Murphree (Component Flow Rates)
Flow Rates			
Errors detected, %	86	88	90
Erroneous identifications, %	14	17	17
Mole Fractions			
Errors detected, %	99	97	98
Erroneous identifications, %	5.6	5.4	5.9
Flash Temperature			
Errors detected, %	86	100	96
Erroneous identifications, %	18	5.3	13
All Variables			
Errors detected, %	91	91	91
Erroneous identifications, %	12	13	13
Mean Efficiency %	75.7	75.3	75.3
Maximum Efficiency, %	126.6	153.7	140.4
Minimum Efficiency, %	46.5	39.8	28.0
Standard Deviation, %	4.3	5.1	4.6

detected errors. Undetected errors toward the smaller end of the size range were frequently reconciled quite satisfactorily.

The fact that the calculated mean efficiency was close to the true value in all cases does not, in itself, demonstrate the effectiveness of the error detection algorithm in producing unbiased estimates. Because both positive and negative systematic errors were generated randomly, nearly unbiased estimates of the efficiency were obtained even when the error detection algorithm was bypassed. To investigate the effectiveness of the algorithm in this respect, additional runs were made with one-sided systematic errors in selected variables. The results in Table 9 show that the algorithm was quite effective in reducing the bias in the estimated efficiency.

Conclusions

Use of the vaporization and modified Murphree efficiencies can be recommended to avoid problems associated with the lack of independence among the component efficiencies and the computational difficulties involved with introducing equilibrium quantities into the MESH equations. Effective error detection and efficiency estimation can be achieved with these efficiencies. In this study, the best error detection was achieved using the modified Murphree efficiency based on mole fractions.

Table 9. Results with One-Sided Systematic Errors: 75% Modified Murphree Efficiency

Variables with Systematic Errors	Mean Efficiency, %		Standard Deviation, %	
	Without Error Detection	With Error Detection	Without Error Detection	With Error Detection
Flash Temp.	80.6	74.9	3.3	2.6
Liquid Flow Rate	73.6	75.0	2.8	2.5
Feed Mole Fraction Propane	77.7	74.9	2.7	2.7
Flash Temp. and Liquid Flow Rate	83.3	74.9	3.2	2.4
Flash Temp. and Feed Mole Fraction Propane	80.0	74.9	3.5	2.3
Liquid Flow Rate and Feed Mole Fraction Propane	71.0	74.9	2.9	2.3

Notation

C = number of components
F = feed flow rate, kmol/s
G = vector of MESH functions
H = enthalpy, kJ/kmol
K = K value
L = liquid flow rate, kmol/s
P = pressure, kPa
Q = heat rate, kJ/s
R = covariance matrix of random measurement errors
T = temperature, K
V = vapor flow rate, kmol/s
x = liquid mole fraction
y = vapor mole fraction
z = feed mole fraction
Z = vector of process variables

Greek letters

θ = flash efficiency
χ = vector of process parameters

Subscripts

F = feed stream
i = component index
L = liquid stream
m = measured value
V = vapor stream

Superscripts

H = Hausen efficiency
HC = Hausen efficiency based on component flow rates
M = Murphree efficiency
MC = modified Murphree efficiency based on component flow rates
MM = modified Murphree efficiency
V = vaporization efficiency
$*$ = equilibrium flash value
$+$ = Murphree Temperature Convention
$\hat{}$ = adjusted (reconciled) value

Literature Cited

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

Graham, J. P., J. W. Fulton, M. S. Kuk, and C. D. Holland, "Predictive Methods for the Determination of Vaporization Efficiencies," *Chem. Eng. Sci.*, **28**, 473 (1973).

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

Holland, C. D., and K. S. McMahon, "Comparison of Vaporization Efficiencies with Murphree-type Efficiencies in Distillation," *Chem. Eng. Sci.*, **25**, 431 (1970).

Hughes, R. R., H. D. Evans, and C. V. Sternling, "Flash Vaporization," *Chem. Eng. Prog.*, **49**, 78 (1953).

King, C. J., *Separation Processes*, 2nd ed., McGraw-Hill, New York (1980).

MacDonald, R. J., and C. S. Howat, "Data Reconciliation and Parameter Estimation in Plant Performance Analysis," *AIChE J.*, **34**, 1 (1988).

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

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

Standart, G., "Studies on Distillation: V. Generalized Definition of a Theoretical Plate or Stage of Contacting Equipment," *Chem. Eng. Sci.*, **20**, 611 (1965).

Manuscript received Sept. 28, 1992, and revision received Feb. 8, 1993.