PROCESS ENGINEERING AND DESIGN

Estimation of Distillation Compositions from Multiple Temperature Measurements Using Partial-Least-Squares Regression

Thor Mejdell and Sigurd Skogestad*

Chemical Engineering, University of Trondheim, NTH, N-7034 Trondheim, Norway

This paper addresses the use of temperature measurements to estimate product compositions in distillation columns. A simple linear multivariate calibration procedure based on steady-state data is used, which requires minimal modeling effort. It is found that these principal-component-regression (PCR) and partial-least-squares (PLS) estimators perform well, even for multicomponent mixtures, pressure variations, and nonlinearity caused by changes in operating conditions. The use of weighting functions, additional factors, and logarithmic transformations improve the estimates and counteract nonlinearities, provided there is not too much noise on the temperatures. In the paper we also compare more generally regression methods based on singular-value decomposition (SVD; generalized least squares), PCR, and PLS.

1. Introduction

Reliable and accurate measurement of product compositions is one of the main difficulties in distillation column control. Most product analyzers, like gas chromatographs, suffer from large measurement delays and high investment and maintenance costs. The overall measurement delay is typically 10 to 20 min. This imposes severe limitations on achievable control performance. However, the reliability of the analyzers is perhaps their weakest point, and this also results in high maintenance costs in terms of manpower and expensive back-up systems. One employee per every three GC analyzers is common in industry. The most popular alternative to analyzers is single-temperature control, i.e., control of a given tray temperature. Temperature measurements are reliable and inexpensive and have negligible measurement delays. However, they are not accurate indicators of product composition. Nevertheless, in most cases temperature control is preferred. For example, Kister (1990) recommends using temperature control unless the difference in boiling point between the key components is very small or there are substantial economical benefits in keeping tight control of the product compositions.

This paper addresses two-product columns where we make a split between two defined key components, denoted the light key (LK) and heavy key (HK) component. We can make two independent specifications to define the split between these components. The distribution of the remaining components, denoted the off-key or nonkey (NK) components, may not be specified. In this paper the specifications are chosen as the product mole fractions, y'_D and x'_B, of the light component on a pseudobinary basis.

1.1. Problems with Single-Temperature Control. Figure 1 displays typical steady-state profiles for the binary example column (The column is described in section 2.1). For this binary mixture at constant pressure, the temperature at the column end is an exact indicator of composition. However, as seen from Figure 1, the temperature variation is very small at the column ends and may be difficult to distinguish from measurement noise. Therefore, temperatures further removed from the ends are

preferred (Nisenfeld and Seeman, 1981). However, even with this precaution, the use of a single temperature to indicate product composition is generally not reliable because of the following.

(a) Even for binary mixtures the relationship between a temperature inside the column and a product composition depends on the feed composition and also on the product composition at the other end of the column.

(b) For multicomponent mixtures the presence of off-key components implies that even at the column ends temperature is not an exact indicator of composition. The effect of variations in off-key components is largest near the feed and at the column ends.

(c) Associated with the measurement and data treatment device are (1) random high-frequency noise that will usually have the same magnitude on all trays and (2) low-frequency measurement offsets, for example, due to fouling or changes in the ice-point temperature compensation.

(d) Flow pulses and improper mixing on the trays cause temperature variations which have a peak at intermediate frequencies and are largest in column sections with large temperature gradients.

(e) Pressure changes cause temperature variations. The total pressure has a similar effect on all temperatures, while the pressure drop, which is influenced by the throughput, has its largest temperature effect near the bottom. Varying liquid holdups and tray performance may give additional local pressure variations.

Some measures may be taken to counteract these problems. The high-frequency noise may easily be filtered. Pressure variations may be compensated for using pressure measurements or using differential temperatures. The effect of nonkey components may be reduced by locating the temperature measurement in regions of the column where their composition is nearly constant (Rademaker et. al., 1975).

However, some of these problems may not be corrected. In particular, keeping a temperature constant on a tray some distance away will not keep the product composition constant. For example, consider Figure 2, where it is shown that the temperatures in the lower part of the
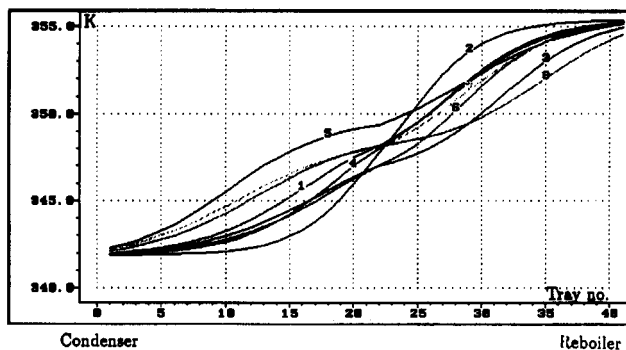
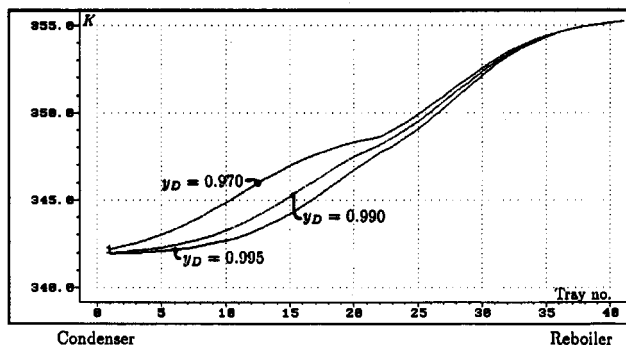**Figure 1.** Steady-state profiles of the example column (binary mixture).

Figure 2. Effect of changes in top composition on the temperature profile. The feed and bottom compositions are constant.
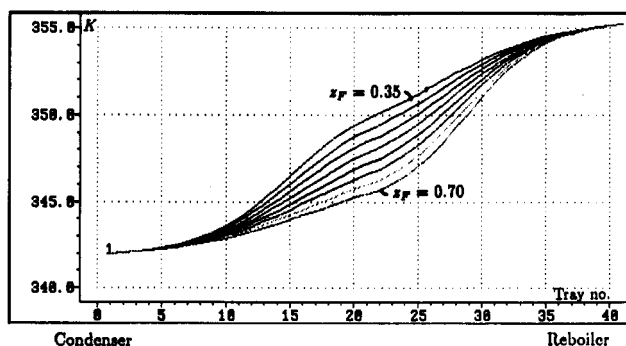

Figure 3. Effect of changes in feed composition on temperature profile. Top and bottom compositions are constant.

column are affected by changes in top composition, and Figure 3, where it is shown how the temperatures are affected by different feed compositions.

1.2. Multiple Temperatures. One solution to these problems is to use multiple temperature measurements to infer the product composition. There has recently been reports from industry on successful implementations of somewhat ad hoc approaches. Whitehead and Parnis (1987) used a weighted temperature average of differential temperatures on a C_2 splitter. Bozenhardt (1988) used multiple temperatures to track the maximum temperature difference between two trays in an alcohol/water/ether column. He found the position of this maximum difference to be strongly correlated to the product composition.

A more rigorous approach is to use a temperature-based composition estimator. In another paper (Mejdell and Skogestad, 1991a), we compared rigorously three different estimators using *linear* data for the binary example column. These estimators were the dynamic Kalman–Bucy filter (Kalman and Bucy, 1961), the static Brosilow inferential estimator (Weber and Brosilow, 1972; Joseph and Brosilow, 1978), and the static principal-component-regression (PCR) estimator. It was found that for feedback

control the static PCR estimator performed almost as well as the Kalman filter. The reason is that the temperatures and the compositions have similar dynamic responses. The Brosilow estimator was very sensitive to model error for this ill-conditioned plant with large relative gain array (RGA) values. Mejdell and Skogestad (1991a) therefore recommended using the simple regression estimator, which is obtained simply by considering corresponding values of temperatures and composition. Such data sets are most easily obtained from simulations but may be complemented with experimental data.

In this paper we consider some additional aspects of using regression estimators for composition estimation, including nonlinearities in the column, pressure variations, multicomponent mixtures, and measurement noise. We discuss different ways of handling these problems, such as variable transformations and scaling.

In this paper, temperature measurements on all trays are used. This is not because it is strictly necessary but to emphasize that all available measurements should be used and to exclude the influence of measurement selection, which otherwise would bias the results. Note that other measurements, such as flow rates, are not used. This is based on Mejdell and Skogestad (1991a) who found that the static estimate was not significantly improved by adding such measurements and that the dynamic estimate became worse.

The results in this paper are based on simulations only. In a companion paper (Mejdell and Skogestad, 1991b) we present results from a pilot-scale distillation column. These results demonstrate that simple static regression estimators, such as PLS, also work in practice.

2. Problem Definition

The following problem is treated in this paper: Given the temperatures on all trays in a distillation column, find a good *static* estimator for the product compositions. The example column described below is used to illustrate the main issues.

2.1. Example Column. The column has 40 theoretical stages (including the reboiler) and a total condenser. The feed stream enters the column at stage 20 as saturated liquid. Two cases are considered: (1) binary mixture with constant relative volatility of 1.5; (2) multicomponent mixture consisting of one heavy and one light nonkey in addition to the binary components in case 1. The column in case 1 is "column A" (Skogestad and Morari, 1988b), which is the same example as used by Mejdell and Skogestad (1991a). Data for the column and for the mixtures are given in Table I. The difference in boiling point between the two pure key components is only 13 °C. This is approximately the lower limit for what is usually recommended when a single-temperature control is used (Riggs, 1990). Typical temperature profiles for the binary case are shown in Figures 1–4.

2.2. The Estimation Problem. Consider the case with binary mixture, constant pressure, and feed and reflux as saturated liquid. Then specifying different values of feed composition z_F, distillate composition y_D, and bottom product x_B yields unique steady-state profiles of the 41 temperatures θ. This may be expressed as

$$\theta^{41\times1} = f(z_F, y_D, x_B) \tag{1}$$

We want to find the inverse relation

$$\hat{y}^{2\times1} = \begin{pmatrix} \hat{y}_D \\ \hat{x}_B \end{pmatrix} = g(\theta^{41\times1}) \tag{2}$$

Table I. Data for Distillation Column Example[a]

A. Binary Mixture

i	compt	z_F	y_D	x_B	$\alpha_{i,i+1}$	T^b/K	Antoine parameters		
							A	B	C
1	LK	0.500	0.99	0.01	1.5	341.9	15.836 60	2697.55	-48.78
2	HK	0.500	0.01	0.99		355.4	15.431 13	2697.55	-48.78

B. Multicomponent Mixture

i	compt	z_F	y_D	x_B	$\alpha_{i,i+1}$	T^b/K	Antoine parameters		
							A	B	C
1	LNK	0.050	0.125	0.000	2.0	321.4	16.529 75	2697.55	-48.78
2	LK	0.350	0.866	0.006	1.5	341.9	15.836 60	2697.55	-48.78
3	HK	0.350	0.009	0.577	2.0	355.4	15.431 13	2697.55	-48.78
4	HNK	0.250	0.000	0.417		381.6	14.737 99	3697.55	-48.78

[a] Conditions and parameters: $\ln P(\text{mmHg}) = A - B/((\theta(\text{K})) + C)$; feed is saturated liquid; constant molar flows; ideal VLE using Raoult's law; constant pressure, 1 atm; $N = 40$, $N_F = 21$ (theoretical trays); holdup on all trays, $M_i/F = 0.5$ min; flow dynamics are neglected.
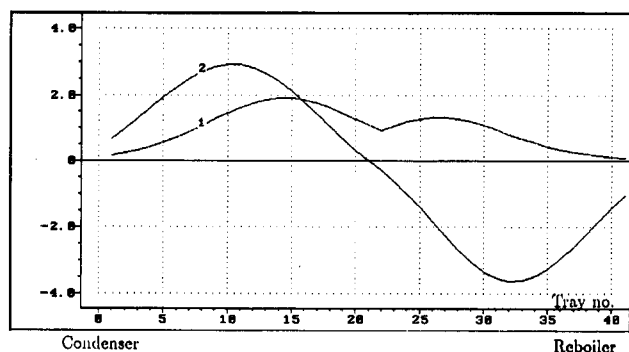

Figure 4. Changes in temperature profile caused by the following: (1) increase in external flows ($\Delta D/F = 0.01$), V is kept constant; (2) decrease in internal flows ($\Delta L/F = \Delta V/F = -1$). D and B are kept constant.

The simplified linear estimator used in this paper may be written (in terms of deviation variables) as

$$\hat{y} = \mathbf{K}\theta \tag{3}$$

where the matrix $\mathbf{K}$ is of dimension 2×41. The problem is to find optimal values of these 82 parameters. These parameters cannot be determined independently because (1) the temperatures are not independent and (2) there is not enough degrees of freedom in the excitations. The first fact is illustrated by Figure 1, where we see that temperatures close to each other change in nearly the same way. The second fact is illustrated by eq 1, where we see that there are only three degrees of freedom in the excitations, and in a nonlinear sense the temperatures have only three different ways to vary. However, note that the number of directions in the *linear* temperature space may be larger than three if large perturbations are used in z_F, y_D, and x_B. Pressure variations and off-key components also increase the degrees of freedom, but they are for the moment assumed to be constant.

The estimation problem may be divided in two steps.

(1) Temperature data are reduced into k latent variables $\mathbf{t}$ (also denoted factors):

$$\mathbf{t}^{k \times 1} = \mathbf{p}(\theta^{41 \times 1}) \tag{4}$$

(2) An estimator is obtained by finding a relationship $\mathbf{g}$ between the latent variables and the product composition.

$$\begin{pmatrix} \hat{y}_D \\ \hat{x}_B \end{pmatrix} = \mathbf{g}(\mathbf{t}^{k \times 1}) \tag{5}$$

The key question is how to find suitable latent variables, in order to make the second regression step easy. Pref-

erably the latent variables should be independent, and they should contain all the original information relevant for estimating the compositions.

(a) The simplest approach is to delete measurements and select only a few which are mutually independent. Then the problem of optimal measurement selection becomes a key issue. A lot of articles have been published on this subject, for example, Joseph and Brosilow (1978), Morari and Stephanopoulos (1980), and Moore (1987).

(b) Using unknown disturbances as latent variables is a procedure that has been proposed by Brosilow and co-workers and employed in their inferential control, e.g. Weber and Brosilow (1972) and Joseph and Brosilow (1978).

(c) Gilles and Retzbach (1980) and later Marquardt (1989) used some geometric shape factors of the temperature profile as latent variables. Here the first factor is the location of the steepest temperature gradient.

(d) The principal components (PCR method) or the partial-least-squares factors (PLS method) can be used as latent variables. These methods have been introduced for distillation columns by Mejdell and Skogestad (1989) and are further outlined in the next section.

The simplest of the above methods is of course to reduce the number of measurements. However, this is not optimal, because additional measurements will improve the estimate as they contain more independent information (at least for few measurements), reduce the effect of measurement noise, make the estimator less sensitive to measurement selection and changes in operating conditions, and better capture the effect of nonlinearities.

We have briefly investigated Marquardt's method and found it to perform very well with perfect measurements, but it was sensitive to measurement noise for our column. It might perform better for distillation columns with sharp profiles, for which it was originally developed.

The Brosilow estimator was studied in detail by Mejdell and Skogestad (1991a). We found that it performed poorly for the binary distillation example. The reason is that for ill-conditioned plants, like distillation columns, the estimate is very sensitive to small errors. This sensitivity has indeed been pointed out also by Brosilow and co-workers.

Mejdell and Skogestad (1991a) studied the static PCR estimator for the example column and found that it performed almost as well as an optimally tuned Kalman filter. In this paper we shall use the static regression estimators, PLS and PCR.

2.3. Multicomponent Mixture. The multicomponent mixture is obtained by extending the original binary with one light nonkey (LNK) and one heavy nonkey (HNK) component (See Table I). The control objective for the separation is still the split with respect to the key com-
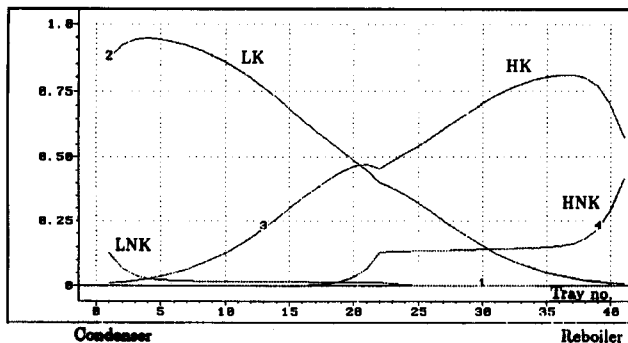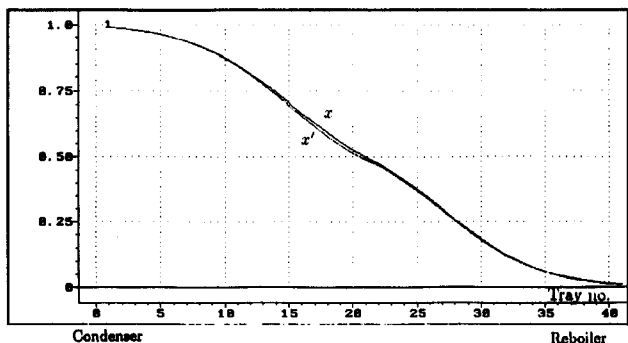
Figure 5. Concentration profiles of the multicomponent mixture.

Figure 6. Concentration profiles for binary mole fraction (x) and multicomponent pseudobinary mole fraction (x').

ponents. On a given tray x' denotes the pseudobinary mole fraction (based on key components) of the light key component, i.e.

$$x' = \frac{x_L}{x_L + x_H} \tag{6}$$

The product specification for both the binary and multicomponent case is $y'_D = 0.99$ and $x'_B = 0.01$.

The composition profile for the multicomponent mixture at the nominal operating point is displayed in Figure 5. The off-key components are almost constant in the column except at the column ends and around the feed tray. The composition profile on a pseudobinary basis, x', is compared with that of the binary example column, x, in Figure 6, and the corresponding temperature profiles are shown in Figure 7. The concentration profiles are almost identical, whereas the temperature profiles are quite different.

2.4. Use of Transformed Variables. The composition and temperature profiles are nonlinear functions of the operating variables. One way to deal with nonlinearity is to find other variables (factors) which can capture the nonlinearity. The Marquardt (1989) method is one such method. A simpler method is to use nonlinear transformations on each variable. Logarithmic transformation of the product compositions have been proposed by several authors (e.g., Joseph et al., 1976; Skogestad and Morari, 1988a) as an effective way to linearize the dynamic and static *response* (with L, V, F, z_F, etc. as independent variables). For binary mixtures the following relationships are defined:

$$Y_D = \ln (1 - y_D); \quad X_B = \ln x_B \tag{7}$$

These transformations apply also for multicomponent mixtures if pseudobinary compositions are used. The composition *profile* (with tray number as an independent variable) may also be linearized using similar transformations. Ryskamp (1981) plotted the compositions on a probability scale. It is also common to use the logarithm
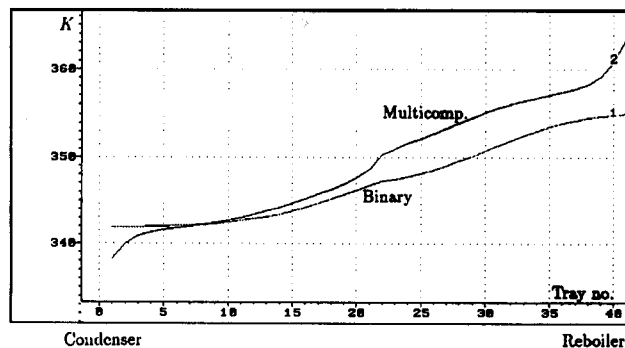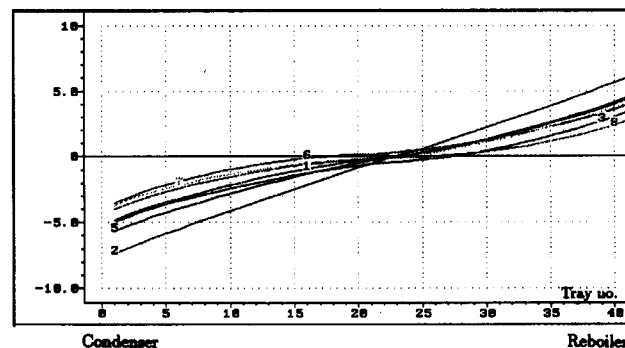

Figure 7. Temperature profiles for the binary and multicomponent cases.

Figure 8. Temperature profiles in terms of logarithmic transformed temperatures, L_T.

of the separation parameter on each tray (e.g., PROCESS, 1981)

$$X = \ln \left(\frac{x_L}{x_H}\right) \tag{8}$$

Here x_L and x_H are mole fractions of light and heavy components on a given tray (the subscript i is omitted to simplify notation). Note that since most columns have $y_D = x_{DL} \approx 1$ and $1 - x_B = x_{BH} \approx 1$, (8) may be viewed as a generalization of (7). In Appendixes 1 and 2, we show that this transformation in addition to linearizing the *profile* also linearizes the *dynamic response*.

Temperature is often a nearly linear function of composition. On the basis of eq 8, we therefore propose to use the following transformation to linearize the temperature response and profile (Appendix 1)

$$L_T = \ln \left(\frac{\theta - T_L^b}{T_H^b - \theta}\right) \tag{9}$$

Here θ is the tray temperature, and T_L^b and T_H^b are the boiling temperatures of pure light and heavy components, respectively. For our example column, this results in a nearly linear profile, as shown in Figure 8 (compare with the unscaled profiles in Figure 1). A column with pinch zones around the feed will not have a linear profile. Note that for the binary case $X \approx -L_T$.

Instead of using boiling temperatures, one may use the transformation

$$L_\theta = \ln \left(\frac{\theta - \theta_L}{\theta_H - \theta}\right) \tag{10}$$

where θ_L and θ_H are some *reference temperatures* (usually measured) at the top and the bottom of the column, respectively. For binary mixtures, one may use the temperature at the column end, which is very close to the boiling temperature. Note that the transformation L_T may

Table II. Specifications Used in Simulations To Obtain Static Temperature Profiles

z_F	y_D	x_B
0.5000	0.990 00	0.010 00
0.5375	0.991 30	0.026 20
0.4250	0.973 80	0.015 10
0.5250	0.970 00	0.013 20
0.4125	0.980 10	0.005 80
0.6000	0.984 90	0.004 40
0.5125	0.994 20	0.006 60
0.5500	0.982 70	0.007 60
0.4875	0.996 20	0.018 90
0.4750	0.995 60	0.008 70
0.5625	0.993 40	0.011 50
0.4625	0.977 20	0.030 00
0.4375	0.995 00	0.003 80
0.4500	0.992 40	0.017 30
0.5750	0.986 80	0.022 80
0.5875	0.988 50	0.005 00
0.4000	0.970 00	0.030 00
0.4000	0.970 00	0.003 33
0.4000	0.996 67	0.030 00
0.4000	0.996 67	0.003 33
0.6000	0.970 00	0.030 00
0.6000	0.970 00	0.003 33
0.6000	0.996 67	0.030 00
0.6000	0.996 67	0.003 33
0.4500	0.982 68	0.017 32
0.4500	0.982 68	0.005 77
0.4500	0.994 23	0.017 32
0.4500	0.994 23	0.005 77
0.5500	0.982 68	0.017 32
0.500	0.982 68	0.005 77
0.5500	0.994 23	0.017 32
0.5500	0.994 23	0.005 77

be used even when we have only one temperature measurement, θ, whereas the transformation L_θ usually needs in addition two measured reference temperatures.

For multicomponent mixtures, one must select the reference temperature at a location some distance away from the ends where the off-key compositions are almost constant from one tray to another (see Figure 5). To avoid taking logarithms of negative numbers, the temperatures between the reference locations and column ends must be treated separately, for instance by using the absolute value of the temperature differences in eq 10. To avoid large effects of noise on the temperatures closest to the reference temperatures, one should also specify a lower permitted limit on the difference termperatures in eqs 9 and 10.

Using reference temperatures instead of boiling point temperatures also provides pressure compensation of the temperature measurements, as well as off-key compensation in the multicomponent case.

2.5. Calibration Set for PCR and PLS. The calibration set used in this paper consists of 32 different simulation runs. The outputs x_B and y_D and the feed composition z_F (disturbance) were specified, and the corresponding steady-state temperature profiles were obtained using a nonlinear column model. The 32 values are listed in Table II. The data were spread with equal distances around z_F, x_B, and y_D. The first run is the nominal operating point; the next 15 runs were randomly chosen. The last 16 runs were selected by a 2-composite design in 4 levels. The 32 runs correspond to maximum variations in column end temperatures of about 0.4 °C and of interior temperatures of about 5 °C.

Different versions of the 32-calibration set was made, which included total pressure variations of ±0.1 atm spread randomly on the various runs, normal distributed random noise of magnitude 0.1 or 0.2 °C on all temperatures, and nonkey components. The pseudobinary compositions were the same as those in the binary case (Table II); but in

addition the feed mole fraction of the light nonkey component was varied between 0.025 and 0.075, and that of the heavy nonkey, between 0.15 and 0.35. In this case all temperatures were rounded to one decimal, which may be viewed as a (colored) noise source.

2.6. Evaluation Criteria. The main criterion used to evaluate the performance of the estimators is the mean square error of prediction (MSEP) (Martens and Næs, 1989, p 250), which is a measure of the expected error of *future* predictions. An estimate of MSEP was performed by a cross-validation procedure as follows: The 32 calibration runs were divided into seven groups. Then the calibration procedure (i.e., finding the estimator $\mathbf{K}$) was performed seven times, each time with six groups used for calibration and one for testing. The mean square error of all test predictions was then obtained from

$$\text{MSEP}(k) = \frac{1}{32} \sum_{i=1}^{32} (\hat{y}_i(k) - y_i)^2 \qquad (11)$$

Here k is the number of factors (or principal components) used in the calibration, and y_i is y_D or x_B for the ith test run. MSEP will generally *not* go to zero when the number of factors increases, since the test runs are independent from the ones used in the calibration. Rather, MSEP will increase when factors containing only noise are included.

We then compute the explained prediction variance in percent

$$\text{EPV}(k) = 100 \left(1 - \frac{\text{MSEP}(k)}{\text{MSEP}(0)} \right) \qquad (12)$$

MSEP(0) is approximately the variance in the calibration data.

3. Data Treatment and Multivariate Regression

We want to estimate p outputs ($\mathbf{y}$) from q known variables (θ). Multivariate regression is a linear statistical technique for obtaining the matrix $\mathbf{K}$ in

$$\hat{\mathbf{y}} = \mathbf{K}\theta \qquad (13)$$

Both vectors $\mathbf{y}$ and θ are centered (deviation variables), so there is no constant term. We obtain a "training set" consisting of n calibration runs of corresponding values of $\mathbf{y}$ and θ and place these as rows in the matrices $\mathbf{Y}^{n \times p}$ and $\Theta^{n \times q}$. (It might seem more reasonable to place $\mathbf{y}$ and θ as columns in the matrices, but we shall here use the standard notation in statistics.) We have the desired relation

$$\mathbf{Y} = \Theta \mathbf{K}^T \qquad (14)$$

The general least-squares solution for $\mathbf{K}$ is (Strang, 1980, p 139)

$$\mathbf{K}^T = \Theta^\dagger \mathbf{Y} \qquad (15)$$

The pseudoinverse $\Theta^\dagger$ is most easily obtained using the SVD of Θ. In addition to minimizing $(y - \hat{y})^2$ this solution minimizes the norm of $\mathbf{K}$.

3.1. Singular-Value Decomposition (SVD). The SVD of Θ may be written

$$\Theta = \mathbf{U} \Sigma \mathbf{V}^T \qquad (16)$$

or as a sum of m rank 1 - matrices (of decreasing importance)

$$\Theta = \mathbf{u}_1 \sigma_1 \mathbf{v}_1^T + \mathbf{u}_2 \sigma_2 \mathbf{v}_2^T + \dots + \mathbf{u}_m \sigma_m \mathbf{v}_m^T \qquad (17)$$

where m is the rank of Θ. If $m < \min(n, q)$, then both columns and rows in Θ are linearly dependent (the matrix is singular). The column vectors of $\mathbf{U}$, $\mathbf{u}_1 \dots \mathbf{u}_m$, are the orthonormal eigenvectors of $(\Theta \Theta^T)$, and the column vectors of $\mathbf{V}$, $\mathbf{v}_1 \dots \mathbf{v}_m$, are the eigenvectors of $(\Theta^T \Theta)$. σ_i are the

singular values; $\sigma_1{}^2$ is the largest eigenvalue of $(\Theta\Theta^T)$ (or $(\Theta^T\Theta)$), $\sigma_2{}^2$ the second largest, and so on. $\sigma_m{}^2$ is the smallest nonzero eigenvalue. The pseudoinverse of Θ is

$$\Theta^\dagger = V\Sigma^{-1}U^T \qquad (18)$$

Here Σ^{-1} is diag $(\sigma_1{}^{-1}, \sigma_2{}^{-1}, ..., \sigma_m{}^{-1})$. Note that the smallest singular value, σ_m, becomes the largest in the pseudoinverse. Consequently, the sensitivity of the pseudoinverse to small errors (e.g., noise) in Θ may be large if σ_m is small, that is, if the condition number

$$\gamma(\Theta) = \frac{\sigma_1}{\sigma_m} \qquad (19)$$

is large. The key idea of the PCR estimator is to reduce this sensitivity by retaining only $k \leq m$ terms in the sum in eq 17.

3.2. Principal Component Regression (PCR). The SVD in eq 17 is written in the alternative form

$$\Theta = t_1 p_1{}^T + t_2 p_2{}^T + ... + t_m p_m{}^T \qquad (20)$$

Here $t_i = u_i \sigma_i$ is the score vector (or latent variable) and $p_i = v_i$ is the loading vector for principal component i. In PCR, only the k first terms which may be distinguished from measurement noise are kept, and the matrices $P^{q \times k}$ and $T^{n \times k}$ include only these k most important directions. Then $\Theta \approx \Theta_k = TP^T$. The latent variables for a given temperature vector, θ, are then given by $t^{k \times 1} = P^T \theta^{q \times 1}$. The least-squares solution to $Y = TK_t{}^T$ becomes $K_t = Y^T T[T^T T]^{-1}$. The condition number of T, $\gamma(T) = \sigma_1(T)/\sigma_k(T)$, may be adjusted by selecting the number k. Since $P^{-1} = P^T$ (P is orthonormal), the overall estimator gain matrix then becomes

$$K_{PCR} = Y^T(\Theta_k^\dagger)^T = Y^T T[T^T T]^{-1}P^T \qquad (21)$$

3.3. Partial-Least-Squares (PLS) Regression. This is a variation of the PCR method which recently has become popular among analytical chemists. The latent variables are here determined in order to have the greatest covariance with the y variables. It is an iterative process, which (see Höskuldsson, 1988) may be described as follows.

For Θ and Y these parameters are found:

(1) the largest eigenvalue a_1 and corresponding
   eigenvector w_1 of $\Theta^T Y Y^T \Theta$ (scale w_1 to length one)

(2) the scores $\qquad t_1 = \Theta w_1$

(3) the Θ loadings $\quad p_1 = \Theta^T t_1/(t_1{}^T t_1)$

(4) the Θ residual matrix $\quad E_1 = \Theta - t_1 p_1{}^T$

(5) the Y loadings $\quad q_1 = Y^T t_1/(t_1{}^T t_1)$

(6) the Y residual matrix $\quad F_1 = Y - t_1 q_1{}^T$

The procedure is started from the top with the residual matrices E_1 and F_1 instead of Θ and Y, and continued until the matrix $E_k{}^T F_k F_k{}^T E_k$ has only small eigenvalues left. For a more exact description of the algorithm and its different versions, see Martens and Næs, 1989.

The estimator based on k factors is

$$K_{PLS} = Q(P^T W)^{-1}W^T \qquad (22)$$

where the matrices $Q^{p \times k}$, $P^{q \times k}$, and $W^{q \times k}$ are formed by the vectors q, p, and w introduced above.

The main advantage of the PLS algorithm compared to PCR is that it selects the directions in Θ which have the largest covariance with y and thus ensures that these directions are treated first.

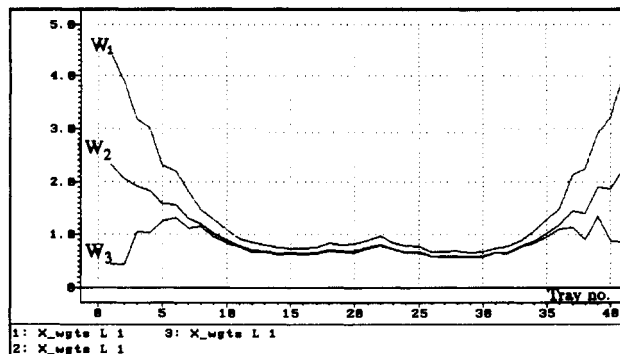### 3.4. Scaling of Variables (Weight Functions). The objective of scaling is to improve the estimate by giving

Figure 9. Weight functions W_1, W_2, and W_3 for case n_2 (untransformed data with 0.2 °C random noise).

each temperature a weight corresponding to the inherent prediction ability.

The most common weight is the inverse of the standard deviation. This ensures that variable scaling (or variable transformations) do not bias the results. The weight for the ith temperature is

$$W_{1i} = 1/s_{ci} \qquad (23)$$

We use s_{ci} as the standard deviations of the calibration set for temperature i, that is, it is the square root of the ith diagonal element of $(1/n)(\Theta^T\Theta)$. This weight simply scales all temperatures such that their changes are of the same magnitude. Since the change in terms of unscaled temperatures is small toward the end of the column, this means that measurements close to the ends will have a large weight, W_1. However, we know that this may not be a good approach because the noise is large (in relative terms) close to the ends. In order to take noise into account in the weighting, Martens and Næs (1989) suggest using the weight

$$W_{2i} = \frac{1}{s_{ci} + s_{e_ki}} \qquad (24)$$

Here s_{e_ki} is an estimate of the noise level, defined as the square root of the residual variance after k factors, i.e., the square root of the ith diagonal element of $(1/n)(E_k{}^T E_k)$. This variance includes all contributions of model/data mismatch, i.e., both noise and mismatch due to nonlinearities, etc. Note that $s_{ci} = s_{e_0i} \geq s_{e_ki}$. Finally, we propose a new weight function

$$W_{3i} = \frac{1}{s_{ci}} \frac{s_{ci} - s_{e_ki}}{s_{ci}} \qquad (25)$$

This weight is equal to W_1 when there is no noise but gives zero weight to measurements where all variations are unexplained (due to noise).

To calculate the weights, we first performed the calibration once without weights. In the weight functions $k = 3$ was used for the binary mixture, and $k = 4$, in the multicomponent mixture. Typical examples of weight functions are given in Figure 9. We see that weight W_3 puts less weight on the column ends than W_1 and W_2.

3.5. Estimators for the Example Column. The different PLS and PCR estimators are identified with a code sequence given in Table III. An "M" denotes a multicomponent mixture. A "P" denotes that different pressure levels of ±0.1 atm are included randomly in the calibration set. L_Y denotes an estimator with logarithmic compositions and untransformed temperatures; L_T denotes logarithmic compositions and logarithmic temperatures; L_θ denotes the same, but with trays 1 and 41 (binary mixture), or trays 4 and 37 (multicomponent mixture) as reference temper-

Linear study

n_0 (PCR)

Nonlinear study

n_0 (PLS)

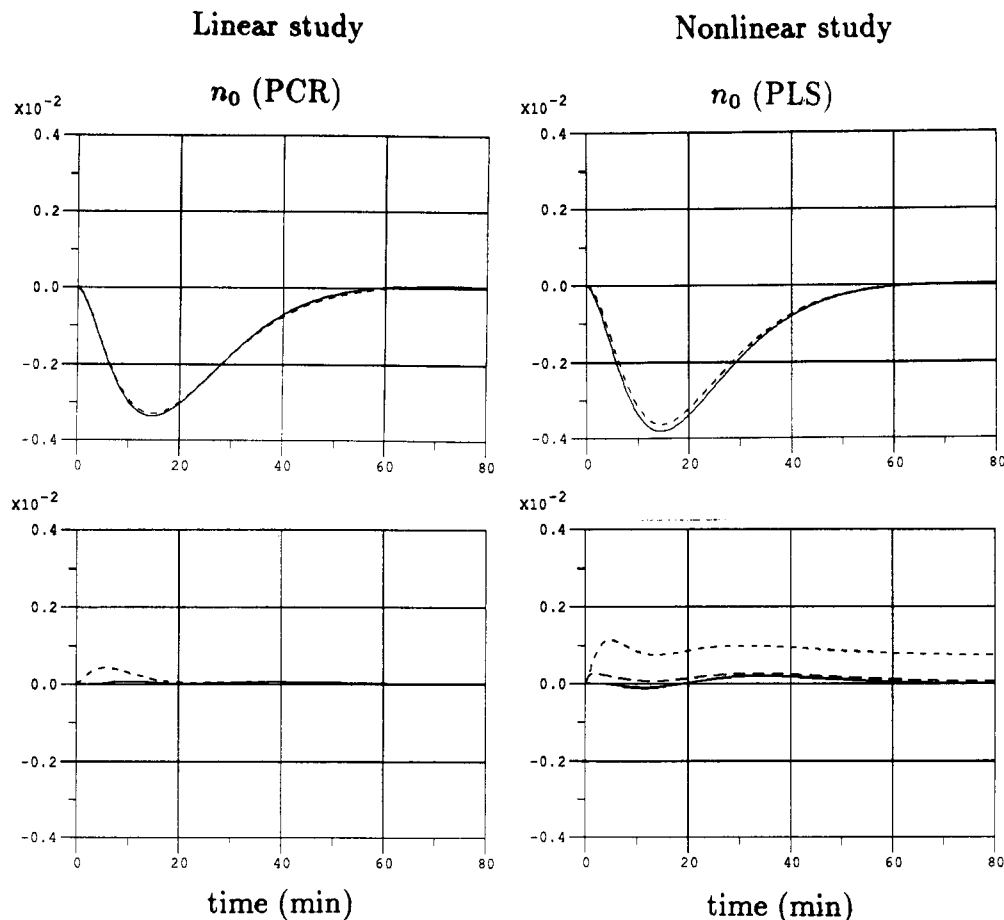time (min)   time (min)

Figure 10. Closed-loop responses of y_D for steps in feed rate (upper) and feed composition (lower) for binary mixture. Input to the controller: perfect measurement (solid line), estimator with three factors (dotted line), and estimator with seven factors (long dotted line).

Table III. Estimator Cases

M	multicomponent mixture
P	pressure variations of ± 0.1 atm
L_Y	logarithmic transformed composition
L_T	logarithmic transformed composition and temperature (reference: boiling temperature)
L_θ	logarithmic transformed composition and temperature (reference: tray temperature)
n_0	no noise
n_1	0.1 °C noise
n_2	0.2 °C noise
W_n	weight function n

atures, rather than the pure component boiling temperatures. Trays 4 and 37 are selected because they have minimum variance in the calibration set and are just outside the nonkey separation area; see Figure 5. The logarithm of the absolute value is used for temperatures 1–3 and 38–41. n_0, n_1, and n_2 denote estimators with zero, 0.1 °C, and 0.2 °C normal distributed noise added to the calibration set. Noise is added to each measurement except for cases P where we add noise on the temperature differences $\theta - \theta_L$ and $\theta_H - \theta$. Note that estimators based on L_θ have a varying reference temperature, which will be corrupted by noise, whereas L_T and PL_θ do not. Finally, W_1, W_2, and W_2 denote the different temperature weight functions in section 3.4.

4. Results

The static explained prediction variance (EPV) (see section 2.6) for a number of cases are summarized in Table IV, and we shall study these results in more detail below. All estimators use PLS regression unless otherwise stated.

4.1. Dynamic and Static Estimation. We shall first discuss the use of PLS estimator n_0 (no noise) for the

binary column. This uses no transformations on y and θ and no weighting. For the linear case (if we use a linear column model at the operating point) with three factors these estimates are identical to those obtained with the PCR estimator studied by Mejdell and Skogestad (1991a). They found the performance of this static estimator to be excellent also when used dynamically and for feedback control. Typical simulation results for the linear case are shown in the left part of Figure 10. The corresponding nonlinear simulations in the right part of Figure 10 show similar dynamic responses. From the figure the main problem seems to be the static prediction capability of the estimator, and we shall therefore use mainly static arguments to evaluate the estimators in the following.

In the dynamic simulations we use a column model with 41 states for the binary case and 123 states for the multicomponent case (not including flow dynamics) and a control system with two PID controllers using LV configuration.

4.2. Effect of Nonlinearity on PLS Factors. From the simulations in the right part of Figure 10 we see that increasing the number of factors in the PLS estimator from three to seven almost removes the effect of nonlinearity. The same conclusion is obtained by considering the static EPV values in Table IV. Since the column itself has only 3 degrees of freedom, three factors would account for 100% of the variance if the column were linear (small perturbations). However, because of nonlinearity, the actual EPV with three factors is only 94%, and the EPV increases by adding factors.

4.3. Influence of Measurement Noise. The simulations above and the EPV values are for the ideal case with no noise, and in practice the results with many factors will not be as good. Figure 11 compares EPV for different

Table IV. EPV for Different PLS Estimators

| | EPV for given no. of factors | | | | | | |
	1	2	3	4	5	6	7
no noise:							
n_0	19.61	81.97	94.17	97.18	98.18	99.43	99.96
$n_0 L_Y$	21.78	80.14	92.86	93.45	96.92	97.20	98.80
$n_0 L_T$	19.72	96.38	99.97	99.98	99.99	100.00	100.00
$n_0 L_\theta$	32.90	95.39	99.95	99.96	99.97	100.00	100.00
$n_0 P$	-2.31	46.77	75.71	92.19	95.50	95.95	97.59
$n_0 L_\theta P$	19.16	94.90	99.94	99.97	99.97	100.00	100.00
Mn_0	10.66	48.02	56.78	88.35	89.58	91.50	91.94
$Mn_0 W_2$	11.14	48.44	57.40	87.76	90.61	91.84	92.25
$Mn_0 L_Y$	18.52	48.88	52.59	83.66	84.10	87.84	87.43
$Mn_0 L_\theta$	22.73	77.52	94.10	95.81	96.27	97.61	97.55
$Mn_0 L_\theta W_2$	20.69	73.75	93.81	95.87	96.75	97.68	97.66
0.1 °C noise							
n_1	19.59	81.78	93.81	97.00	97.59	97.86	98.05
$n_1 L_T W_3$	23.10	91.42	98.89	98.92	98.92	98.82	98.80
$n_1 L_\theta W_3$	23.18	84.60	91.09	94.01	96.16	96.69	97.06
$n_1 L_\theta W_3 P$	26.89	91.05	98.25	98.78	99.04	99.08	99.04
0.2 °C noise							
n_2	18.87	81.10	93.78	94.92	94.49	94.42	94.47
$n_2 W_1$	17.06	86.04	95.75	94.66	94.39	94.18	93.64
$n_2 W_2$	20.40	84.65	95.66	95.36	94.74	94.70	94.12
$n_2 W_3$	19.78	83.34	95.08	95.59	95.40	95.25	95.04
$n_2 L_Y W_3$	18.14	80.14	91.53	92.57	91.73	91.29	90.53
$n_2 L_T$	12.50	78.41	81.90	84.04	87.59	90.63	91.16
$n_2 L_T W_1$	22.65	86.65	95.44	96.25	96.70	97.22	97.33
$n_2 L_T W_2$	22.78	87.66	96.43	96.99	97.30	97.64	97.59
$n_2 L_T W_3$	23.04	89.07	97.49	97.68	97.86	98.14	98.05
$n_2 L_\theta W_3$	26.41	84.53	86.59	87.43	90.07	89.10	88.52
$n_2 L_\theta W_3 P$	20.58	86.30	98.01	98.09	97.89	97.84	97.80
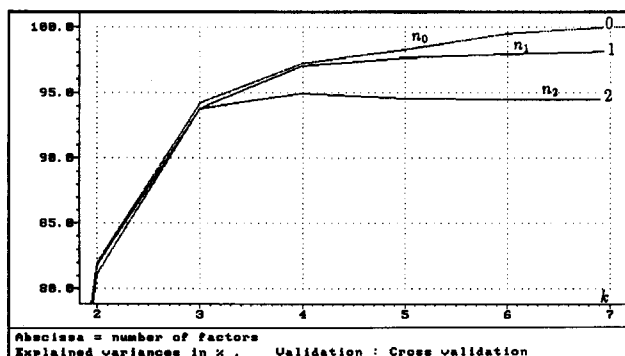
Figure 11. Effect of measurement noise on EPV for estimator n_i: $i = 0$, no noise; $i = 1$, 0.1 °C; $i = 2$, 0.2 °C.

Figure 12. Loading plot (vector p_i) of the first three factors for estimator n_0. Curve identifier: i, factor number.

Figure 13. Loading plot for case $n_0 P$ with pressure variation. Curve identifier: Factor number.

levels of measurement noise on the temperatures in the calibration set. The noise will corrupt the smallest factors, and after three to five factors there is no improvement with the addition of factors, and it may even reduce the prediction ability. We therefore see that in the presence of noisy measurements, it is doubtful to use additional factors to capture nonlinearity.

4.4. Insights about Directions in the Temperature Space. Figure 12 displays a typical plot of the three largest loading vectors; that is how the different measurements are summed up to make the factors (latent variables). By comparing with the temperature profiles in Figure 4, we see that the first factor is due to the changes in external streams, D and B, and reflects movement in the temperature profile up and down the column. The second factor is related to changes in the internal streams, L and V, when D and B are held constant. It reflects the magnitude of separation in the column. The third factor is due to changes in feed composition.

4.5. Effect of Pressure. Figure 13 displays the loading plot for the case with total column pressure variations of 10% (case $n_0 P$ in Table IV). We see that the first factor, which has no predictive ability, mainly represents the
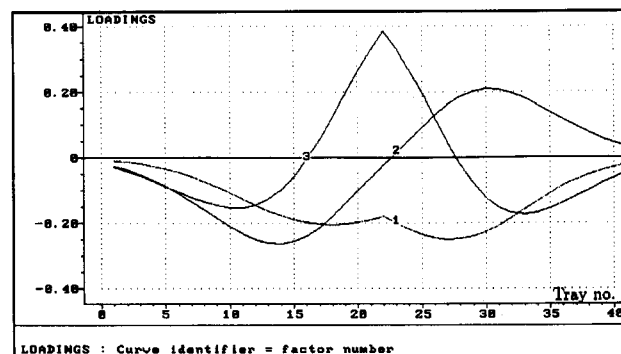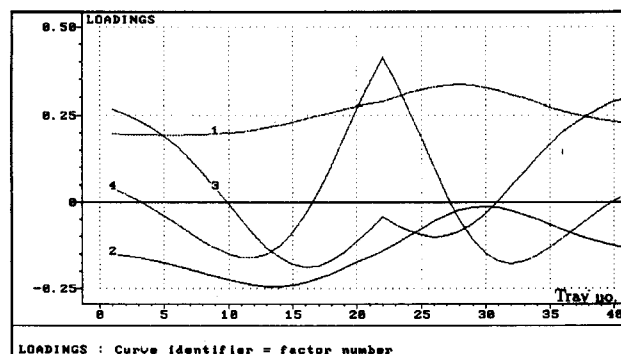
pressure variation. To get good predictive properties, we need at least five factors. The pressure variation may alternatively be taken care of by using differential temperatures.

4.6. Use of Logarithmic Transformations. The results in Table IV seem to indicate that use of logarithmic transformed compositions, i.e., $Y_D = \ln(1 - y_D)$ and X_B
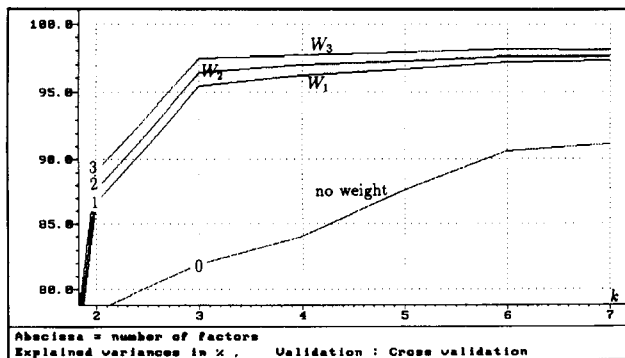
Figure 14. Effect of weights on EPV for estimator $n_2L_TW_i$ with logarithmic temperatures and 0.2 °C noise.

= $\ln x_B$, combined with *un*transformed temperatures (L_Y estimators), generally has a negative effect on the estimate. However, the results are not quite comparable because the EPV is based on Y_D and X_B instead of y_D and x_B. Other tests show that they are comparable in performance, but they have slightly different properties: The accuracy of the logarithmic estimator will be best in the pure region and will never give estimates outside the region 0–1. Furthermore, for feedback control it may be an advantage to use Y_D and X_B because this makes the controller tunings less dependent on operating conditions (e.g., Skogestad and Morari, 1988b).

However, estimator performance is significantly improved by using logarithmic transformations also on the temperatures. We see from Table IV, cases n_0L_θ and n_0L_T, that with no noise EPV is close to 100% after only three factors. This is the case also when pressure variations are included; that is, these are automatically taken care of when transformed temperatures L_θ are used.

4.7. Effect of Weights on Temperatures. We shall consider the case with logarithmic temperatures. In cases n_0L_T and n_0L_θ with no noise, there is no improvement of weighting the measurements, because the logarithmic transform will automatically weigh the temperatures similar to weight W_1. But when noise is added to the calibration set, weighting is very important. Figure 14 compares different weight functions with 0.2 °C noise. Weight W_3 yields the best result.

4.8. Noise of Reference Temperatures. Case $n_2L_\theta W_3$ in Table IV shows that estimators using corrupted (noisy) reference temperatures perform poorly. The values of L_θ close to the location of the reference will then be very sensitive to noise. However, in practice, temperatures close to each other are perhaps more likely to have correlated rather than independent noise. The results with the noise put on the temperature difference instead, such as $L_\theta P$ and L_T, may therefore be more realistic. Anyhow, the results demonstrate that logarithmic transforms may be quite sensitive to noise.

4.9. Multicomponent Mixture. The results in Table IV for case Mn_0L_θ show that with no noise we obtain EPV values in the range 94.1–97.6% for static PLS estimators using logarithmic temperatures and three to six factors. We see that we need more factors compared to those for the binary mixture, mainly due to the additional degrees of freedom caused by the two additional components. However, we see that even with more factors, we do not attain the same prediction capability as for the binary mixture. The improvement from using logarithmic temperatures is however substantial also in this case.

The simulations in Figure 15 illustrate the dynamic performance of the static PLS estimator. The performance is somewhat worse than for the binary mixtures shown in Figure 10. The worst disturbance seems to be changes in
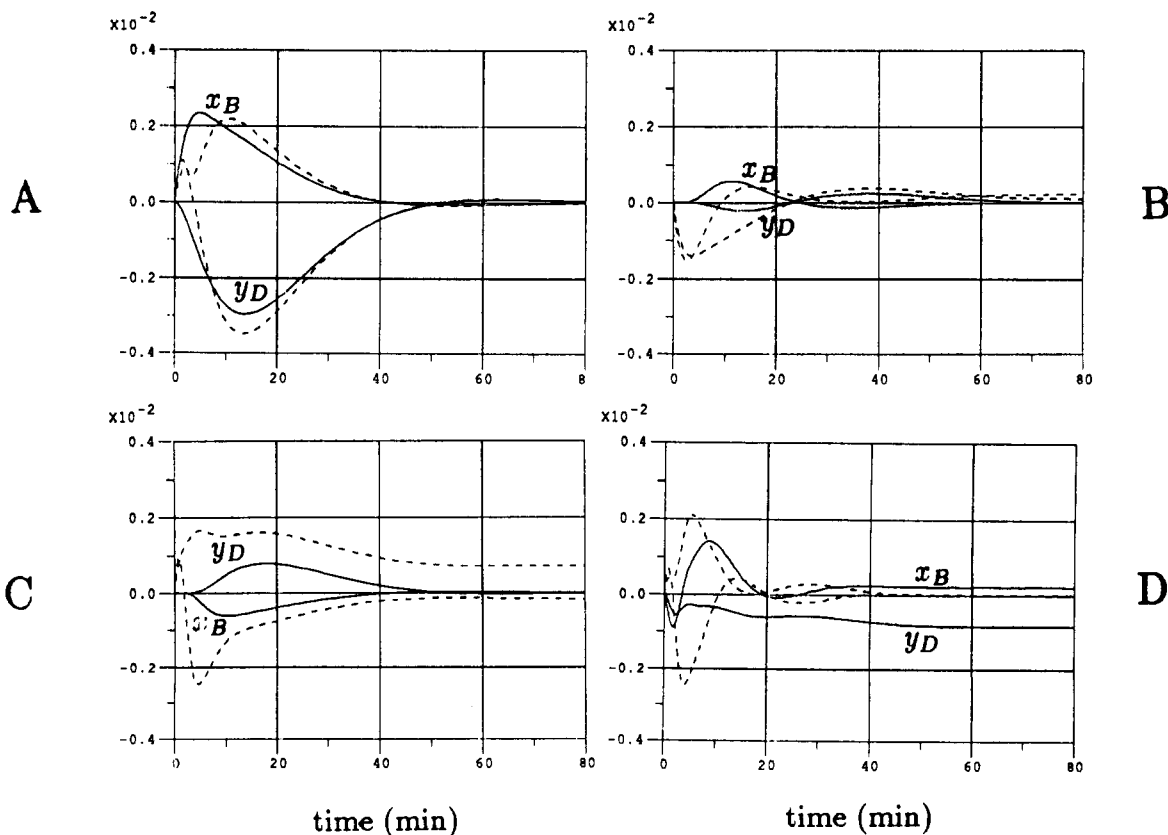

time (min)

Figure 15. Closed-loop responses of compositions y (solid line), and estimates $\hat{y}$ (dotted line), for multicomponent mixture. Estimator is $Mn_0L_\theta W_2$ with four factors. Results are shown for 20% steps in (A) feed rate, (B) feed composition z'_f, and (C, D) heavy nonkey composition. y is used for feedback in cases A–C, whereas $\hat{y}$ is used in case D.

Table V. Improvement of % EPV Using PLS Rather Than PCR Estimator

	EPV for given factor k						
case	1	2	3	4	5	6	7
n_0	-9.64	0.90	0.06	0.14	0.01	0.49	0.00
n_1	-9.76	0.87	0.07	0.06	0.60	0.68	1.16
n_2	-9.63	0.91	0.05	0.03	-0.32	-0.23	-0.35
$n_0 L_Y$	-9.26	1.00	0.01	0.35	4.60	0.86	0.01
$n_0 P$	1.48	-1.70	4.99	0.02	0.13	0.34	0.59
$n_2 L_T W_1$	-11.70	1.37	4.31	2.29	2.14	2.11	2.34
$M n_0 L_\theta$	-5.96	1.64	16.54	4.11	-0.27	1.36	0.33

the heavy component. However, case D in Figure 15 demonstrates that even this seemingly poor estimate may be acceptable for feedback control.

4.10. Comparison of PLS and PCR. In Table V the two methods are compared. The values of practical interest are those for three to five factors. In most cases the difference is small, although PLS is generally somewhat better.

5. Discussion

5.1. Comparison with One Temperature Estimator. All results above were based on using all 41 temperatures for estimation. To compare with the conventional one-temperature control, we used the same calibration data to compute the EPV for an estimator using only one temperature. Different measurement locations were considered for the case with the binary mixture and 0.2 °C noise. To estimate y_D, tray no. 9 was found to be optimal with EPV = 88.7%, and for logarithmic compositions, EPV = 85.4%. This compares to EPV = 97.5% obtained with logarithmic estimator $n_2 L_T W_3$ with three factors and using all temperatures. The relatively good EPV values for the case with one temperature may explain why one-temperature control is popular for composition control. However, the accuracy of the estimate is quite sensitive to the location, and the EPV is reduced 4-5% only two trays off tray 9. In addition to yielding better estimates, the use of multiple temperatures is therefore less sensitive to measurement location and to noise.

5.2. Neglected Effects. The results have demonstrated that it is possible to obtain quite precise estimates with the PLS and PCR regression methods. The results are based on a simulation study where we neglected variations in tray efficiency, liquid fraction in feed, reflux subcooling, local pressure drops, etc. Variations in these will affect the temperature profile to some extent. However, we believe that in many cases of these variations will be too small to be distinguished from noise or will not significantly change the relationship between θ and y.

5.3. Noise Level. Our mixture has a relative volatility of 1.5, corresponding to a temperature difference between the two key components of only 13 °C. Nevertheless, for the binary mixture we were able to obtain a prediction capability (EPV) of 97.5% with only three factors for the case with 0.2 °C noise. This noise level is about 1.5% of the temperature difference. Hence, a mixture with a temperature difference of 40 °C should be able to cope about equally well with a noise level of about 0.6 °C.

5.4. Coping with Nonlinearity and Noise. A major problem for the estimator is that the temperatures at the column ends, which are most representative for the product streams (at least in the binary case), also are most affected by nonlinearity and noise.

We have proposed three different methods to deal with this problem: use more factors, weight according to temperature variation and noise, and logarithmic transforms.

Using a larger number of factors than the number of degrees of freedom is helpful when the noise level is not too high. The reason why this may help is that the product compositions have different nonlinear relations to different temperatures. These differences will appear as extra directions (factors) in the linear temperature space. However some of these directions may be too small to be distinguished from noise. A useful rule is to increase the number of factors until they no longer have any significant positive effect on the EPV. For our distillation column, the typical optimal number of factors is three to five.

Weighting of variables is commonly used for dealing with different kinds of measurements. This is to prevent domination by measurements with the largest nominal changes. In our distillation column only temperature measurements are used, so one might think that weighting is unnecessary. Nevertheless, weighting proved useful. The reason is that the temperature changes are very small at the column ends (in the calibration set their standard deviation is only 6% of the temperature with the largest variation), but even in the presence of noise they do contain useful information about the end compositions. By using weights we avoid discarding the use of these temperatures with the PLS method. However, the noise should also be taken into account, and weight functions W_2 and even more W_3, which include information about the noise, yield better results.

Compared to PCR, there is a kind of weighting inherent in the PLS method (Höskuldsson, 1988). This follows since it searches for directions in the $\Theta^T Y Y^T \Theta$, instead of only $\Theta^T \Theta$. The weighting with the matrix $Y Y^T$ gives temperatures which have the largest covariance with y larger weight when the factors are made. The comparison between PLS and PCR shows that these "weightings" improved the estimates in some cases, although the difference was quite small.

The use of logarithmic transformations of temperatures was clearly the single method with the greatest effect. It appears to be a very powerful method to cope with the nonlinearity in distillation columns and also automatically gives the temperature measurement at the ends a greater weight. However, when the temperatures are corrupted with noise, the noise will also be transformed and will have a relatively large effect on these end temperatures. Therefore, as seen from Figure 14, it is absolutely necessary to weight the transformed data to take this into account. As already mentioned, one should also take action to avoid very small or negative temperatures differences before transforming the temperatures.

5.5. Choice of Reference Temperature. For multicomponent mixtures the reference temperatures should be located some distance away from the ends. (1) They should be in the section where the concentration of the off-key component is almost constant (See Figure 5). Much of the off-key component's contribution to the temperature will then be cancelled. (2) They should be located as far out to the ends as possible to capture the nonlinearity. A simple method which combines these two criteria is to select the temperature with the least variation in the calibration set.
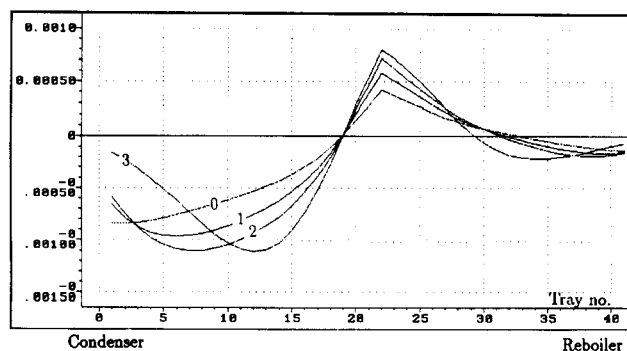
Figure 16. Estimator vector **K** (y_D) for PLS estimator $n_i W_2$ with three factors. Curve identifier i: 0, no noise; 1, 0.1 °C; 2, 0.2 °C; 3, ∞ °C (the same as no weighting).

5.6. Measurement Selection. The results above are based on using *all* temperatures as measurements. This is of course not necessary. However, the number of measurements should at least be equal to the number of factors needed for prediction. For example, to capture three factors, we need at least three to five temperatures. The highest number applies to estimators which use differential temperatures, for example, L_θ. Additional temperatures will mainly reduce the effect of measurement noise. As a simple method to select the location of temperature measurements, we recommend identifying the peak elements in the **K** matrix for the weighted (scaled) variables. The number of peaks is usually the same as the number of factors. As an illustration, consider Figure 16, which displays the elements in **K** for the weighted untransformed temperatures for y_D using PLS estimators $n_0 W_2$, $n_1 W_2$, $n_2 W_2$, and n_0 with three factors (the weight W_2 is different in each case as it depends on the noise level). We see that when the noise is increased it seems better to locate the temperatures further from the end.

A simple procedure for measurement selection and obtaining the estimator is to (1) find all possible measurement locations and include all these in steps 2–7, (2) determine expected magnitudes of outputs (y) and all variables (disturbances) affecting the system, (3) perform simulations that include the expected variations (using factorial design and making sure the entire output space is spanned), (4) add random noise to all measurements, (5) transform the variables and perform the PLS regression, (6) determine the number of optimal factors and weigh the variables with a suitable weight function, for example, W_3, (7) do the PLS regression once more with the weighted variables, but without noise on the measurements, (8) find from the weighted **K** matrix which of the possible measurements to select, and (9) perform the final calibration to obtain the estimator with the selected set of weighted measurements.

The advantage with this procedure, compared for instance to that of Moore (1987), is that for each measurements it explicitly takes care of both its correlation to the outputs and its noise level.

6. Conclusions

The problem addressed in this paper belongs to a broad class of problems concerning how to handle multiple measurements in an estimator. Depending on the magnitude of correlation with the estimated variable and its sensitivity to noise, the different sensors should be weighted to give the best estimate. The paper has shown that this indeed also applies for sensors of the same type such as the temperatures in a distillation column.

For distillation columns the main difficulty of using linear estimators is the nonlinearity in the process. It is found that the logarithmic transforms of compositions and temperatures proposed in this paper is a highly powerful means of coping with this nonlinearity. Together with weight functions that place less weight on sensors with large noise, these transforms are found to give a substantial improvement in the prediction ability.

Use of logarithmic temperatures, L_θ, which makes use of differential temperatures, gives the additional benefit of counteracting pressure variations.

The results for multicomponent mixtures indicate that the estimator may perform well in a wide range of applications. Using section reference temperatures at locations with the smallest temperature variance will make logarithmic transforms useful also here.

Besides being an efficient method of obtaining estimators, the standard multivariable calibration techniques yield added benefits, such as insight in the process, good statistical information about the prediction ability, and a method for sensor location.

Acknowledgment

Nomenclature (Also see Table III)
B = bottom product flow rate
D = distillate flow rate
E_k = residual matrix for Θ using k factors
EPV = explained prediction variance (see eq 12), %
F = feed rate
k = number of factors used in estimator
K = estimator constant
L = reflux flow rate
L_T = logarithmic temperature based on boiling points
L_θ = logarithmic temperature based on reference temperatures
t = vector of latent variables
T = matrix of latent variables (**t**) for calibration runs
T^b = boiling temperature of pure component
V = boilup from reboiler
x_B = mole fraction of light component in bottom product
y_D = mole fraction of light component in distillate
y = output vector $(y_D, x_B)^T$
Y = matrix of outputs (**y**) for calibration runs
z_F = mole fraction of light component in feed
W_1, W_2, W_3 = weight functions for measurement (temperature) scaling

Greek Symbols

$\alpha_{i,j}$ = relative volatility between components i and j
γ = condition number
σ_i = ith singular value
θ = tray temperature
$\boldsymbol{\theta}$ = temperature vector
θ_L, θ_H = reference temperatures in top and bottom of column
Θ = matrix of temperatures (θ) for calibration runs

Subscripts

H = heavy key component
i = tray number
L = light key component

Superscripts

$'$ = pseudobinary basis

Appendix 1: Linearizing Effect of Logarithmic Transform on Profile

(1) For the condition of a binary mixture and assumed constant relative volatility, then

$$\frac{y_i(1 - x_i)}{x_i(1 - y_i)} = \alpha \qquad (26)$$

At infinite (total) reflux we have $y_i = x_{i+1}$ and we get

$$\ln\left(\frac{x_{i+1}}{1 - x_{i+1}}\right) - \ln\left(\frac{x_i}{1 - x_i}\right) = \ln \alpha \qquad (27)$$

If the logarithmic mole fraction is introduced

$$X_i = \ln\left(\frac{x_i}{1 - x_i}\right) \qquad (28)$$

Then

$$X_{i+1} - X_i = \ln \alpha \qquad (29)$$

That is, the profile in terms of X_i is linear.

(2) At finite reflux we derive a similar expression: In the bottom part the material balance is

$$L_B x_{i+1} = V y_i + B x_B \qquad (30)$$

where $L_B = L + F$ is the liquid flow in the bottom part. Equation 26 then gives

$$\ln\left(\frac{x_{i+1} - (B/L_B)x_B}{1 - (L_B/V)x_{i+1} + (B/V)x_B}\right) - \ln\left(\frac{x_i}{1 - x_i}\right) =$$
$$\ln\left(\alpha\frac{V}{L_B}\right) \qquad (31)$$

In most cases the first term is approximately equal to $\ln[x_{i+1}/(1 - x_{i+1})]$. (The numerator approximation may not apply to the bottom tray if B/L_B is close to 1; the denominator approximation usually does not apply close to the feed tray where x_{i+1} is not small.) We then get

$$X_{i+1} - X_i \approx \ln\left(\alpha\frac{V}{L_B}\right) \qquad (32)$$

(Here $\alpha(V/L_B)$ is the ratio between the slope of the equilibrium line and operating line in the bottom section.) A similar expression but with $\alpha(V/L_B)$ replaced by $\alpha(L/V_T)$ is derived for the top. Around the feed tray the expression does not apply, and the profile in terms of X_i may not be linear.

(3) To derive a linearizing transform in terms of temperatures, we assume that the boiling temperature is a linear function of x_i, that is, for the binary case

$$\theta_i = T_L^b x_i + T_H^b(1 - x_i) \qquad (33)$$

and we derive

$$\frac{x_i}{1 - x_i} = \frac{T_H^b - \theta_i}{\theta_i - T_L^b} \qquad (34)$$

and we have

$$L_{Ti} = \ln\left(\frac{\theta_i - T_L^b}{T_H^b - \theta_i}\right) = -X_i \qquad (35)$$

Appendix 2: Linearizing Effect of Logarithmic Transforms on Dynamic Response

(Also see Skogestad and Morari, 1988a.)

With consideration of the component material balance on a given tray with vapor flow V and liquid flow L (as-suming constant molar flows),

$$M_i \frac{dx_i}{dt} = L(x_{i+1} - x_i) + V(y_{i-1} - y_i) \qquad (36)$$

If the column is assumed to be at steady-state and the initial response to a small change in L is considered, we get

$$M_i \frac{dx_i}{dt} = \Delta L(x_{i+1} - x_i) \qquad (37)$$

Here $x_{i+1} - x_i$ varies strongly with tray number i and with the operating point. However, if both sides are divided by x_i, then

$$M_i \frac{d \ln x_i}{dt} = \Delta L\left(\frac{x_{i+1}}{x_i} - 1\right) \qquad (38)$$

In the bottom part of the column the ratio $x_{i+1}/x_i \approx \alpha(V/L_B)$ and is approximately constant (Appendix 1). That is, in terms of logarithmic compositions the *initial dynamic response* is almost independent of tray number and operating point. A similar expression is derived for changes in V.

Similar expressions in terms of $d \ln (1 - x_i)$ are derived for the top part. Since in the bottom $1 - x_i \approx 1$ and in the top $x_i \approx 1$, we may combine these transforms and use for the entire column the following transform

$$X_i = \ln\left(\frac{x_i}{1 - x_i}\right) \qquad (39)$$

This transform also has some linearizing effect in the static case, but to a lesser extent (Skogestad and Morari, 1988a).

Literature Cited

Bozenhardt, H. F. Modern control tricks solve distillation problems. *Hydrocarbon Process.* 1988, 6, 47-50.

Gilles, E. D.; Retzbach, B. Reduced models and control of distillation columns with sharp temperature profiles. *Proc. 19th IEEE Conf. Des. Control* 1980, 865-870.

Höskuldsson, A. PLS Regression Methods. *J. Chemom.* 1988, 2, 211-228.

Joseph, B.; Brosilow, C. B. Inferential Control of Processes. *AIChE J.* 1978, 24, 485-509.

Joseph, B.; Brosilow, C. B.; Howell, J. C.; Kerr, W. R. D. Multi-temps give better control. *Hydrocarbon Process.* 1976, 3, 127-131.

Kalman, R. E.; Bucy, R. S. New Results in Linear Filtering and Prediction Theory, ASME. *J. Basic Eng.* 1961, 83, 95-108.

Kister, H. Z. *Distillation Operation*; McGraw-Hill, New York, 1990; p 545. Marquardt, W. Nonlinear model reduction for binary distillation. *DYCORD 86*, 1986, 123-128.

Marquardt, W. Concentration profile estimation and control in binary distillation. In *Proceedings of the IFAC Workshop, Model Based Process Control*, Atlanta, GA, 13-14 June 1988; McAvoy, T. J., et al., Eds.; International Federation of Automatic Control, Pergamon Press: Oxford, U.K., 1989.

Martens, H.; Næs, T. *Multivariate Calibration*; John Wiley & Sons: New York, 1989.

Mejdell, T.; Skogestad, S. Estimate of process outputs from multiple secondary measurements. *Proc. Am. Control Conf.* 1989, 2112-2121.

Mejdell, T.; Skogestad, S. Output estimation for ill-conditioned plants using multiple secondary measurements: High-purity distillation. Submitted for publication in *Automatica*, 1991a (also presented at the 1990 AIChE meeting, Chicago; Paper 23g).

Mejdell, T.; Skogestad, S. Composition Estimator in a Pilot-Plant Distillation Column Using Multiple Temperatures. *Ind. Eng. Chem. Res.* 1991b, following paper in this issue.

Moore, C.; Hackney, J.; Canter, D. Selecting Sensor Location and Type for Multivariable Processes. In *Shell Proceedings of the Control Workshop*, Houston, TX, 1986; Prett, D. M., Morari, M., Eds.; Butterworth: Boston, 1987.

Morari, M.; Stephanopoulos, G. Optimal Selection of Secondary Measurements within the Framework of State Estimation in the

Presence of Persistent Unknown Disturbances. *AIChE J.* 1980, *26*, 247.

Nisenfeld, A. E.; Seeman, R. C. *Distillation Columns*; ISA Monograph Series 2; Instrument Society of America: Research Triangle Park, NC, 1981.

PROCESS Reference Manual; Simulation Science Inc.: Fullerton, CA, 1981; pp 9.43–9.44.

Rademaker, O.; Rijnsdorp, J. E.; Marrleveld, A. Dynamics and Control of Continous Distillation Units; Elsevier: Amsterdam, 1975.

Riggs, J. Letter to the editor. *AIChE J.* 1990, *36* (7), 1124–1125.

Ryskamp, C. Using Probability Axis for Plotting Composition Profiles. *Chem. Eng. Prog.* 1981, *77*, 42–47.

Skogestad, S.; Morari, M. Understanding the Dynamic Behavior of Distillation Columns. *Ind. Eng. Chem. Res.* **1988a**, *27*, 1848–1862.

Skogestad, S.; Morari, M. LV-Control of a High-Purity Distillation Column. *Chem. Eng. Sci.* **1988b**, *43* (1), 33–48.

Strang, G. Linear Algebra and its Applications; Harcourt Brace Jovanovich: New York, 1980.

Weber, R.; Brosilow, C. B. The Use of Secondary Measurements to Improve Control. *AIChE J.* 1972, *18*, 614–623.

Whitehead, D. B.; Parnis, M. Computer control improves ethylene plant operation. *Hydrocarbon Process.* 1987, 105–108.

Composition Estimator in a Pilot-Plant Distillation Column Using Multiple Temperatures

Thor Mejdell and Sigurd Skogestad*

Chemical Engineering, University of Trondheim, NTH, N-7034 Trondheim, Norway

Results are given for the implementation of a static partial least-squares (PLS) regression estimator for product compositions on a high-purity pilot-plant distillation column. Temperatures on all 11 trays are used as inputs to the estimator. Several estimators were tested off line to compare their performance, and one estimator was used on line for dual composition control. It was found that the estimators perform very well when appropriate logarithmic transforms and scalings are used. Since the estimator is static, the implementation is straightforward. An estimator based only on experimental data gave excellent performance over a wide range of operating points. Estimators based on simulations did not perform quite as well, and the bias had to be adjusted when a change was made from one operating point to another. Nevertheless, since it may be difficult to obtain good experimental data in an industrial setting, this estimator is probably most useful in practice. In this paper we also discuss how to combine information from simulations (basic modeling) and experiments.

1. Introduction

Product composition analyzers for distillation columns, such as gas chromatographs, have large investment and maintenance costs, in addition to unfavorably large measurement delays. The most popular means of product control is therefore temperature control (Kister, 1990), which provides an easy, fast, and inexpensive means of composition control.

The temperature selected for control is usually located at a tray some distance away from the column ends, because the products may be extremely pure, and the temperature variations are then small compared to the noise. Furthermore, pressure variations and off-key components will interfere with the relationship between product composition and temperature and locating the measurements away from the ends is favored (Rademaker et al., 1975, p 421). However, if the measurement is located too far from the end the temperature will be strongly influenced by the composition of the feed and of the product at the other column end.

An important issue in conventional temperature control is therefore to find the best measurement location by making proper compromises between these considerations.

However, some of the interferences may be handled by using more measurements. For example, since the column pressure has about the same effect on all temperatures in the column, the pressure variation may be compensated for using temperature differences. This requires an additional temperature measurement which preferably is located at a tray where the composition is almost constant. Along the same line of thought are the proposals to use double differential temperatures. Yu and Luyben (1984) proposed use of the other differential temperature for off-key-component compensation, while Luyben (1969) and Boyd (1975) proposed using it for column-pressure-drop compensation. However, these ideas do not seem to be widely applied.

On the other hand, for the case of high-purity columns with large relative volatility between the components, the use of multiple temperatures has found some applications because the conventional temperature control is difficult. In these columns the main temperature drop will take place in a small region consisting of only a few trays. Quite small deviations from the normal operating point may lead to a control temperature outside this region. On the other hand, the location of this temperature front (region) is usually closely correlated to the compositions and may alternatively be the control objective. Bozenhart (1988) located the front by scanning multiple temperatures for the maximum temperature drop between two trays. Luyben (1971) suggested tracking the temperature front by using an average of many tray temperatures. Whitehead and Parnis (1987) used a weighted average of many differential temperatures in a C_2 splitter.

A more rigorous means of using multiple temperatures is to provide an estimator for product compositions. Many approaches have been proposed, e.g., by Brosilow and co-workers (Weber and Brosilow, 1972; Joseph and Brosilow, 1978), who used temperatures together with stream measurements and a linearized process model, and by Marquardt (1989), who used a state space observer for the location of the temperature front.