

Output Estimation Using Multiple Secondary Measurements: High-Purity Distillation

Thor Mejdell and Sigurd Skogestad

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

Measurements of temperatures (secondary outputs) and flows (inputs) are used to estimate product compositions (outputs) in a distillation column. The problem is characterized by strong collinearity (correlation) between temperature measurements and an ill-conditioned model from inputs to outputs. In a linear study, three estimator methods, the Kalman-Bucy filter, Brosilow's inferential estimator, and principal component regression (PCR), are tested for performance with μ -analysis. One can achieve remarkably good control performance with the static PCR estimator, which is almost as good as the dynamic Kalman filter. The quality of the estimate for these two estimators is improved by additional temperature measurements, although the improvement is only minor for more than about five measurements. On the other hand, the performance of the Brosilow inferential estimator may not improve by adding measurements due to sensitivity to modeling errors. For all estimators, the use of flow (input) measurements does not improve the estimator performance and does in fact damage the performance if a static estimator is used.

Introduction

This article addresses estimation of unmeasured process outputs based on multiple secondary measurements. The application chosen here is the use of temperature and flow measurements to estimate the product compositions in a distillation column, which has been studied previously by a number of authors, for example, Joseph and Brosilow (1978a,b). This is an important industrial problem since on-line composition measurements are expensive and unreliable. Furthermore, it provides an interesting estimation problem which features a large number of strongly coupled measurements, and several disturbances and inputs with similar effects on the outputs resulting in an ill-conditioned plant model.

This article compares and discusses various dynamic and static estimators. Although it is based on a particular distillation column example, the treatment in this article is rather general. Readers who want more information about the distillation example and how temperatures may be used to estimate compositions in distillation columns should consult Mejdell and Skogestad (1991a). In that article we also discuss in detail the use of the static PCR or PLS estimators for the example column, and we introduce means coping with non-

linearity. In Mejdell and Skogestad (1991b) we consider the implementation of such static estimators on a pilot-plant column, and we also discuss the use of experimental data rather than models to obtain the estimators.

Problem definition

The objective is to obtain the best estimate \hat{y} of the outputs (product compositions in our application) using all available information, $\bar{\theta}$. In terms of deviation variables the linear estimator may be written:

$$\hat{y}(s) = K(s)\bar{\theta}(s) \quad (1)$$

This estimate should be obtained based on a description of the process (nominal model and expected uncertainty), the expected noise and disturbances, and a more precise definition of what we mean by "best." In the general case, $\bar{\theta}$ should include all measured dependent variables (primary measurements, y , and secondary measurements, θ) and all known independent variables (manipulated inputs, u , and measured disturbances, d). In this article, we usually have $\bar{\theta} = \theta$, that is, the estimate is based on only the secondary temperature measurements. The reason is that we assume no primary measure-

Present address of T. Mejdell: SINTEF Applied Chemistry, N-7034 Trondheim, Norway.

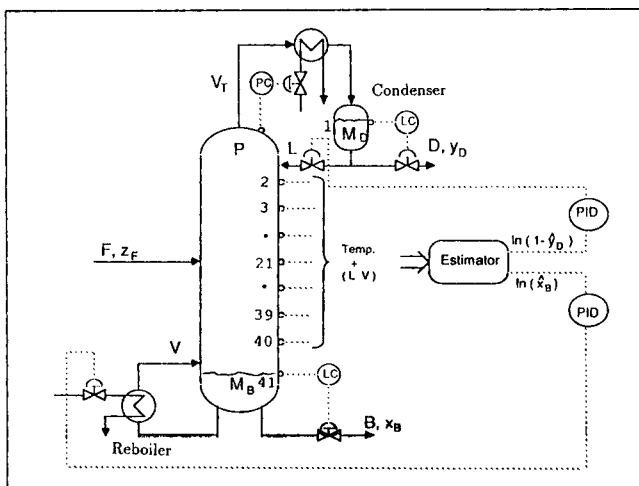


Figure 1. Control scheme based on LV configuration.

ments, no measured disturbances, and we shall show for our case that the additional information contained in u is of limited value. This estimation problem is usually called "inferential" estimation in the process control literature.

Use of separate estimator

An estimator-based control scheme for the distillation column is shown in Figure 1. Note that we are implicitly assuming that the controller should be separated into two parts: one estimator which condenses all the measurements into a few estimated outputs, and a "small" (in terms of number of inputs) controller which uses these estimates for feedback control. The motivation for doing this is reliability, design simplicity and robustness.

In this article, we consider three different approaches to the problem: the Kalman-Bucy filter, Brosilow's inferential estimator, and principal component regression (PCR). In the last two cases, we use a steady-state (static) estimator such that K is a constant matrix.

Use all available measurements

The statement in the problem definition above that the best estimate should be based on *all* available measurements is not as obvious as one should think. For example, our example column has 41 temperature measurements. That is, we need to determine 41 parameters in K for each output if all temperatures are used. However, since the temperatures are strongly coupled, all these parameters cannot be determined independently. To avoid this overparameterization problem, the general consensus within the process control community has been until quite recently that one should use only a few (typically, two or three) temperatures (for example, Weber and Brosilow, 1972; Joseph and Brosilow, 1978a,b; Morari and Stephanopoulos, 1980; Patke et al., 1982; Yu and Luyben, 1986; Moore et al., 1987, Keller and Bonvin, 1987). In 1987 and 1988, we started looking at multivariable regression techniques, PCR and the closely related PLS (partial least square), used by the analytical chemists, where it is stressed that *all* available measurements should be used, and this made us question the conventional wisdom. We then set out to do a rigorous linear comparison of the inferential estimator presented by

Brosilow and coworkers (denoted the Brosilow estimator in the following), with a simple regression estimator (PCR) and the standard quadratic state estimation scheme (Kalman filter). The results from this comparison are presented in this article. The analysis includes both measurement noise and model uncertainty, and the structured singular value, μ , is used as a tool. The conclusion is that for the two latter estimators the estimate is improved by including more measurements (specifically, it reduces the effect of measurement noise), and that the earlier finding that more measurements could deteriorate the estimate was a result of the particular methods used in those studies and not an inherent property of the problem.

The fact that the goodness of the *estimate* generally improves when measurements are added does not mean that one necessarily should use all measurements. Additional measurements are costly and may be unreliable, and may therefore not be needed since:

- (1) The relationship between the measurement and the output is insignificant or uncertain.
- (2) The measurements are strongly correlated, and additional measurements provide only minor improvements for the estimate. In our case, there seems to be little need to implement more than about five temperature measurements.
- (3) The measurement noise is so great that the measurements contain little information. In our case, this may be a problem if we use input measurements, because the gain from inputs to outputs is large. Then, the compositions (outputs) are very sensitive to small changes in the flows (inputs), and the input flow measurements are not sufficiently accurate to provide useful information.
- (4) The dynamic response of the measurement is different from that of the outputs and for a static estimator the estimate using input measurements is poor dynamically. In our case, the dynamic lag from inputs to outputs cause the static PCR- and Brosilow-estimators to perform poorly when input flow measurements are used (even in the absence of measurement error).

We will discuss the three last items in more detail. Note that items 3 and 4 provide two different reasons for why we may not want to use flow measurements.

Latent variables

In the PCR method, the available data are smoothed by obtaining a smaller number, k , of "latent variables," t , which are not coupled and contain most of the original information. These latent variables may be written $t = P_i \theta$, where P_i is the projection matrix. The estimator then becomes $\hat{y} = K_i t$ where K_i is a "small" matrix with k parameters for each output (typically k is the number of degrees of freedom, 3 in our examples), and the overparameterization in the regression step is avoided. In the Brosilow method, the (estimated) disturbances may be viewed as latent variables, but this may not be a good choice from a numerical point of view. In the Kalman estimator, the states may be regarded as latent variables. The effect of measurement noise on the states is smoothed by the strong couplings in the model used in the Kalman estimator.

Distillation Column Application

Example column

As an example, we use the distillation column A studied by

Table 1. Data for Distillation Column Example

α	N	N_F	z_F	y_D	x_B	D/F	L/F
1.5	40	21	0.50	0.99	0.01	0.500	2.706
• Liquid feed		• Const. molar flows					
• Ideal VLE		• Pressure 1 atm					
• Holdup on each tray; $M_i/F=0.5$ min							
Antoine parameters:		A	B	C			
Light component		15.8366	2,697.55	-48.78			
Heavy component		15.4311	2,697.55	-48.78			

Skogestad and Morari (1988). The column separates a binary mixture with relative volatility 1.5 and has 40 theoretical stages, including the reboiler, plus a total condenser. Other column data are given in Table 1.

The difference in boiling points of the two pure components is 13°C. In Figure 2 some steady-state temperature profiles for the column are displayed. We note that variations in temperature are small toward the ends of the column. Thus, although for our binary mixture with constant pressure there is a one-to-one relationship between the product mole fraction and the temperature at the corresponding column end, it may be difficult to use these measurements for estimation because they are not sensitive enough to changes. On the other hand, inside the column where the sensitivity is large, changes in feed composition have a large effect on the temperatures even with constant product compositions. This demonstrates the limitations of single temperature control and motivates the use of an estimator where multiple temperatures are used.

In this study, we use a linear model of the column. To simplify the model the liquid holdup is assumed constant, that is, the flow dynamics are neglected. This gives rise to a 41st-order linear model in terms of the mole fraction of the light component on each tray, which yields the following transfer function model:

$$y(s) = G_u(s)u(s) + G_d(s)d(s) \quad (2)$$

$$\theta(s) = F_u(s)u(s) + F_d(s)d(s) \quad (3)$$

Here the inputs are $u = [L, V]^T$, the disturbances are $d = [z_F, F]^T$, the outputs $y = [y_D, x_B]^T$ and the (secondary) measurements are $\theta =$ temperatures on all trays. The dominant time constant (corresponding to the smallest eigenvalue) of the column model is 194 min. At steady state,

$$G_u(0) = \begin{pmatrix} 87.8 & -86.4 \\ 108.2 & -109.6 \end{pmatrix} \quad (4)$$

Here, the gain matrix has been scaled such that the outputs (compositions) are in relative units (an output of magnitude 1 corresponds to 0.01 mole fraction units), and the inputs are scaled relative to a feed rate of $F=1$. The large elements in this matrix should be noted as they make it difficult to make use of input data when estimating product compositions. Such large gain elements are indeed common to all distillation columns with both products of high purity. For example, we find for changes in external flows when we assume that the overall separation in the column is constant (Skogestad and Morari, 1987b):

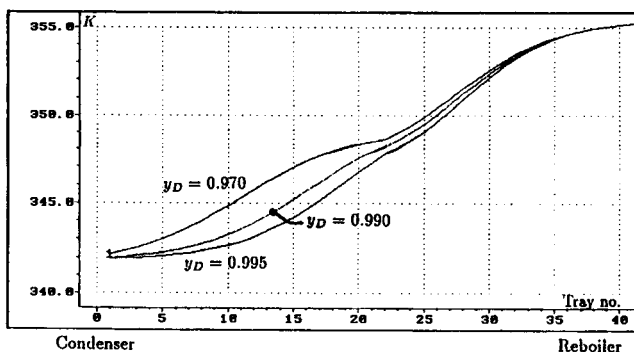
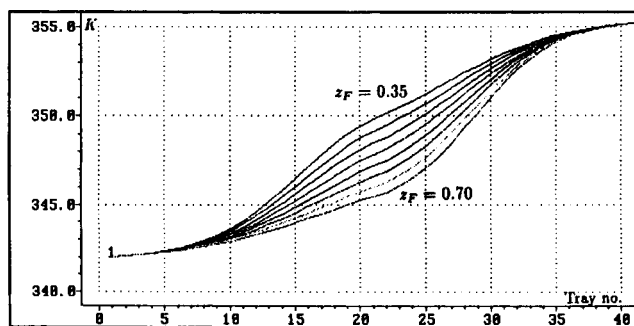


Figure 2. Column temperature profiles.

Upper profiles for different feed compositions when y_D and x_B are held constant. Lower profiles for different top product compositions when z_F and x_B are held constant.

$$\left(\frac{1}{x_B} \frac{\partial x_B}{\partial (L/F)} \right)_v = \left(\frac{\partial \ln x_B}{\partial (L/F)} \right)_v \approx \frac{1}{(B/F)x_B + (D/F)(1-y_D)} \quad (5)$$

Applying this expression to our example where $x_B=1 - y_D=0.01$ yields that all elements in the matrix in Eq. 4 should be 100 in magnitude, which is in good agreement with the exact model.

The condition number of the matrix in Eq. 4 is 141.7, and the diagonal RGA elements are 35.5. This plant is consequently ill-conditioned (Skogestad and Morari, 1987a). The other matrices in Eqs. 2 and 3 are also ill-conditioned and have large elements at steady state.

Estimators

Block diagrams of the three estimators are shown in Figure 3.

Kalman filter

In the Kalman filter scheme (Kalman and Bucy, 1961), a dynamic state space model is used in parallel with the process, and the deviation between the outputs from the process and the model is used as feedback to the model through a filter gain K_f . Write the linear state space model on the following standard form:

$$\dot{x} = Ax + Bu + Ev \quad (6)$$

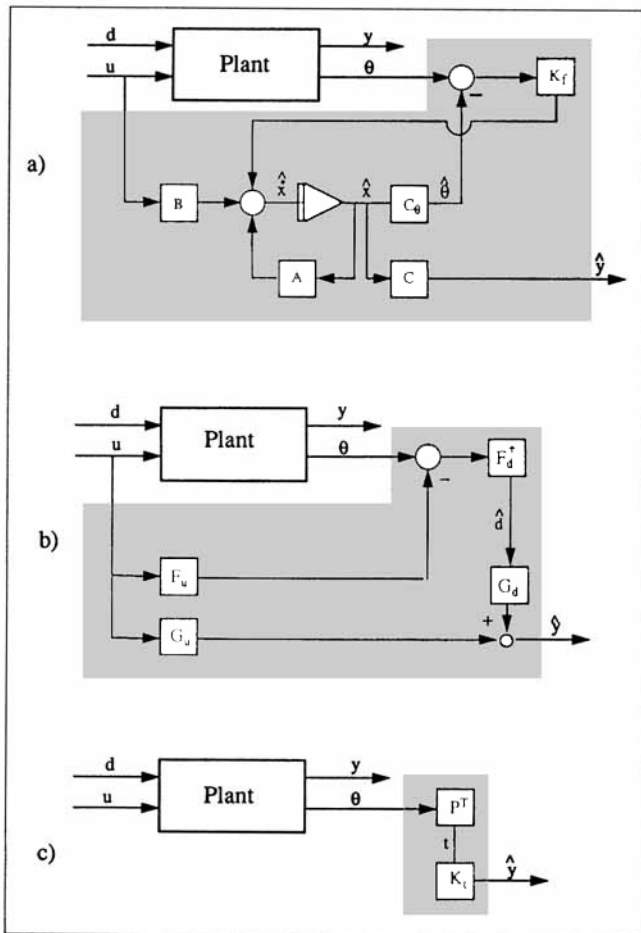


Figure 3. Block diagrams for the three estimators: (a) Kalman filter; (b) static Brosilow estimator; (c) static PCR-estimator based on output measurement only.

$$y = Cx \quad (7)$$

$$\theta = C_\theta x + w \quad (8)$$

Here, v and w are assumed to be white noise processes with covariance matrices V and W . Minimizing the expected variance $\theta - \hat{\theta}$ yields the estimated states:

$$\hat{x} = A\hat{x} + Bu + K_f(\theta - C_\theta\hat{x}) \quad (9)$$

$$= (A - K_f C_\theta)\hat{x} + Bu + K_f\theta \quad (10)$$

where filter gain K_f is:

$$K_f = X C_\theta^T \mathcal{W}^{-1} \quad (11)$$

Here, X , the covariance matrix of \hat{x} , is found from solving the algebraic matrix Riccati equation:

$$AX + XA^T - X C_\theta^T \mathcal{W}^{-1} C_\theta X + E^T V E = 0 \quad (12)$$

The overall Kalman estimator then becomes:

$$\hat{y}(s) = C(sI - A + K_f C_\theta)^{-1} [K_f \theta(s) + Bu(s)] \quad (13)$$

The covariance matrix of the measurement noise W was set to $0.04 I$ for our example column (I is the identity matrix). This corresponds to 0.2°C noise on each temperature. The process noise (disturbance) is here $v^T = [L, V, F, z_F]$ (reflux, boilup, feed rate, and feed composition). Its covariance matrix, V , was assumed diagonal and was varied in order to tune the Kalman filter. Four different values of the variance on L and V were selected (Table 2), and the corresponding filter gain matrices are denoted $K1$ to $K4$. The assumption of white noise process disturbances is somewhat unrealistic in a distillation column, and we might add an integrator and use a nonstationary disturbance $d = (1/s)v$. The estimator, however, is not significantly improved by such changes, although it will remove the steady-state offset for the case with perfect temperature measurements.

Brosilow estimator

The following linear steady-state model of the column in terms of deviation variables is used in the Brosilow estimator (Weber and Brosilow, 1972; Joseph and Brosilow, 1978a):

$$y = G_u u + G_d d \quad (14)$$

$$\theta = F_u u + F_d d \quad (15)$$

where d denotes the disturbances. The matrices above are of course related to those used in the state space description in the Kalman filter. For example, $G_u = -CA^{-1}B$, and for the case $v = d$ we have $F_d = -C_\theta A^{-1}E$. The main idea is now to use Eq. 15 to estimate the disturbances. We get:

$$\hat{d} = F_d^+ (\theta - F_u u) \quad (16)$$

where the pseudoinverse F_d^+ is the optimal inverse in the general least-square sense. Except for the dynamics, the main difference from the Kalman filter is that the presence of measurement noise and input noise/disturbances are neglected.

The inversion in Eq. 16 may be impossible, or at least numerically ill-conditioned, when there are collinearities among the variables (Weber and Brosilow, 1972; Joseph and Brosilow, 1978a). To avoid some of these problems one should obtain the pseudoinverse, F_d^+ , from a singular value decomposition (SVD) of F_d by deleting directions with singular values equal or close to zero (for example, see Strang, 1980, p. 142). Combining Eqs. 14 and 16 yields the Brosilow estimator

$$\hat{y} = K_B \theta + (G_u - K_B F_u) u \quad (17)$$

where

Table 2. Process Disturbance Covariance Matrix of Kalman Filter Gains: $\mathcal{W} = 0.04I$

Case	\mathcal{V}				
	L	V	F	z_F	
K1	diag { 200	200	0.01	0.01	}
K2	diag { 0.10	0.10	0.01	0.01	}
K3	diag { 0.01	0.01	0.01	0.01	}
K4	diag { 0.0	0.0	0.01	0.01	}

$$K_B = G_d F_d^T \quad (18)$$

For our example column $d^T = [z_F, F]$ and $u^T = [L, V]$. The matrices F_d, F_u, G_d and G_u were found by linearizing the model at the nominal operating point.

PCR estimator

The main idea is here to consider directly a regression between the q measurements (θ) and the p outputs (y) to be estimated. The problem is then to obtain the matrix K in

$$\hat{y} = K\theta \quad (19)$$

To this end obtain n "calibration" runs of corresponding values of y and θ , and place these as rows in the matrices $Y^{n \times p}$ and $\Theta^{n \times q}$, respectively. (It may seem more reasonable to place y and θ as columns in the matrices, but we shall here use the standard convention from the statistics literature. This convention is the reason for the reverse order in Eqs. 19 and 20.) If the estimator is perfect we will have:

$$Y = \Theta K^T \quad (20)$$

Given Y and Θ the general least-square solution to Eq. 20 is (for example, Strang, 1981, p. 139)

$$K_{CLS} = Y^T [\Theta^T]^\dagger \quad (21)$$

In addition to minimizing $(y - \hat{y})^2$ this solution minimizes the norm of K . The pseudo inverse is obtained from a SVD of Θ . Using standard notation from the statistics literature (see for example, Wold et al. 1987), the SVD of Θ is written

$$\Theta = t_1 p_1^T + t_2 p_2^T + \dots + t_m p_m^T \quad (22)$$

where $m \leq \min(n, q)$ is the rank of Θ . Here, p_1 is the eigenvector corresponding to the largest eigenvalue of $\Theta^T \Theta$, (the square of the largest singular value of Θ), and p_2 is the eigenvector corresponding to the second largest eigenvalue, and so on. These "loading vectors" (p 's) give the directions of the principal components, while the "scores" (t 's) give the magnitude. If all m terms in Eq. 22 are retained we obtain the generalized pseudoinverse in Eq. 21. However, in PCR we select only those first k principal components that can be distinguished from the measurement noise. Let the matrices $P^{q \times k}$ and $T^{n \times k}$ include only these k most important directions. Define the new latent variables as $t = P^T \theta$. Note that $P^T = P^{-1}$ since P is orthonormal. The least-square solution to $y = K t$ becomes $K_t = Y^T T [T^T T]^{-1}$, and the overall estimator gain matrix becomes:

$$K_{PCR} = Y^T [\Theta_k^T]^\dagger = Y^T T [T^T T]^{-1} P^T \quad (23)$$

In the general case, θ may be replaced by $\tilde{\theta}$ which includes also the inputs and measured disturbances.

The calibration sets for the example column were obtained from a linear steady-state column model. Note that there are only three degrees of freedom for our example column but additional runs were added to better study the effect of measurement noise. Thus, a factorial design method was used to

Table 3. Data to Simulate Stationary Temperature Profile:
 $q_F = 1.0, P = 1.0$ atm

z_f	y_d	x_b	z_f	y_d	x_b
0.4000	0.9810	0.0190	0.4000	0.9810	0.0010
0.4000	0.9990	0.0190	0.4000	0.9990	0.0010
0.6000	0.9810	0.0190	0.6000	0.9810	0.0010
0.6000	0.9990	0.0190	0.6000	0.9990	0.0010
0.4500	0.9855	0.0145	0.4500	0.9855	0.0055
0.4500	0.9945	0.0145	0.4500	0.9945	0.0055
0.5500	0.9855	0.0145	0.5500	0.9855	0.0055
0.5500	0.9945	0.0145	0.5500	0.9945	0.0055

select 16 different runs (sets of corresponding values of θ and y) around the operating point. The independent variables for generating these data were chosen as the outputs y_D and x_B and the feed composition z_F , see Table 3. Since the column conditions in the simulation model are independent of the load, it is not necessary to simulate different feed rates.

It is important to note that this approach allows us to freely vary the outputs (y_D and x_B), and we are thus able to span all directions in the output space. This is different from other approaches, for example, the Brosilow estimator, which is based on an open-loop model in terms of the inputs (L, V, F, z_F), and where the output space may not be properly spanned for ill-conditioned plants with strongly coupled outputs.

When stated random noise of magnitude 0.1°C was added on all temperatures in the calibration set, but the default is no noise. During the regression step the temperature data were reduced to three latent variables and K_{PCR} was computed from Eq. 23.

Analysis of the Estimators

The objective is to evaluate the different estimation methods described above. In this section, we define our criteria and tools for the evaluation. We shall consider both the "open-loop" estimation error, $e_{ol} = y - \hat{y}$, and the "closed-loop" (when the estimate is used for closed-loop control) control error, $e_{cl} = y - y_s$.

μ -analysis

Our main analysis tool is the structural singular value (μ) (Doyle, 1982). In this framework, we rearrange the system to fit the general form shown in Figure 4. Here the interconnection

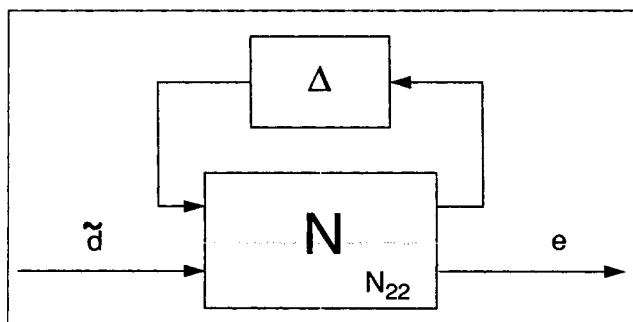


Figure 4. Structure singular value analysis.

e may be the "open loop" estimation error $e_{ol} = y - \hat{y}$, or the "closed-loop" control error, $e_{cl} = y - y_s$.

matrix N includes the plant, the controller, the estimator and the weights. \bar{d} denotes external inputs (disturbances, noise and setpoint changes), and e is the "error" we want to keep small.

We have a separate Δ -block loop to represent the model uncertainty. Weights are used to scale the signals, \bar{d} and e , and the uncertainty Δ to be less than 1. In the μ analysis we evaluate the worst case amplification from \bar{d} to e at each frequency, and the performance requirement for the error, e_{ol} or e_{cl} , is satisfied if μ is less than one at all frequencies. Nominal performance (with $\Delta=0$) is satisfied if $\mu(N_{22}) < 1$, and robust performance (for all allowed Δ 's) is satisfied if $\mu(N) < 1$. In the article, we plot μ as a function of frequency, and estimators with small μ values are preferred. More details about the methods used for μ analysis may be found in Mejdell and Skogestad (1989) and Mejdell (1990).

Uncertainty Weights. The most important source of uncertainty is assumed to be on the inputs (L and V). We shall use the same multiplicative input uncertainty weight as Skogestad et al. (1988), which is given by:

$$w_I(s) = 0.2 \frac{5s + 1}{0.5s + 1} \quad (24)$$

In the low frequency range the weight allows for 20% input gain uncertainty and the uncertainty reaches 100% at about $\omega = 1 \text{ min}^{-1}$.

Weights for External Inputs. The external inputs to the systems (the \bar{d} 's in Figure 4) consist of setpoints, as well as disturbances and noise. The maximum setpoint changes are set to ± 0.01 mole fraction units. The disturbances in the feed rate F and the feed composition z_F are set to 20%, that is, z_F may vary from 0.4 to 0.6 mole fraction units. Noise was generated by adding a constant vector of random values with normal distribution and a standard deviation of 0.2°C to all 41 temperatures. Noise is included in the μ analysis only if stated.

Evaluation criteria

Open-Loop Evaluation (OL). One obvious criteria for evaluating the different estimators is the estimation error $e_{ol} = y - \hat{y}$. The system is assumed to operate under feedback, since this is closer to a real situation than a pure open loop test where it may "drift away." The term "open loop" is still used since the controller uses the actual y , that is, there is no feedback from the estimate \hat{y} . We use single-loop PID controllers since this is the most common choice in practice. The tunings were obtained for optimal robust performance (minimize μ) when the estimate is exact. To make our results less dependent on the controller used, we shall only consider the nominal performance in this test, that is, without any uncertainty. This makes the comparison independent of the robust stability requirement of the system which depends strongly on the controller. The performance weight for the open-loop test is:

$$w_{Pol}(s) = \frac{10}{4s + 1} \quad (25)$$

This weight requires less than 10% estimation error at steady state ($\omega \leq 0.1 \text{ min}^{-1}$). At higher frequencies the weight increases

to one at $\omega = 2.5 \text{ min}^{-1}$. This allows for an error greater than 100% at frequencies above 2.5 min^{-1} .

Closed-Loop Evaluation (CL). The main objective of the estimator is to replace the primary measurement of y with the estimate \hat{y} for feedback control. The error of interest to be minimized, is then the control error, $e_{cl} = y - y_s$, that is, the difference between the actual outputs and the setpoints (reference signals). In this case, we include the uncertainty and consider the robust performance. We use the same controller as for the open-loop comparison. One disadvantage with this test is that using the same PID controller for all estimators will bias the comparison somewhat, since the optimal controller in each case will depend on the estimator used. The performance weight for the closed-loop test is

$$w_{Pcl}(s) = 5 \frac{10s + 1}{100s + 1} \quad (26)$$

This implies that the control error, $y - y_s$, should be within 20% at steady state. Our feedback system should be effective up to about $\omega = 0.05 \text{ min}^{-1}$ and the amplification at high frequencies should never exceed 2. Except for the allowed steady-state offset, this weight is the same as the one used by Skogestad et al. (1988).

Results

Comparison of Kalman filter and static PCR estimator

In Figure 5, we compare the μ -plots of the Kalman and PCR estimators, using 41 temperatures. The first thing to note is how well the simple static estimator $\hat{y} = K_{PCR}\theta$ performs in the closed-loop test. The main reason is that the dynamic responses of the temperatures θ and the compositions y are very similar. This will be the case for most distillation columns, at least for sections of the column, but may of course not be the case for other applications.

In the "open loop" analysis, the Kalman filter is significantly better than the PCR estimator at higher frequencies. This is due to the dynamics included in this estimator. On the other hand, the "closed loop" test shows that the estimators will perform about equally well when used for feedback, and also as well as using perfect measurements. Actually, for some frequencies, the PCR estimator is even better than using perfect measurements. The reason is that the temperatures in the middle of the column generally change slightly faster than at the ends, and the steady-state estimator will therefore have a small inherent "feedforward" effect. The linear simulation responses in Figure 6 confirm that the PCR-estimate is almost equal to the true value. One exception is for feed composition disturbances, where it shows a small inverse response.

The PCR estimator uses only temperatures, but we did also evaluate the effect of adding input flow measurements. However, even at steady state with exact measurements (no model error) the improvement in estimator performance was very small and with measurement noise it was not significant. Furthermore, the dynamic behavior of a static estimator is much worse when inputs are used unless a dynamic lag (low-pass filter) is used on the flow measurement.

Different Kalman filters and use of inputs in estimator

Figure 7 shows μ plots for the Kalman filters obtained using

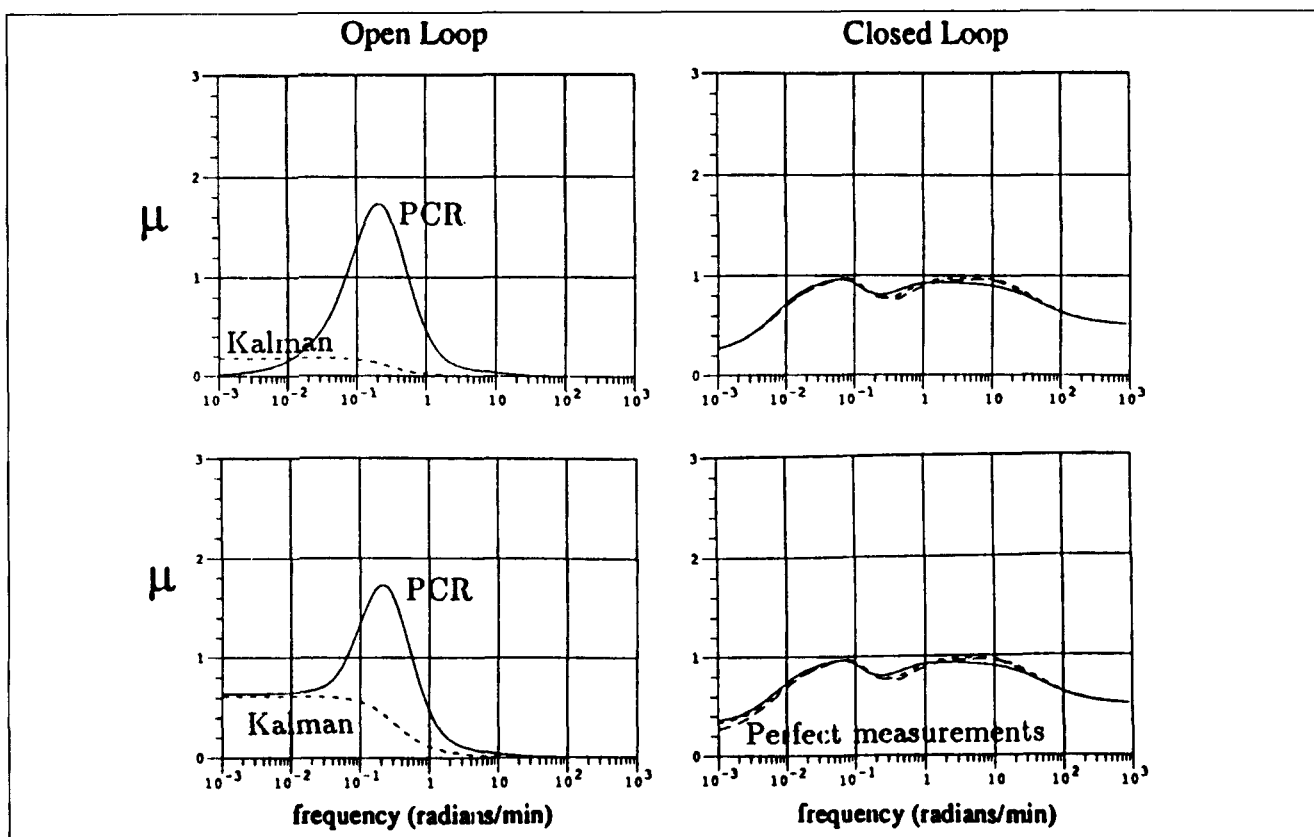


Figure 5. Comparison of Kalman (case K1) and PCR estimator with 41 temperatures.

Upper: without noise in μ analysis; lower: with noise.

the four different levels of process noise on L and V in Table 2. The best Kalman filter, K1, is the one that was compared with PCR above. The remarkable thing with this best estimator is the very large assumed variance on the inputs $u(L$ and $V)$. In effect, this variance is so large that the transfer function from u to \hat{y} in Eq. 13 is approximately zero, that is, the estimator does not use the information about the input signals.

The worst Kalman filter, K4, assumes disturbances (noise) of magnitude 0.1 for F and z_F , but assumes no disturbances on the inputs. This estimator performs reasonably well in the open-loop μ test when there is no uncertainty (left in Figure 7). However, it is extremely poor when input uncertainty is added in the closed-loop test (right). Similar differences are found for the open-loop estimation error when there is uncertainty.

For the case with no measurement noise the steady-state offset of the Kalman Filter may be removed by including non-stationary noise as discussed earlier, but otherwise this change has very little effect.

Brosilow estimator

The μ plots for the Brosilow inferential estimator, Eq. 17, with different numbers of temperature measurements are shown in Figure 8. The estimator performs poorly, and its performance does not improve with increasing number of measurements. The open-loop test shows that the estimator nominally works well at very low frequencies ($\omega < 0.001 \text{ min}^{-1}$). The poor

dynamic performance at intermediate frequencies is due to the fact that the estimator uses the input signals $u(L$ and $V)$, as shown in Eq. 17. As discussed in the introduction, the dynamic behavior of u and the compositions y are very different and using a constant matrix $G_u - K_B F_u$ does not work well. This problem might have been corrected by using a low-pass filter on the inputs with a large time constant of about 194 min (that is, add dynamics to G_u and F_u) (Joseph and Brosilow, 1978b). However, even this estimator would not perform well, since we find that with model uncertainty the performance is poor even at low frequencies (for example, see the closed-loop plot in Figure 8). This is mainly due to the input error, that is, that the actual values of L and V are different from those used by the estimator. As mentioned above, the Kalman filter suffers from the same problem when it is tuned such that it makes use of the input measurements.

A better approach for our case is to regard the inputs (L and V) as additional unknown disturbances. This gives rise to a "modified" Brosilow estimator which uses the temperature measurements only. For the case with perfect model this estimator is identical to the PCR-estimator with no noise. However, because the matrices G_u , G_d , F_u and F_d are ill-conditioned for our example, small modeling errors may in practice cause also this modified estimator to perform poorly (Mejdell, 1990). In particular, this will be the case if the model is obtained experimentally. Brosilow and coworkers discuss the ill-conditioning of F_d caused by using too many measurements (Weber and Brosilow, 1972; Joseph and Brosilow, 1978a), but in ad-

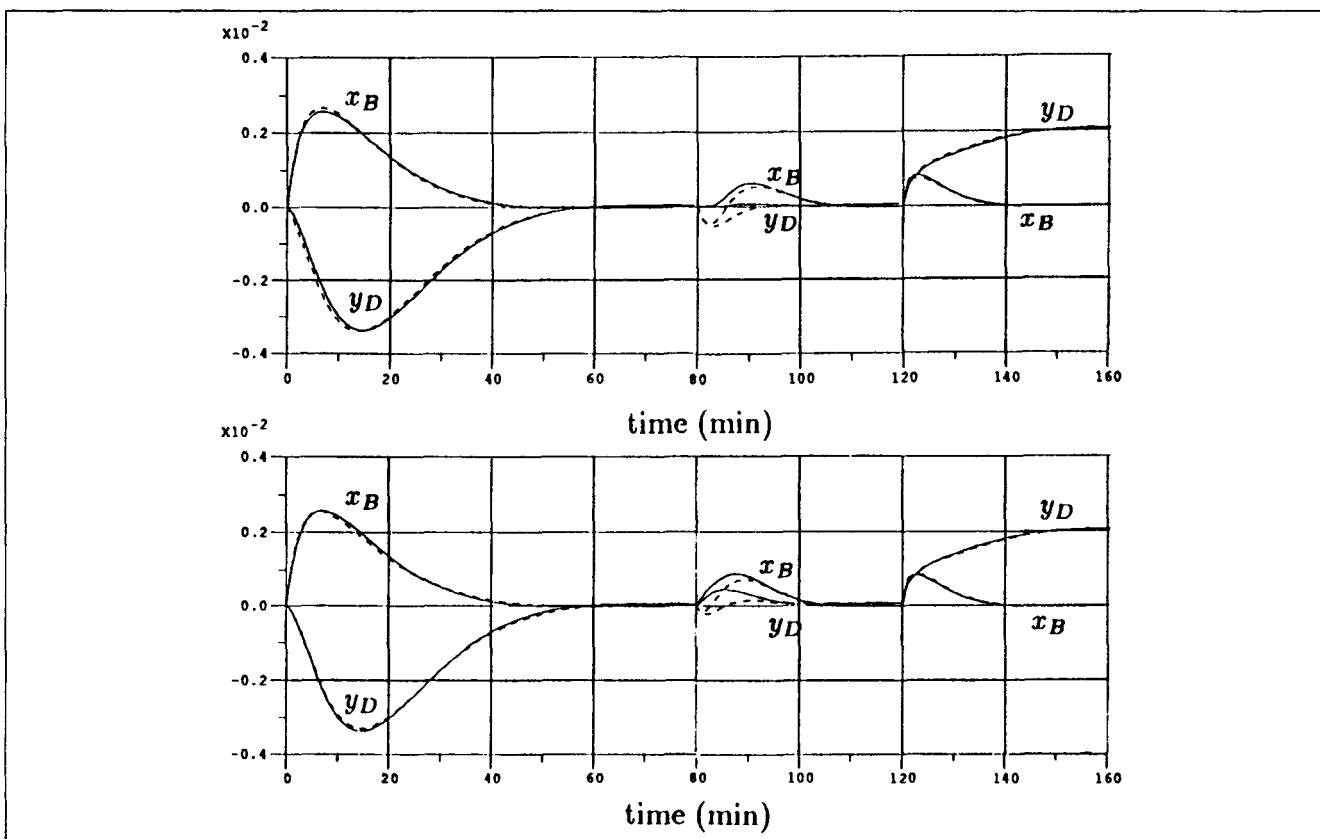


Figure 6. Comparison of actual output $y(t)$ (—) and PCR-estimate $\hat{y}(t)$ (· · · ·).

Responses under feedback control are shown for a 20% increase in feed rate at $t=0$, a 20% increase in feed composition at $t=80$ min, and a setpoint change in y_D at $t=120$ min. The PCR estimate uses all temperature measurements. Upper: y used for feedback control; Lower: \hat{y} used for feedback control.

dition some plants, for example, our high-purity distillation column, are inherently ill-conditioned and the matrices G_u , G_d , F_u and F_d are all ill-conditioned even for the case with few measurements. The PCR-estimator does not suffer from these robustness problems since it deals directly with the relationship between θ and y .

Number of measurements and their location

The μ plots in Figure 9 for the PCR estimator shows the effect of using varying numbers of measurements. (Note that the noise in this case is put on the temperatures in the calibration set and not in the μ analysis.) Figure 9 demonstrates that adding temperature measurements improves the estimates

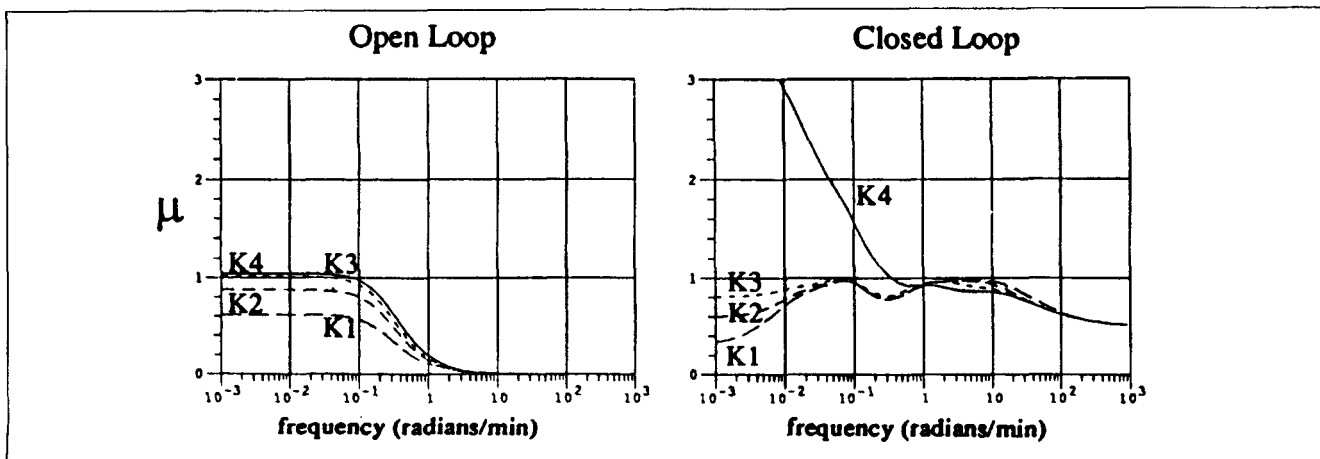


Figure 7. μ plots for different Kalman filter gains (Table 2).

41 temperatures and noise are included in the μ analysis.

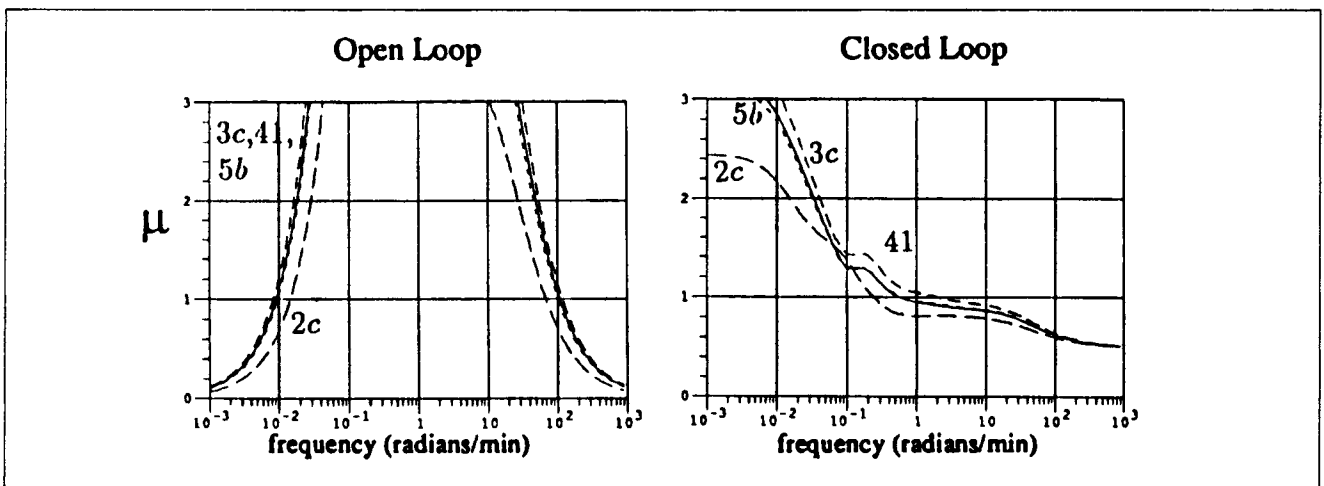


Figure 8. Brosilow inferential estimator for various number of measurements.

No noise. Tray locations: 2c:9, 33; 3c:10, 22, 33; 5b:10, 15, 22, 29, 33.

and the control performance. The main difference is between two and three measurements. Similar results also apply for the Kalman filter.

When using more temperatures than necessary (three for our example column with three degrees of freedom), the additional temperatures will reduce the effect of the measurement noise and, more importantly, make the performance less sensitive to measurement location. Figure 10 clearly illustrates this.

Use of Inputs in Estimator

We found above that when the inputs are used explicitly by the estimator, the Kalman filter (case K4) and the Brosilow estimators are very sensitive to input errors, even at steady state. In this section, we shall provide some simple explanations for these results.

The inputs \hat{u} used for estimation may be either a measurement of the inputs (u_m) or the controller output (u_c) (see Figure 11). To understand the effect of input error, consider a simple estimator $K = G_u$ based only on the inputs:

$$\hat{y} = G_u(s)\hat{u} \quad (27)$$

and the estimation error becomes:

$$y - \hat{y} = G_u(u - \hat{u}) \quad (28)$$

Measured inputs ($\hat{u} = u_m$). In this case, the estimation error becomes:

$$y - \hat{y} = G_u u_n \quad (29)$$

where u_n is the input error (noise). We here assume that the input noise and the outputs have been normalized such that their expected/allowed variations are in the range -1 to 1 . We then note that the estimation error may be very large if the elements in the matrix $G_u(j\omega)$ are much larger than 1 , and in such cases little will be gained by using the measured inputs for estimation. In particular, this applies to ill-conditioned plants under feedback control where the inputs (L and V) are correlated, whereas the measurement noise (μ_n) is uncorrelated,

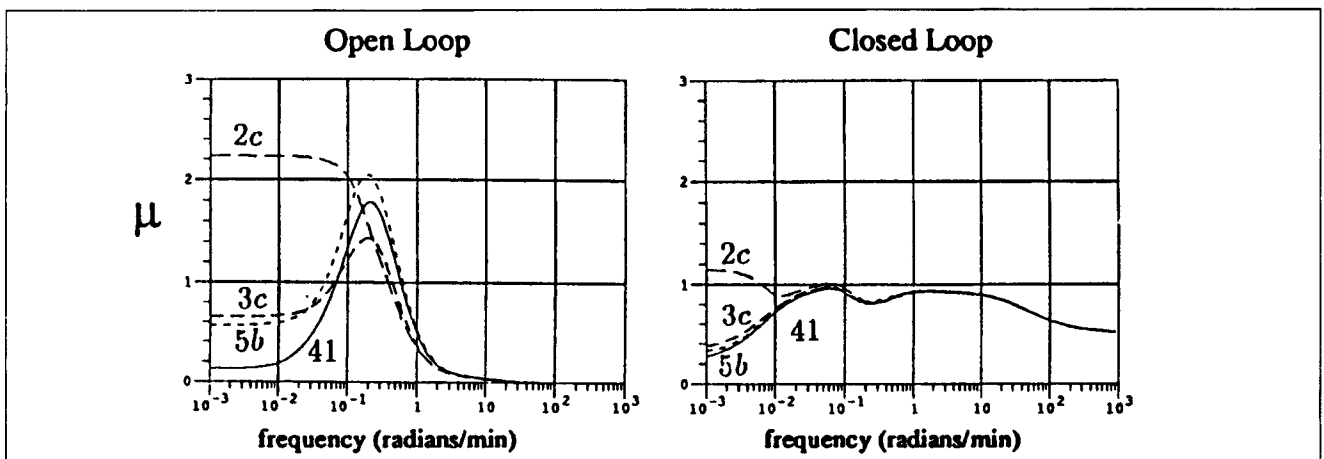


Figure 9. Effect on μ of number of temperatures for PCR estimator.

The temperatures in the calibration set are corrupted with 0.1°C noise. Tray locations: 2c:9, 33; 3c:10, 22, 33; 5b:10, 15, 22, 29, 33.

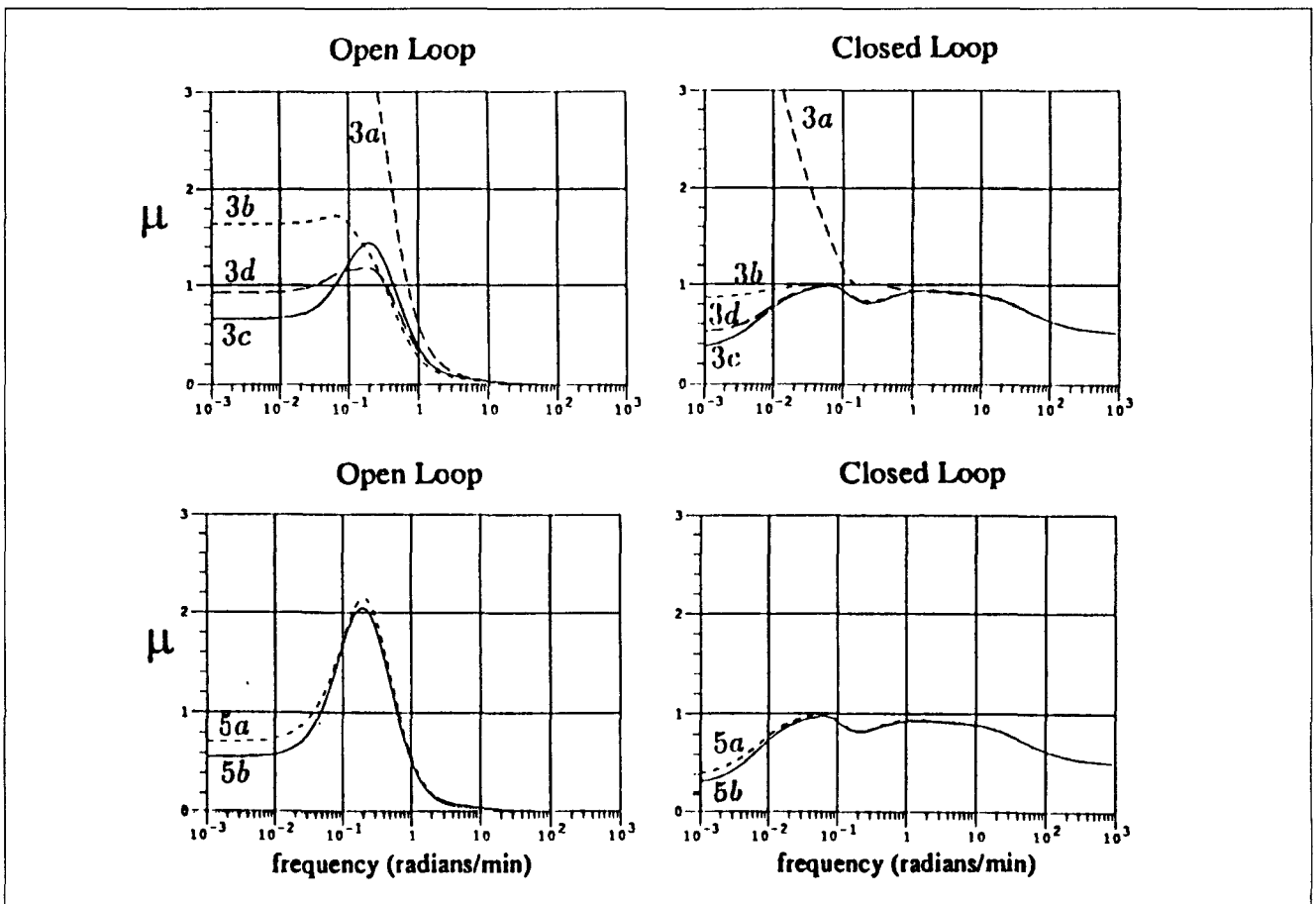


Figure 10. Effect on μ of the measurement location.

PCR with 0.1°C noise in calibration set. Tray locations: 3a:2, 22, 41; 3b:6, 22, 36; 3c:10, 22, 33; 3d:10, 17, 33; 5a:1, 12, 21, 30, 41; 5b:10, 15, 22, 29, 33.

such that the measurement error may easily exceed 100% in the direction of the largest singular value of $G_u \mu$. Specifically, for our example column, the elements in $G_u(0)$ (appropriate scaled) are about 100, and an error in one flow measurement of only 0.01 input unit (0.3% of the total flow in L and V) results in an estimation error of 1.0 output units (0.01 mole fraction units, which is 100% of the nominal impurity).

Controlled outputs ($\hat{u} = u_c$). The actual inputs are:

$$u = (I + \Delta_I) u_c \quad (30)$$

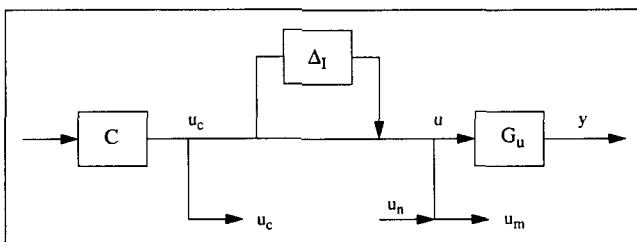


Figure 11. Actual input, u , will differ from value, u_c , computed by the controller and from the measured output, μ_m .

where the uncertainty matrix $\Delta_I = \text{diag}\{\Delta_j\}$ is a diagonal matrix consisting of the relative input errors on each input channel j . The estimation error becomes $y - \hat{y} = G_u \Delta_I u_c$, and using Eq. 28 we get:

$$y - \hat{y} = G_u \Delta_I G_u^{-1} \hat{y} \quad (31)$$

Skogestad and Morari (1987a) found that the i th diagonal element of the term $G_u \Delta_I G_u^{-1}$ is given by $\sum_j \lambda_{ij} (G_u) \Delta_j$, where λ_{ij} denotes the ij th RGA elements. Consequently, in the presence of input uncertainty, the estimation error, $y - \hat{y}$, is likely to be very large for plants with large RGA elements. Note that this result is independent of scaling and of the controller used. The model, Eq. 4, used throughout this article has diagonal RGA values of 35.1 and with an input gain error of 20% ($\Delta_1 = -\Delta_2 = 0.2$) the diagonal elements of $G_u \Delta_I G_u^{-1}$ are about 14. This large value explains the large μ values that were obtained for the Kalman K4-estimator for the "closed-loop" case in Figure 7 which includes input uncertainty.

To avoid such large RGA-values one may consider using as inputs, for example, D and V (DV configuration) rather than L and V . This follows since all RGA elements in this case are less than 1, and the sensitivity to input uncertainty may be less. However, in this case the gain elements for changes in D

will be large as well (after appropriate scaling), and the use of D will be sensitive to input disturbances/noise.

The conclusion for our distillation column is to base the estimate on temperature measurements only. Flow measurements do not improve the estimate significantly because of sensitivity to small measurement errors, poor dynamic response when used in a static estimator, and the fact that the temperatures contain so much information that the estimate is not significantly improved even at steady state with no model error. The above arguments apply to a system with constant tray efficiencies. In practice, the tray efficiency may vary with column load, and improved estimates may be obtained by using flow measurements to infer the efficiency.

Discussion

Kalman filter

Model uncertainty is not included explicitly when obtaining the Kalman filter, and it may require physically unrealistic values of the noise weights, \mathcal{V} and \mathcal{W} , to obtain the best estimator when uncertainty is included. This is illustrated by the large value needed for noise (disturbances) on the inputs to obtain the best Kalman filter, K1. Otherwise, the Kalman Filter performed well in the μ tests and was undoubtedly the best estimator in the open-loop μ test. The main reason is its inherent dynamic model. Furthermore, it is flexible, because of the weights, and it may be tuned to perform well for ill-conditioned plants as well. In this article, we have used a dynamic Kalman filter, but separate results show that for our application the static version, with $s = 0$ in Eq. 13, is very similar to the static PCR estimator.

Brosilow estimator

The Brosilow inferential estimator performs poorly for our example plants mainly because it uses information about the manipulated inputs u . One may consider a modified Brosilow estimator that treats the inputs as disturbances as discussed earlier. However, even in this case the method relies on first estimating the disturbances, and for ill-conditioned plants this approach is poor numerically. Rather, a direct regression between the measurements (θ) and the outputs (y) for the PCR estimator is preferred.

Principal component regression (PCR) estimator

The PCR estimator does not have the same weaknesses as the Brosilow estimator. First, the estimator used here does not use the input values. Second and more importantly, its numerical properties are much better. The matrix to invert in PCR, the score matrix T in Eq. 23, is generally much better conditioned than F_d used for the Brosilow estimator. For example, for our column the condition number of T is 4.7, whereas the condition number of F_d is 50. To get a well-conditioned T one must ensure that excitations of the weak directions are included in the calibration set. To ensure such excitations, one should use data with specified outputs (Table 3). One should be careful about using open-loop data such as step responses, which will excite only the strong directions. (The gain matrices in Brosilow's scheme are typically obtained from such excitations.)

Simple static PCR-estimators should perform well also for other applications with multiple measurements when the measurement dynamics are similar to those of the outputs to be estimated. Conceptually, it is simple to generalize the static estimator to obtain a dynamic estimator. This may be done by using the PCR method to derive the model relating time series data for θ and y .

Partial least-square (PLS) estimator

The PLS estimator is an alternative regression estimator, which also takes into account the directions in Y when finding the approximate pseudo inverse of Θ (Höskuldson, 1988). In the PLS method, this is done by considering the eigenvalues of $\Theta^T Y Y^T \Theta$ rather than of $\Theta^T \Theta$ used in PCR. This takes into account the directions in Θ which have the largest covariance with Y , and thus ensures that these directions are not deleted. For the linear distillation example studied in this article, the PCR and PLS methods gave almost identical results. When nonlinear data were used, however, we found PLS to be somewhat better (Mejdell and Skogestad, 1991a).

Obtaining and implementing the estimators

Both the Kalman filter and the Brosilow estimator require a linear open-loop model. On the other hand, the PCR approach only deals with the data which may or may not be obtained from a model. Using the data is an advantage, especially when experimental data are used, but also when we do have a good simulation model, as in this article, since we save a significant effort in obtaining the linear model matrices.

To obtain the Kalman filter one must specify weighting matrices for noise and disturbances. These may be difficult to determine *a priori*, especially since the best value of these weights may not be physically meaningful. We experienced no numerical problems in obtaining the Kalman filter gain matrix, even for our high order system with 41 states. The Brosilow estimator has the advantage of having essentially no tuning parameters, but this makes it inflexible and does not work well for ill-conditioned plants. Although not discussed here, the PCR/PLS estimator depends strongly on the scaling of the variables. These scalings are then effective tuning parameters, which are used primarily to reflect the measurement noise. No variable scaling was applied here.

As for implementation, a static estimator is, of course, much simpler than a dynamic estimator.

μ -analysis of estimators

The structured singular value, μ , is a powerful tool for comparing multivariable linear systems with unknown disturbances and uncertainty, without having to perform a large number of simulations. Since μ is a worst-case measure, this tool determines explicitly the weak spots in a system. For example, it would have been much more difficult to determine the estimators' sensitivity to input uncertainty from simulations. However, the test requires additional modeling effort to capture the uncertainty in an adequate way.

When comparing different estimators, both the performance in the "open loop" (estimation error) and in the "closed loop" (control error) should be considered. One disadvantage with

the closed-loop test is that it may depend strongly on the controller chosen.

While using the μ analysis we encountered problems with how to include measurement noise. Modeling it as independent disturbances would give a worst-case combination which would be extremely unlikely to occur when there are many temperature measurements. Therefore, in the μ analysis we added the noise as $n = kn_0$, where n_0 is a constant random vector, k is a scalar, and the noise is assumed equal for all frequencies $|k(j\omega)| < 1$. This approach works well when comparing estimators with the same location and number of measurements. In other cases, however, the specific value of the random numbers in the noise vector n_0 may be important and may bias the μ values. When comparing various PCR estimators we therefore did not include noise in the μ analysis, but rather included noise in the calibration sets.

Although the μ analysis has some difficulties of representing noise, it was found to be most suitable for studying the performance of the estimators for this ill-conditioned plant.

Nonlinearity and real implementation

All models used in this article are linear. This simplifies the problem and is necessary for using the μ analysis. Also, a system that does not perform well in the linear case will generally not perform well in the nonlinear case, and a linear study is therefore a good first step in a performance evaluation. Nevertheless, distillation columns are known to be strongly nonlinear, so the effect of nonlinearity should be taken into consideration. Mejdell and Skogestad (1991a,b) show that with both simulations and implementation on a pilot-plant column, a linear PCR/PLS estimator works very well also for real nonlinear distillation columns. Logarithmic transforms on the temperatures were used to reduce the nonlinearity.

We have also implemented a linear static PLS regression estimator on an industrial C_2 - C_3 -splitter and have obtained excellent results. The estimator was based on historical data using ten temperature measurements as well as a measurement of the bottom flow. The flow measurement was filtered through a large dynamic lag and was included mainly to take into account variations in the column behavior with load. Logarithmic transforms on the compositions were used to reduce the effect of nonlinearity. Another implementation has been performed on an industrial butane-splitter using nine temperature measurements. For both these implementations it was found that the static estimator had the additional advantage of providing values that are a few minutes ahead of the actual compositions. (The reason for this is that the internal tray temperatures respond somewhat faster than those at the column ends.) Similar results are reported for the pilot-plant column studied by Mejdell and Skogestad (1991b). Pilot-plant data on the same column also show that temperature measurements on the external column wall, which are much cheaper to implement, may give very good results when experimental data are used to obtain a PCR/PLS estimator.

Conclusions

We found for our distillation example that one can achieve remarkably good performance with the static PCR estimator, and almost as good as a Kalman filter based on a detailed dynamic model.

From a theoretical point of view, one may always improve the estimate by the use of additional information (measurements). In some cases, however, the value of additional information may be minor and must be traded off against the cost of obtaining the measurements and the increased risk of failures.

In distillation columns, one usually should use all the available temperature measurements. More temperature measurements make the estimator less sensitive to nonoptimal sensor locations and measurement noise. For our high-purity distillation example, the addition of input (flow) measurements is not recommended, since it improves the estimate only insignificantly, and with a static estimator the dynamic estimate is deteriorated.

The reason why the static PCR estimator performed so well is that there is a very close relationship, also dynamically, between the temperature measurements (secondary outputs) and composition (primary output). Another suitable application with multiple correlated measurements is the use of temperature profile data from chemical reactors.

Notation

- d = disturbances (= $[z_F, F]^T$ in most cases)
- \tilde{d} = external inputs in μ -analysis
- D = distillate flow rate
- e_{cl} = $y - y_s$, "closed-loop" control error
- e_{ol} = $y - \hat{y}$, "open-loop" estimation error
- F = feed flow rate
- F_u, F_d = gain matrices from inputs (u and d) to secondary measurements (θ)
- G_u, G_d = gain matrices from inputs (u and d) to primary outputs (y)
- K = estimator block
- L = reflux flow rate
- PCR = principal component regression
- RGA = relative gain array, = $G(j\omega) \times [G^{-1}(j\omega)]^T$ where \times denotes element-by-element multiplication
- u = manipulated inputs (= $[L, V]^T$)
- v = process noise (disturbances) = $[L, V, F, z_F]^T$
- \mathcal{V} = process noise covariance matrix
- V = boilup rate from reboiler
- \mathcal{W} = measurement noise covariance matrix
- x = state vector
- x_B = mole fraction of light component in bottom product
- y = primary output vector = $[y_D, x_B]^T$
- \hat{y} = estimated primary outputs
- y_D = mole fraction of light component in distillate
- y_s = setpoint or reference signal for y
- z_F = mole fraction of light component in feed

Greek letters

- α = relative volatility
- Δ = uncertainty block
- θ = secondary measurements (temperature vector)
- $\tilde{\theta}$ = vector of all available information = $[\theta, u]$
- Θ = data matrix of θ
- μ = structured singular value

Literature Cited

- Doyle, J. C., "Analysis of Feedback Systems with Structured Uncertainties," *IEEE Proc.*, **129D**, 242 (1982).
- Höskuldsson, A., "PLS Regression Methods," *J. of Chemometrics*, **2**, 211 (1988).
- Joseph, B. and C. B. Brosilow "Inferential Control of Processes: I. Steady State Analysis and Design," *AIChE J.*, **24**, 485 (1978a).
- Joseph, B., and C. B. Brosilow, "Inferential Control of Processes:

3. Construction of Optimal and Suboptimal Dynamic Estimators," *AIChE J.*, **24**, 485 (1978b).
- Kalman, R. E., and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *Trans. ASME, J. Basic Eng.*, Ser. D83, 95 (1961).
- Keller, J. P., and D. Bonvin, "Selection of Inputs for the Purpose of Model Reduction and Controller Design," *World Cong. on Aut. Cont. (IFAC)*, Munich (1987).
- Mejdell, T., "Estimators for Product Composition in Distillation Columns," PhD Thesis, University of Trondheim—NTH (1990).
- Mejdell, T., and S. Skogestad, "Estimate of Process Outputs from Multiple Secondary Measurements," *Proc. Amer. Cont. Conf.*, Pittsburgh, 2112 (1989).
- Mejdell, T., and S. Skogestad, "Estimation of Distillation Compositions from Multiple Temperature Measurements using PLS Regression," *Ind. Eng. Chem. Res.*, **30**, 2543 (1991a).
- Mejdell, T., and S. Skogestad, "Composition Estimator in a Pilot Plant Distillation Column Using Multiple Temperatures," *Ind. Eng. Chem. Res.*, **30**, 2555 (1991b).
- Moore, C., J. Hackney, and D. Canter, "Selecting Sensor Location and Type for Multivariable Processes," *Proc. of the Shell Cont. Workshop*, D. M. Pretz and M. Morari, eds., Butterworths, Boston (1987).
- Morari, M., and G. Stephanopoulos, "Optimal Selection of Secondary Measurements within the Framework of State Estimation in the Presence of Persistent Unknown Disturbances," *AIChE J.*, **26**, 247 (1980).
- Patke, N. G., and P. B. Deshpande, "Experimental Evaluation of the Inferential System for Distillation Control," *Chem. Eng. Commun.*, **13**, 343 (1982).
- Skogestad, S., and M. Morari, "Implication of Large RGA-Elements on Control Performance," *Ind. & Eng. Chem. Res.*, **26**(11) 2121 (1987a).
- Skogestad, S., and M. Morari, "A Systematic Approach to Distillation Column Control," *Distillation 87, Brighton, I. Chem. E. Symp. Ser.*, **104**, A71 (1987b).
- Skogestad, S., and M. Morari, "Understanding the Dynamic Behavior of Distillation Columns," *Ind. & Eng. Chem. Res.*, **27**(10) 1848 (1988).
- Skogestad, S., M. Morari, and J. C. Doyle, "Robust Control of Ill-Conditioned Plants: High-Purity Distillation," *IEEE Autom. Cont.*, **33**(12), 1092 (1988).
- Strang, G., *Linear Algebra and its Applications*, Harcourt Brace Jovanovich, New York (1980).
- Weber, R., and C. Brosilow, "The Use of Secondary Measurements to Improve Control," *AIChE J.*, **18**, 614 (1972).
- Wold, S., K. Espensen, and P. Geladi, "Principal Component Analysis," *Chemometrics and Intel Lab Syst.*, **2**, 37 (1987).
- Yu, C. C., and W. L. Luyben, "Control of Multicomponent Distillation Columns using Rigorous Composition Estimators," *I. Chem. E. Symp. Ser.*, **104**, 29 (1987).

Manuscript received Aug. 24, 1992, and revision received Dec. 28, 1992.