Frequency-Domain Closed-Loop Identification of Multivariable Systems for Feedback Control

Wei Li and Jay H. Lee
Dept. of Chemical Engineering, Auburn University, Auburn, AL 36849

A frequency-domain identification method for multivariable systems is presented. The method emphasizes accurate identification of the process's gain directionality that is shown to be important for feedback control. The method requires closed-loop testing in addition to conventional open-loop testing, and combines open-loop and closed-loop data such that the derived model matches both the system's frequency-response matrix and its inverse accurately. The method is used to identify a model for a nonideal high-purity distillation column and is shown to capture the gain directionality of the process more accurately than the conventional method.

Introduction

The use of model-based control in the process industries has proliferated recently (Garcia et al., 1989), due to the demand for higher efficiencies. The incentive for using model-based control is especially high in multivariable control problems, for which the effectiveness of proportional integral derivative (PID) control is severely limited due to issues like variable interaction, constraints, and the time-varying, competing nature of control objectives. The model-based control paradigm offers a systematic framework to design a multivariable control system in the face of these issues. However, a "good" model is an essential ingredient of the paradigm. While intensive research has yielded a plethora of design algorithms, the same kind of progress and effort have been lacking on the subject of obtaining multivariable models for feedback control.

Since any reasonable model used in control is likely to be a gross approximation of a real process, it is important that model parameters are adjusted to fit those aspects of the process that are important for feedback control. For example, in the single-input, single-output (SISO) case, the critical information to match is the frequency response around the desired crossover frequency. Errors in the low-frequency region are less visible under feedback due to the high loop gain applied at these frequencies (so long as the sign of the gain is correct). This conclusion carries over to the multiple-input, multiple-output (MIMO) case, but not in terms of individual transfer functions. This is because, when a model-based controller is applied with feedback, certain combinations of errors in the SISO models have a much more profound effect on the closed-loop behavior than other combinations of errors of similar magnitudes. Hence, accuracy of each individual frequency response is not a good measure for assessing the quality of a model to be used in feedback control. Another explanation is that a multivariable model used in feedback control should accurately predict output responses for all input directions, not just when inputs are changed one at a time. The dependence of the system gain on the input direction will be referred to as the "gain directionality" hereafter. The point to be made is that, for highly interactive systems, insignificant errors in the individual SISO models can cause substantial errors in the gain directionality.

A conventional way to construct a multivariable model is to identify each transfer function independently and combine the SISO models into a MIMO model. When SISO models are used together for feedback control, however, small errors in the individual SISO models can cause significant performance degradation and even instability. Realization of this has led to a series of publications over the last several years (Andersen et al., 1989, 1992a,b; Koung and MacGregor, 1993; Li and Lee, 1996; Jacobsen, 1994; Varga and Jorgensen, 1994). The first significant conclusion to be drawn from these studies is that, in order to identify the gain directionality correctly, one must perform multiple-input, single-output (MISO) or MIMO identification, wherein inputs are changed simultaneously instead of one at a time. While MISO identification fits parameters for different outputs independently, MIMO identification identifies the entire MIMO model simultaneously, and therefore utilizes the output correlations as well as the input correlations in the data.

Correspondence concerning this article should be addressed to J. H. Lee.

The second significant conclusion from these studies is that one must use data that are informative in all system directions. Conventionally designed pseudorandom binary sequence (PRBS) signals applied to highly interactive systems in an open-loop manner give output data that are unbalanced, that is, nearly colinear along the high-gain directions. High-gain directions here refer to the input–output directions corresponding to relatively large gains and are defined by the singular value decomposition (SVD) of the gain matrix or the frequency-response matrix. This implies that (1) such data generally suffer from a poor signal-to-noise ratio along the low-gain direction, and (2) model bias is likely to affect the fit in the low-gain direction much more than the fit in the high-gain direction, if the parameters are determined based on the prediction error minimization. Koung and MacGregor (1993) proposed a geometric approach in which inputs for different orthogonal directions (defined by the prior known input singular vectors) are designed according to the ratio of the respective gain (defined by the singular values). This gives data that are evenly excited along the orthogonal directions of the output space. Several researchers also suggested that closing one or more loops during the identification makes it easier to excite all the directions of the outputs (Andersen and Kummel, 1992b; Koung and MacGregor, 1993; Jacobsen, 1994). Li and Lee (1996) provided a formal analysis of this, based on the idea that dynamics for the low-gain directions of a system are strongly reflected in the inverse of the gain matrix and a SISO testing with remaining outputs under closed-loop control enables direct estimation of the inverse. Varga and Jorgensen (1994) proposed on-line adaptation of model parameters through recursive identification as an alternative. Others proposed weighting the prediction error through prefiltering and postfiltering so that model bias would be distributed according to the closed-loop requirement (Gaikwad and Rivera, 1994).

In this article, we draw upon the just-mentioned insights to develop a frequency-domain identification method for multivariable, dynamic systems. Since gain directionality of a dynamic system is defined on a frequency-by-frequency basis, it is convenient to pose the gain directionality estimation problem in the frequency domain. The results from our previous work (when extended to other frequencies) state that a model with accurate gain directionality should match both the real plant's frequency response and the matrix inverse closely. In order to develop such a model, we propose to identify the frequency-response matrix and its inverse through open-loop and closed-loop experiments, respectively (i.e., under conditions that yield informative data). Finding a nonparametric transfer matrix that best fits both frequency-response data can be formulated as a frequency-by-frequency least-squares problem. The frequency-response model so obtained can be converted either into an impulse-response model or into a parametric model. It is also possible to fit a low-order parametric model directly to the frequency-response data. In this case, weighting between the two frequency-response data determines the bias distribution between the high-gain directions and low-gain directions.

The method is used to identify a linear, dynamic model for a nonideal, high-purity distillation column studied by Chien and Ogunnaike (1994). The resulting model is then compared with the models obtained by Chien and Ogunnaike in terms of both gain directionality and closed-loop performance.

Impact of Gain Directionality Fit on Closed-Loop Performance

Before we propose an identification method, we present a simple example designed to illustrate the importance of gain directionality fit in feedback control and to draw insights helpful for further development of this article.

Illustrative examples

Consider the following single time-constant model of a 41-tray high-purity distillation column (see Skogestad et al. (1988) for detailed specifications of the column):

$$\begin{bmatrix} y'_D(s) \\ x'_B(s) \end{bmatrix} = \underbrace{\frac{1}{194s+1}\begin{pmatrix} 0.8747 & -0.8610 \\ 1.0853 & -1.0990 \end{pmatrix}}_{G_1(s)} \begin{bmatrix} L'(s) \\ V'(s) \end{bmatrix}. \quad (1)$$

The preceding model relates changes in the reflux and boilup flow rates to changes in the top and bottom product compositions. The process is highly interactive (the 1,1 element of the relative gain array (RGA) of the preceding gain matrix is 35.80), and the control problem has served as a benchmark problem for identification and robust control.

Let us suppose that one obtained the following gain matrix from a SISO identification:

$$\begin{bmatrix} y'_D(s) \\ x'_B(s) \end{bmatrix} = \underbrace{\frac{1}{194s+1}\begin{pmatrix} 0.8591 & -0.8668 \\ 1.1059 & -1.0862 \end{pmatrix}}_{G_2(s)} \begin{bmatrix} L'(s) \\ V'(s) \end{bmatrix}. \quad (2)$$

Comparing Eq. 2 with Eq. 1, we note that relative errors for the individual gain elements are quite small. When we consider the effect of nonlinearity and various disturbances/measurement noise occurring during a typical identification experiment, it is difficult to imagine obtaining a model that is more accurate on an element-by-element basis. On the other hand, closed-loop simulation reveals that a dynamic matrix control (DMC) controller designed using this model gives an unstable closed-loop system regardless of the choice of tuning parameters (Li and Lee, 1996).

Next suppose that SISO identification gave exact steady-state gains, but individual time constants were off by a little bit:

$$\begin{bmatrix} y'_D(s) \\ x'_B(s) \end{bmatrix} = \underbrace{\begin{pmatrix} \dfrac{0.8747}{197.9s+1} & \dfrac{-0.8610}{190.1s+1} \\ \dfrac{1.0853}{190.1s+1} & \dfrac{-1.0990}{190.1s+1} \end{pmatrix}}_{G_3(s)} \begin{bmatrix} L'(s) \\ V'(s) \end{bmatrix}. \quad (3)$$

The problem of SISO identification giving an inconsistent number of dominant poles is well documented in Jacobsen and Skogestad (1994). Figure 1 compares the frequency response of the preceding model with that of the plant in Eq. 1. The fit is excellent; however, closed-loop simulation (with a DMC controller tuned in a usual manner) again showed instability.

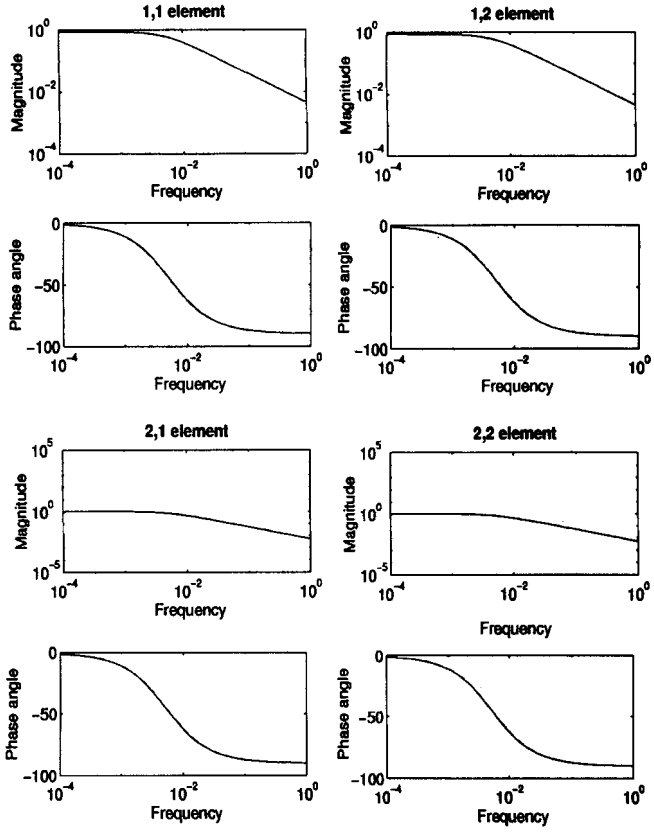

Figure 1. Frequency responses of the plant G_1 and the model G_3.

Solid—G_1; dash—G_3.

Analysis

The preceding example elucidates the shortcomings of conventional SISO identification in developing a MIMO model. Putting emphasis on the accuracy of individual elements does not lead to good estimation of gain directionality in general. To understand the reason for this, let us perform some analysis of the preceding example.

Gain directionality is best analyzed using SVD, which shows the gains and corresponding output directions for different orthogonal input directions. The SVDs of the theoretical model G_1 and the identified model G_2 are given by

$$\underbrace{\begin{pmatrix}0.8747 & -0.8610\\ 1.0853 & -1.0990\end{pmatrix}}_{G_1(0)} = \underbrace{\begin{pmatrix}0.6221 & -0.7829\\ 0.7829 & 0.6221\end{pmatrix}}_{U}\underbrace{\begin{pmatrix}1.9728 & 0\\ 0 & 0.0136\end{pmatrix}}_{\Sigma}\underbrace{\begin{pmatrix}0.7065 & -0.7077\\ -0.7077 & -0.7065\end{pmatrix}}_{V^T} \quad (4)$$

$$\underbrace{\begin{pmatrix}0.8591 & -0.8668\\ 1.1059 & -1.0862\end{pmatrix}}_{G_2(0)} = \underbrace{\begin{pmatrix}0.6186 & -0.7857\\ 0.7857 & 0.6186\end{pmatrix}}_{\tilde{U}}\underbrace{\begin{pmatrix}1.9728 & 0\\ 0 & 0.0129\end{pmatrix}}_{\tilde{\Sigma}}\underbrace{\begin{pmatrix}0.7098 & -0.7044\\ 0.7044 & 0.7098\end{pmatrix}}_{\tilde{V}^T} \quad (5)$$

Comparing Eq. 5 with Eq. 4, we find that both singular values of the plant are well matched by the model. However, the input singular vector for the low-gain direction shows what is essentially a 180-deg mismatch, implying a sign error in the gain for this direction. It is a well-known fact that a sign mismatch between a model and a plant leads to closed-loop instability when a controller with integral action is used. Note that the sign mismatch in the low-gain direction contributes very little to the errors in the individual gain elements. Therefore, open-loop SISO identification does not give a correct sign for the low-gain direction on a consistent basis.

Next, let us examine the SVD of model G_3. Since the steady-state gains are matched perfectly, the gain directionality is also matched perfectly at the zero frequency. On the other hand, the errors in the pole location give significant mismatch in the high-frequency region. For instance, the SVDs of G_1 and G_3 at $\omega = 0.1$ are, respectively,

$$G_1(j\omega)|_{\omega=0.1} = \begin{bmatrix}0.031-j0.614 & 0.177-j0.769\\ 0.041-j0.788 & -0.138+j0.599\end{bmatrix} \times \begin{bmatrix}0.1028 & 0\\ 0 & 0.0003\end{bmatrix}\begin{bmatrix}0.701 & -0.713\\ -0.713 & -0.701\end{bmatrix}$$

$$G_3(j\omega)|_{\omega=0.1} = \begin{bmatrix}0.032-j0.621 & 0.040-j0.782\\ 0.040-j0.782 & -0.032+j0.621\end{bmatrix} \times \begin{bmatrix}0.1016 & 0\\ 0 & 0.0007\end{bmatrix}\begin{bmatrix}0.7065 & -0.7077\\ 0.7077 & 0.7065\end{bmatrix}.$$

The mismatch in the weak direction should be reflected strongly in the inverse of the gain matrix and frequency-response matrix. Indeed, the inverses of the gain matrices for G_1 and G_2 are very different, as shown below:

$$\begin{pmatrix}0.8747 & -0.8610\\ 1.0853 & -1.0990\end{pmatrix}^{-1} = \begin{pmatrix}40.928 & -32.065\\ 40.418 & -32.575\end{pmatrix} \quad (6)$$

$$\begin{pmatrix}0.8591 & -0.8668\\ 1.1059 & -1.0862\end{pmatrix}^{-1} = \begin{pmatrix}-42.697 & 34.073\\ -43.471 & 33.770\end{pmatrix}. \quad (7)$$

We have also plotted the reciprocal of the elements of G_1^{-1} and G_3^{-1}, as a function of frequency in Figure 2. Any significant error in the gain directionality should show up as substantial mismatch between the frequency plots. Indeed, as the frequency increases, we see the increasing mismatch between them.

One conclusion from the preceding analysis is that, while the elements of the open-loop frequency-response matrix are sensitive only to errors in the high-gain directions, its inverse is sensitive to any change in the low-gain directions. Hence, accurate SISO estimation of the frequency response, and the matrix inverse implies accurate estimation of the gain directionality.

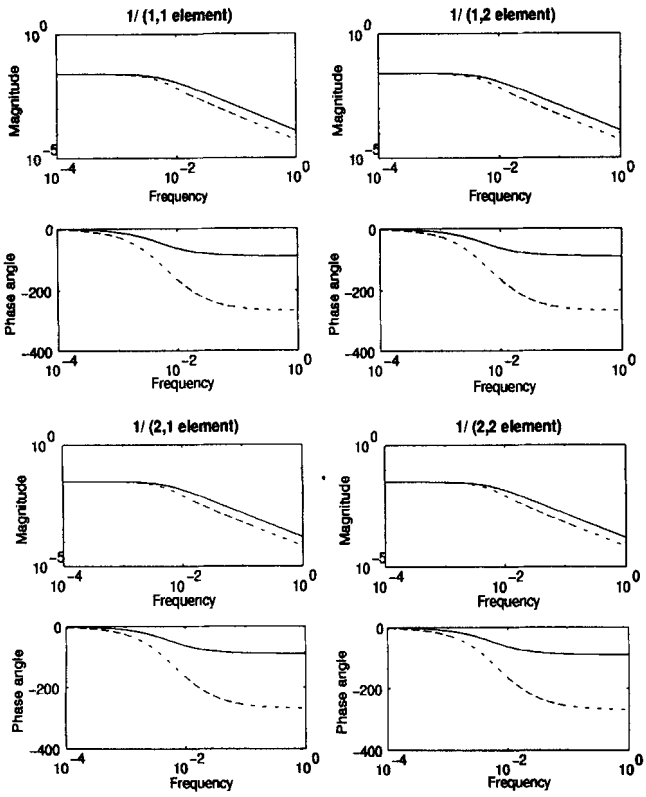

Figure 2. Inverse of the frequency response matrices for the plant G_1 and the model G_3.

Solid—reciprocal of the elements of G_1^{-1}; dash—reciprocal of the elements of G_3^{-1}.

Brief Survey of Identification Methods for Multivariable Systems

MISO/MIMO identification methods

Most identification methods in the literature are developed and discussed in the context of SISO systems. The inadequacy of the open-loop SISO identification for highly interactive MIMO systems has been well documented by several authors (Andersen et al., 1989, 1992a,b; Li and Lee, 1996; Jacobsen, 1994). The main point to be drawn from the literature is that, when the identification data are obtained by perturbing one input at a time, estimation of a low-gain direction is greatly affected by both the low signal-to-noise ratio and model bias. This is because, in the output data obtained in such a manner, perturbations along the high-gain directions tend to dominate. In order to ensure that low-gain directions are estimated with equal accuracy, it is essential to use data that are "informative" in these directions as well.

Obtaining such data requires perturbing inputs in a correlated manner either by using some prior information about the system (e.g., information available from previous identification results) or by closing one or more loops. Parametric identification algorithms can be generalized straightforwardly to the case when several inputs are perturbed simultaneously. The advantage of MIMO identification over MISO identification is that correlations in the output data are also utilized, and this can help when disturbances entering different output channels are correlated. However, parameter identification of a MIMO polynomial model (e.g., MIMO ARMAX model) often results in a numerically ill-conditioned, nonconvex optimization with possible local minima. This is probably the reason why one hardly finds an application of true MIMO identification in the literature.

Recently, a state-space identification method called "subspace identification" has been suggested as an alternative. A well-publicized subspace identification algorithm is the N4SID method developed by Van Overschee and De Moor (1994). The only prior knowledge needed to apply this algorithm is an upper bound on the order of the system. In addition, the algorithm is noniterative and consists of numerically stable operations such as the QR decomposition and SVD. N4SID includes an order-reduction scheme similar to the Hankel-norm-based model reduction as part of the identification algorithm. The order reduction can help the user get rid of extra modes created by identification errors, and therefore prevent overfitting the data. However, it is an open-loop-type order-reduction scheme and should be used with caution for highly interactive systems, as important modes for the low-gain directions can be eliminated in the process. Jacobsen (1994) discussed an application of N4SID to a high-purity distillation column example. He found that the algorithm does lead to an improvement in the gain directionality estimation when compared to the SISO or MISO identification methods, but data informative in all system directions are essential.

Correlated input design

MISO or MIMO identification is beneficial only when one uses a data set informative with respect to all system directions. Hence, obtaining such a data set is another important issue. Koung and MacGregor (1993) proposed an SVD-based method for designing correlated inputs for 2×2 systems such that all the system directions are emphasized equally. The main idea is to make the input signal for the low-gain direction be richer (in the energy content) than that for the high-gain direction, by a factor given by the ratio of the respective singular values. This way both output directions are excited at the same degree (in the two-norm sense). Of course, the method requires prior knowledge of the low- and high-gain input directions as well as the ratio of the singular values. Such knowledge may be derived from the previous identification results. However, the necessity of iteration may prove to be a practical barrier. Another potential problem is that the gain directionality of a dynamic system is a function of frequency, which complicates the design. Koung and MacGregor did not discuss how their method extends to the dynamic case.

Closed-loop identification

An alternative to the correlated input design is to close one or more loops before applying the input signals. The main motivation for this approach is that closing loops prior to the identification facilitates the task of exciting the outputs in a desired manner. For instance, Jacobsen (1994) argued that closing a proportional control loop on a 2×2 system makes the system less interactive, and hence facilitates the task of exciting the outputs in desired directions and at desired magnitudes. Andersen and Kummel (1992b) reported similar results in identifying MISO models for a high-purity distillation

column. They recommended using a low-pass filter to improve the estimation in the low- and mid-frequency regions, as high-frequency input perturbations are enhanced under feedback. Li and Lee (1996) gave a slightly different interpretation, that the gain for an input–output pair with the rest of the outputs under tight control (as in the definition of RGA) is the reciprocal of an element of the gain matrix inverse, and therefore contains information mostly in the low-gain directions. Koung and MacGregor (1993) proposed closing both loops on a 2×2 plant using PI controllers. This clearly simplifies the task of designing input signals (set points) that perturb the outputs in a desired manner. However, a drawback is that it may be difficult to design two PI controllers with sufficiently high bandwidths while still maintaining stability. Varga and Jorgensen (1994) suggest on-line adaptation of model parameters as an alternative.

Weighting of prediction error for bias distribution

In addition to signal-to-noise ratio, the issue of bias distribution needs to be considered carefully in order to obtain a model of accurate gain directionality. This is important when one attempts to obtain a low-order model for which model bias can be significant. Although the frequency domain bias distribution for SISO systems has been discussed extensively in the literature (Ljung, 1987), directional distribution of the bias for MIMO systems is seldom addressed. However, it is obvious that, under the minimization of prediction error, the dominance of perturbations along the high-gain direction in the output data will cause the estimation of the low-gain direction to be sacrificed. In reality, it would seldom be practical to obtain data that have equal amounts of excitation along all system directions (since input constraints and nonlinearity usually prohibit implementing input signals that offset the directional gain difference). If the prediction error is minimized in fitting a low-order parametric model to such data, use of weighting would be essential to achieve desired bias distribution. Gaikwad and Rivera (1994) discuss the use of control-relevant weighting in fitting a low-order parametric model to system frequency responses (that are obtained by fitting high-order ARX models to open-loop data). Although not specifically mentioned in the article, when the control requirement calls for control of the entire output space, the effect of these weightings is the distribution of model bias to all the important input–output directions of the system on an equal basis.

Summary

The importance of the problem is evidenced by the large volume of publications. Most published work centers around the idea of exciting the low-gain directions sufficiently in order to achieve good signal-to-noise ratio for all the output directions. This necessitates perturbing all the inputs simultaneously in general, and hence requires the use of MISO or MIMO identification algorithms. While the prediction error method is the usual choice of model fitting, the issue of bias distribution is seldom discussed. An exception to the article by Gaikwad and Rivera (1994) that discusses generation of the frequency response and subsequent fit of a low-order model using a control-relevant weighting. However, one has to first obtain frequency-response data with correct gain directionality. The method proposed subsequently provides a way to produce such frequency-response data.

Frequency-Domain Identification Method

In this section, we present a frequency-domain identification method for MIMO systems based on the preceding analysis. The proposed method is summarized below:

Step 1. Estimate the frequency response and the matrix inverse by performing open-loop and closed-loop SISO identification experiments.

Step 2. Find a nonparametric or parametric transfer matrix that best fits both sets of the frequency-response data through a least-squares optimization.

Some preliminaries

Throughout this section, the system considered is an $n\times n$ open-loop stable, nonsingular transfer-function matrix $G(s)$. Let us use $g_{ij}(s)$ to denote the (i, j)th element of $G(s)$. The "ideal closed-loop transfer function" $k_{ij}(s)$ will be defined as the transfer function from the jth input to the ith output when the rest of the outputs are under perfect control, that is,

$$k_{ij}(s)=\left(\frac{y_i(s)}{u_j(s)}\right)_{y_l=0,l\neq i}, \tag{8}$$

where $k_{ij}(0)$ is the closed-loop gain used to define the RGA (Bristol, 1966). The concept of perfect control used in the preceding definition may not be physically meaningful at all frequencies, but it proves to be useful in defining controller-independent measures of a plant like the RGA. Extending the RGA relation to other frequencies, it is easy to see that the following relationship holds:

$$\Lambda(G(j\omega))=\begin{bmatrix}\dfrac{g_{11}(j\omega)}{k_{11}(j\omega)} & \cdots & \dfrac{g_{1n}(j\omega)}{k_{1n}(j\omega)}\\ \vdots & \ddots & \vdots\\ \dfrac{g_{n1}(j\omega)}{k_{n1}(j\omega)} & \cdots & \dfrac{g_{nn}(j\omega)}{k_{nn}(j\omega)}\end{bmatrix} \tag{9}$$

$$=G(j\omega)\otimes(G^{-1}(j\omega))^T,$$

where $\otimes$ denotes element-by-element multiplication. Hence

$$G^{-1}(j\omega)=\begin{bmatrix}1 & \cdots & \dfrac{1}{k_{n1}(j\omega)}\\ \vdots & \ddots & \vdots\\ \dfrac{1}{k_{1n}(j\omega)} & \cdots & \dfrac{1}{k_{nn}(j\omega)}\end{bmatrix} \tag{10}$$

In other words, the frequency response of the ideal closed-loop transfer function $k_{ij}(s)$ is the reciprocal of the (j, i)th element of the system's frequency-response matrix inverse. This observation suggests a possibility of identifying the ele-

ments of the frequency-response matrix inverse directly through closed-loop SISO identification experiments.

Obtaining the frequency-response data

To simplify the exposition, we will present the idea for a 2×2 case and then discuss how the method extends to higher dimensional cases in a later section.

Obtaining Data for the Frequency-Response Matrix. The open-loop frequency response $g_{ij}(j\omega)$ can be obtained in a conventional way. Since we will perform a separate closed-loop experiment to obtain data that are well excited along the low-gain direction, inputs may be perturbed one at a time (or simultaneously if preferred). The pseudorandom binary signal (PRBS) is the popular choice of input since the PRBS is easy to implement and can be designed to excite all the frequencies of interest. Once the input–output data are obtained, the frequency response can be calculated either through parametric means (e.g., fitting a high-order ARX model to the data using the prediction error method (PEM)) or through nonparametric means (e.g., using the empirical transfer function estimate (ETFE) method described in Ljung (1987)). The latter method is easy to apply and requires little prior knowledge about the system, but suffers from the drawback that the variance of estimate does not decay with the number of data points. Smoothing windows or the finite impulse-response assumption can be used to bypass this problem (see Arifin et al., 1995), but the approach still requires a large data set in general. On the other hand, PEM requires a parametric structure to be specified *a priori*, but parameter fitting can be usually done with a reasonable-size data set.

Obtaining Data for the Frequency-Response Matrix Inverse. The frequency response obtained in the preceding manner may not accurately model the low-gain directions due to the problems described earlier. Hence, we need to gather additional data that are well excited along the system's low-gain directions. This can be done by obtaining data for the elements of the frequency-response matrix inverse through additional closed-loop experiments. The closed-loop experiments that we propose to perform are shown in Figure 3. Note that closed-loop transfer functions are given by

$$f_{11}(s)=g_{11}(s)-h_{22}(s)\frac{g_{12}(s)g_{21}(s)}{g_{22}(s)}$$

$$f_{12}(s)=g_{12}(s)-h_{21}(s)\frac{g_{11}(s)g_{22}(s)}{g_{21}(s)}$$

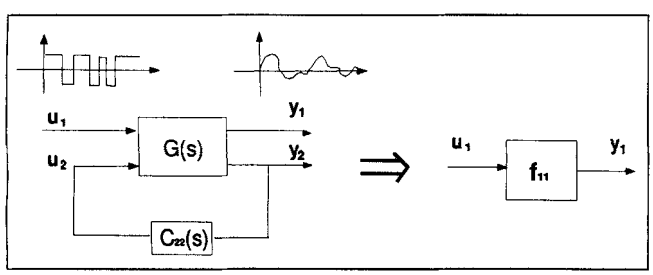

Figure 3. Closed-loop experiment for computing closed-loop transfer function between u_1 and y_1.

$$f_{21}(s)=g_{21}(s)-h_{12}(s)\frac{g_{11}(s)g_{22}(s)}{g_{12}(s)}$$

$$f_{22}(s)=g_{22}(s)-h_{11}(s)\frac{g_{12}(s)g_{21}(s)}{g_{11}(s)}, \qquad (11)$$

where $h_{lk}(s)$ is the complementary sensitivity function for the loop closed between the lth output and the kth input, that is,

$$h_{lk}(s)=\frac{c_{lk}(s)g_{lk}(s)}{1+c_{lk}(s)g_{lk}(s)}. \qquad (12)$$

From Eq. 11 and the definition of RGA, it is straightforward to show that

$$f_{ij}(j\omega)\approx k_{ij}(j\omega)=\frac{1}{[G^{-1}(j\omega)]_{ji}}$$
$$\text{if} \quad h_{lk}(j\omega)\approx 1 \quad \text{for} \quad l\neq i, \quad k\neq j. \qquad (13)$$

In other words, within the bandwidth of the loop, $f_{ij}(j\omega)$ matches $k_{ij}(j\omega)$ very closely and therefore the reciprocal of the (j, i)th element of the frequency-response matrix inverse. Hence, one can obtain a good estimate of $k_{ij}(j\omega)$ for a wide frequency range (say $0\le\omega\le\omega^*$) by performing a SISO frequency-response identification between the jth input and the ith output after closing a tight loop between the other input and output variables (such that $\hat{y}_l(\omega)\approx 0$ for $l\neq i$, $0\le\omega\le\omega^*$). Note that, far beyond the bandwidth of the closed-loop loop, $f_{ij}(j\omega)\approx g_{ij}(j\omega)$ and closed-loop data provide essentially the same information as the open-loop data. However, a SISO loop can usually be tuned quite tightly so that the preceding relation holds within the frequency range relevant for MIMO control.

Reducing the Number of Closed-Loop Experiments Using RGA Relations. The RGA has some useful properties that we can exploit here. One of these is that the sum of the elements of each row and column of the RGA is unity, that is,

$$\sum_{i=1}^{n}\lambda_{ij}(\omega)=\sum_{j=1}^{n}\lambda_{ij}(\omega)=1, \qquad (14)$$

where $\lambda_{ij}(\omega)$ denotes the (i, j)th element of $\Lambda(G(j\omega))$.

The significance of this property is that, for a 2×2 system, only one closed-loop experiment needs to be performed to obtain the data for all the closed-loop transfer functions. For example, suppose that an estimate for $k_{11}(j\omega)$ is obtained for a certain frequency range after closing a tight loop between y_2 and u_2. Then, the rest of the closed-loop frequency response $k_{12}(j\omega)$, $k_{21}(j\omega)$, and $k_{22}(j\omega)$ for the same frequency range can be calculated from the following equations:

$$k_{12}(j\omega)=\frac{g_{12}(j\omega)}{1-\lambda_{11}(\omega)}$$

$$k_{21}(j\omega)=\frac{g_{21}(j\omega)}{1-\lambda_{11}(\omega)}$$

$$k_{22}(j\omega)=\frac{g_{22}(j\omega)}{\lambda_{11}(\omega)}. \tag{15}$$

These relations can be easily derived from Eq. 14 and the definition of the RGA, that is, $\lambda_{ij}=g_{ij}/k_{ij}$. Note that if the estimates for g_{11}, g_{12}, g_{21}, and g_{22} and k_{11} have small relative errors, k_{12}, k_{21}, k_{22} obtained using the relations just listed would also have small relative errors. The only exception is when λ_{11} is close to 1, but this corresponds precisely to the case when there is insignificant interaction among the variables and the open-loop SISO identification is sufficient. In choosing the input–output pair for loop closure, it is important to pick a pair that permits the use of the highest bandwidth (i.e., a pair with negligible delay, smallest time constant, etc.). This way, one can obtain data for the closed-loop frequency response for the widest possible frequency range.

Fitting a model to the open-loop and closed-loop data

In the previous section, we presented an experimental procedure for obtaining data for the elements of the frequency-response matrix and its inverse. The remaining task is to fit a model to these data. We present the following two options: (1) fitting a nonparametric transfer matrix, and (2) fitting a low-order parametric transfer matrix. We will again use a 2×2 system to illustrate the point and defer the discussion of the higher dimensional case to later.

Fitting a Nonparametric Transfer-Function Matrix to the Data. One way to extract a model from the data is to fit a nonparametric transfer matrix (i.e., frequency-response model). This can be formulated as a frequency-by-frequency optimization problem. Let us denote the data sets for the plant frequency-response matrix and its inverse (obtained from the open-loop and closed-loop tests, respectively) as

$$\hat{G}(j\omega)\triangleq\begin{bmatrix}\widehat{g_{11}}(j\omega) & \widehat{g_{12}}(j\omega)\\ \widehat{g_{21}}(j\omega) & \widehat{g_{22}}(j\omega)\end{bmatrix} \tag{16}$$

$$\widehat{G^{-1}}(j\omega)\triangleq\begin{bmatrix}\dfrac{1}{\widehat{k_{11}}(j\omega)} & \dfrac{1}{\widehat{k_{21}}(j\omega)}\\ \dfrac{1}{\widehat{k_{12}}(j\omega)} & \dfrac{1}{\widehat{k_{22}}(j\omega)}\end{bmatrix} \tag{17}$$

The notation $\{\hat{}\}$ is used to emphasize the fact that they are identified values and may be different from the actual plant frequency responses. Our objective is to find a nonparametric transfer matrix model $\tilde{G}(j\omega)$ that fits the both sets of the data in the best possible sense. Note that, since the identification data $\hat{G}(j\omega)$ and $\widehat{G^{-1}}(j\omega)$ contain errors, $\widehat{G^{-1}}(j\omega)\neq\hat{G}^{-1}(j\omega)$. As a matter of fact, they can differ considerably for highly interactive systems, due to the high sensitivity of the inverse elements to the perturbations in the open-loop elements. Hence, we must choose the model such that not only $\tilde{G}(j\omega)$ matches $\hat{G}(j\omega)$ closely, but also $\tilde{G}^{-1}(j\omega)$ matches $\widehat{G^{-1}}(j\omega)$ closely.

In the absence of any quantitative information about the identification errors, it is reasonable to use the following least-squares objective for this purpose:

$$\min_{\tilde{g}_{11}(\omega),\ldots,\tilde{g}_{22}(\omega)}\sum_{j=1}^{2}\sum_{i=1}^{2}\left[\frac{1}{\gamma_{ij}(\omega)}\left|\frac{\widehat{g_{ij}}(j\omega)-\tilde{g}_{ij}(j\omega)}{\widehat{g_{ij}}(j\omega)}\right|^2+\frac{1}{\rho_{ij}(\omega)}\left|\frac{\dfrac{1}{\widehat{k_{ij}}(j\omega)}-\dfrac{1}{\tilde{k}_{ij}(j\omega)}}{\dfrac{1}{\widehat{k_{ij}}(j\omega)}}\right|^2\right] \tag{18}$$

where

$$\begin{aligned}\frac{1}{\tilde{k}_{11}(j\omega)}&=\frac{\tilde{g}_{22}(j\omega)}{\tilde{g}_{11}(j\omega)\tilde{g}_{22}(j\omega)-\tilde{g}_{12}(j\omega)\tilde{g}_{21}(j\omega)}\\ \frac{1}{\tilde{k}_{12}(j\omega)}&=-\frac{\tilde{g}_{21}(j\omega)}{\tilde{g}_{11}(j\omega)\tilde{g}_{22}(j\omega)-\tilde{g}_{12}(j\omega)\tilde{g}_{21}(j\omega)}\\ \frac{1}{\tilde{k}_{21}(j\omega)}&=-\frac{\tilde{g}_{12}(j\omega)}{\tilde{g}_{11}(j\omega)\tilde{g}_{22}(j\omega)-\tilde{g}_{12}(j\omega)\tilde{g}_{21}(j\omega)}\\ \frac{1}{\tilde{k}_{22}(j\omega)}&=\frac{\tilde{g}_{11}(j\omega)}{\tilde{g}_{11}(j\omega)\tilde{g}_{22}(j\omega)-\tilde{g}_{12}(j\omega)\tilde{g}_{21}(j\omega)},\end{aligned} \tag{19}$$

where $\gamma_{ij}(\omega)$ and $\rho_{ij}(\omega)$ are *de facto* tuning parameters that weigh the data points according to their importance and reliability (after normalization). In probabilistic terms, the preceding least-squares solution corresponds to the maximum likelihood estimate (MLE) under the following multiplicative errors on the frequency-response data:

$$\begin{aligned}\widehat{g_{ij}}(j\omega)(1+\epsilon_{ij}(\omega))&=g_{ij}(j\omega)\\ \frac{1}{\widehat{k_{ij}}(j\omega)}(1+w_{ij}(\omega))&=\frac{1}{k_{ij}(j\omega)},\end{aligned} \tag{20}$$

where ϵ_{ij} and w_{ij} are normally distributed variables of zero mean and covariance $\gamma_{ij}I_2$ and $\rho_{ij}I_2$, respectively (when the complex variables are viewed as two-dimensional real random vectors). In the absence of any quantitative information about the identification errors in different data points, it is reasonable to assume that each data point has a similar multiplicative error and assign $\gamma_{ij}(\omega)=\rho_{ij}(\omega)=1$ for all i, j, and ω. In this case, the open-loop data and closed-loop data are given the same weight, and therefore the high-gain direction and the low-gain direction are equally emphasized in fitting the model. If an accurate stochastic model for identification errors is available and the model bias is insignificant, one can develop confidence intervals for the identified frequency-response data (for instance, using the method proposed by Arifin et al., 1995). Then, one can set up an optimal estimation problem based on a well-defined criterion like maximum likelihood in order to fit the nonparametric transfer matrix model.

If closed-loop data are not available for the entire fre-

quency range, one can perform the preceding optimization for the frequency range for which the closed-loop data are available and use the open-loop data for the rest. In formulating constraints in Eq. 19, it was assumed that the data for $k_{ij}(j\omega)$ were obtained under the condition that the other output was under perfect control (within the frequency range of the applied closed-loop data). One can relax this by replacing Eq. 19 with Eq. 11. Note that one can express $h_{ij}(j\omega)$ in terms of the controller and open-loop frequency response through Eq. 12. According to our experience, this turns out to be an unnecessary complication in most cases since one can usually tune a single loop such that its bandwidth exceeds the desired bandwidth for MIMO control.

Fitting a Parametric Model to the Data. It is also possible to fit a low-order parametric model (e.g., continuous or discrete transfer matrix) directly to the data. In this case, model bias may be significant, and its distribution depends on how different data points are weighed in fitting the parameters. We will assume here that a continuous transfer function model is to be fitted. Suppose the transfer matrix to be fitted is parameterized as below:

$$\tilde{G}(s,\theta)=\begin{bmatrix}\tilde{g}_{11}(s,\theta) & \tilde{g}_{12}(s,\theta)\\ \tilde{g}_{21}(s,\theta) & \tilde{g}_{22}(s,\theta)\end{bmatrix}. \tag{21}$$

Then, the least-squares objective is

$$\min_{\theta}\int_0^{\omega^*}\sum_{j=1}^{2}\sum_{i=1}^{2}\left[\frac{1}{\gamma_{ij}(\omega)}\left|\frac{\widehat{g_{ij}}(j\omega)-\tilde{g}_{ij}(j\omega)}{\widehat{g_{ij}}(j\omega)}\right|^2+\frac{1}{\rho_{ij}(\omega)}\left|\frac{\dfrac{1}{\widehat{k_{ij}}(j\omega)}-\dfrac{1}{\tilde{k}_{ij}(j\omega)}}{\dfrac{1}{\widehat{k_{ij}}(j\omega)}}\right|^2\right]d\omega \tag{22}$$

with constraints (in addition to Eq. 19)

$$\tilde{g}_{ij}(j\omega)=\tilde{g}_{ij}(\theta,\omega)\qquad \text{for}\quad 1\le i\le 2,\quad 1\le j\le 2, \tag{23}$$

where ω^* denotes the highest frequency for which the frequency-response data are available. The optimization is nonconvex in general, and one needs a good starting point for convergence. One can, for instance, use the parameters obtained from a conventional method (e.g., PEM) as a starting point. The relative magnitudes of $\gamma_{ij}(\omega)$ and $\rho_{ij}(\omega)$ determine the relative importance of the high-gain direction vs. the low-gain direction in fitting the model, and affect the bias distribution. Normally, both directions are equally important, and hence it is reasonable to choose the weightings such that $\gamma_{ij}(\omega)=\lambda_{ij}(\omega)$ for all i and j. These weighting factors can also be varied with ω to emphasize the fit of the frequency response in certain frequency ranges over others. Again, a more rigorous approach is possible if a model for the identification errors exists.

Extension to higher dimensional systems

Extending the preceding identification procedure and algorithm to higher dimensional systems is conceptually straightforward, but presents some practical barriers. The first obstacle is the curse of dimensionality, that is, the number of experiments required grows exponentially with the system dimension (four experiments are needed for 3×3 systems, nine experiments are needed for 4×4 systems, etc.). In addition, more loops need to be closed in order to perform the closed-loop experiments. Hence, it is unrealistic to expect that the procedure be applied exactly as described to cases larger than 3×3. However, from the preceding analysis, one can speculate that any closed-loop input–output data, regardless of the number of loops closed, serve as useful information for identification.

One can actually systematize this idea by exploiting a property of the RGA. For instance, open-loop tests can be carried out first and an analysis based on the RGA of the identified model can be performed to locate subsystems that are highly interactive. A useful fact in this regard is that the reciprocal of an RGA element is the maximum allowed multiplicative error on the corresponding gain element in order for the matrix to remain nonsingular (Hovd and Skogestad, 1992). Hence, additional closed-loop tests can be carried out on those subsystems with large RGA elements, and the frequency responses for these subsystems based on both the open-loop data and closed-loop data may be constructed via frequency-by-frequency optimization.

Discussion

First and foremost, the proposed method can be viewed as a way of generating input–output data that are informative in all system directions. For highly interactive systems, open-loop SISO tests provide information only about the high-gain direction. The proposed closed-loop test provides additional information about the low-gain direction, to yield a data set with a good signal-to-noise ratio for all the gain directions. Model fitting was performed in a particular way, that is, by estimating the frequency-response matrix and its inverse using the open-loop and closed-loop data, respectively, followed by a least-squares fitting in the frequency domain. We chose the frequency-domain framework because this allowed us to either avoid or treat the issue of bias distribution in a transparent manner.

In fact, the proposed closed-loop test can also be used with conventional time-domain MISO or MIMO identification algorithms as well. In this case, the test facilitates generation of data that are informative in all system directions. For 2×2 systems, two closed-loop tests independently excite two orthogonal directions of the output space (i.e., $[1, 0]^T$ and $[0, 1]^T$ directions). One open-loop test and one closed-loop test can also be designed to achieve a similar effect. However, values of all the inputs (not just the input that is manipulated) need to be recorded during the closed-loop tests.

If a low-order parametric model is fitted to the data through the PEM, one needs to pay particular attention to the issue of bias distribution. Bias distribution under the PEM depends on the level of output excitation, and it may not have been possible to excite the high-gain and the low-gain directions of the system equally. Some weightings would be neces-

Table 1. Data for the High-Purity Distillation Column

Mixture	Methanol–ethanol
Product split	0.01–0.99
Number of trays	27
Top temperature	608.0 R
Bottom temperature	642.24 R
Feed composition	0.5

sary in order that the prediction errors for all the system directions are weighed equally. Subspace algorithms like N4SID (Van Overschee and de Moor, 1994) may bypass this problem (since it starts out assuming a very high-order model), but care must be taken in deciding the model order in the order-reduction step so as not to eliminate the modes corresponding to the low-gain directions. On the other hand, choosing too high an order may leave extra modes that are created by the errors in the high-gain directions. A good option is to construct a high-order model using a subspace algorithm and perform an order reduction using the control-relevant weighting, as discussed in Gaikwad and Rivera (1994), or using the frequency-domain parameter-fitting method that was suggested in this article.

The proposed method can also be interpreted as a way of checking the internal consistency of an open-loop MIMO model and modifying the model accordingly. The closed-loop data can be viewed as additional constraints that need to be satisfied in order for an open-loop model to be "internally consistent." The least-squares problem is set up to modify the model so that the consistency relationships would be satisfied to a reasonable degree. This may be a practical view of the proposed methodology for systems of large dimension.

Numerical Example

In this section, the proposed identification method is applied to a nonideal, high-purity distillation column studied by Chien and Ogunnaike (1994).

Column description

The distillation column studied is a high-purity column separating a binary mixture of methanol–ethanol into product streams of 99% purity. It consists of 25 trays plus a reboiler and a total condenser. The two controlled variables are temperatures on trays 21 and 7. The two manipulated inputs are reflux flow and vapor boilup. Symbols T_{21}, T_7, L, and V will be used to represent the normalized output and input deviations from the steady-state value. The normalization factors are 30 and 6 for the temperatures and flows, respectively. The column information is given in Table 1. The dynamic model used in the nonlinear simulation was first presented by Weishedel and McAvoy (1980). The model includes not only the component mass balances but also the pseudo-steady-state energy balances, and uses the Francis Weir formula for flow dynamics.

The column is ill-conditioned, especially in the low-frequency region, as shown in Figure 4. The strong dependence of the system gain on the direction of input perturbation is also a good indication of ill-conditioning, as illustrated by Chien and Ogunnaike (1994). In addition to these, the column is highly nonlinear and interactive.

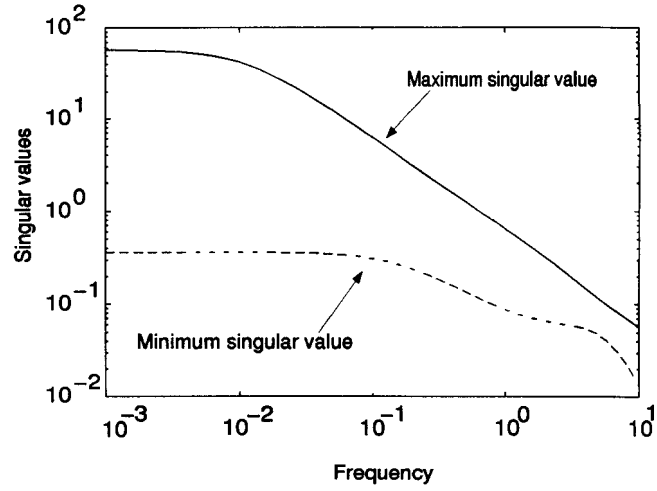

Figure 4. Frequency-dependent plant singular values for the linearized theoretical distillation column model.

Solid—maximum singular value; dash—minimum singular value.

SISO identification

According to Chien and Ogunnaike (1994), the linear models that they obtained from step tests and PRBS tests varied considerably, both in terms of model parameters and RGA. Such dramatic differences in the identified models for the same process made them believe that the highly nonlinear dynamics of the process cannot be modeled accurately with a single linear model. In fact, with linear models giving an unsatisfactory closed-loop performance, they went on to construct nonlinear input–output models to alleviate the difficulties (Ogunnaike and Arkun, 1993). We believe that, although the nonlinearity contributed significantly to the identification errors, the core of the problem is that open-loop SISO identifications were performed on a highly interactive system, leading to very poor estimation of the dynamics along the low-gain direction. Hence, it may have been possible to obtain satisfactory closed-loop performance even with a linear model if special measures were taken in identifying the linear model. We will explore this possibility here.

Chien and Ogunnaike (1994) obtained the following two models by applying the PRBS inputs of different switching time:

$$\tilde{G}_1(s) = \begin{pmatrix} \dfrac{-33.89}{(98.02s+1)(0.42s+1)} & \dfrac{32.63}{(99.60s+1)(0.35s+1)} \\ \dfrac{-18.85}{(75.43s+1)(0.30s+1)} & \dfrac{34.84}{(110.50s+1)(0.03s+1)} \end{pmatrix} \tag{24}$$

$$\tilde{G}_2(s) = \begin{pmatrix} \dfrac{-29.97}{(120.0s+1)(0.20s+1)} & \dfrac{35.64}{(82.86s+1)(0.43s+1)} \\ \dfrac{-35.97}{(105.9s+1)(0.55s+1)} & \dfrac{16.17}{(65.44s+1)(0.03s+1)} \end{pmatrix} \tag{25}$$

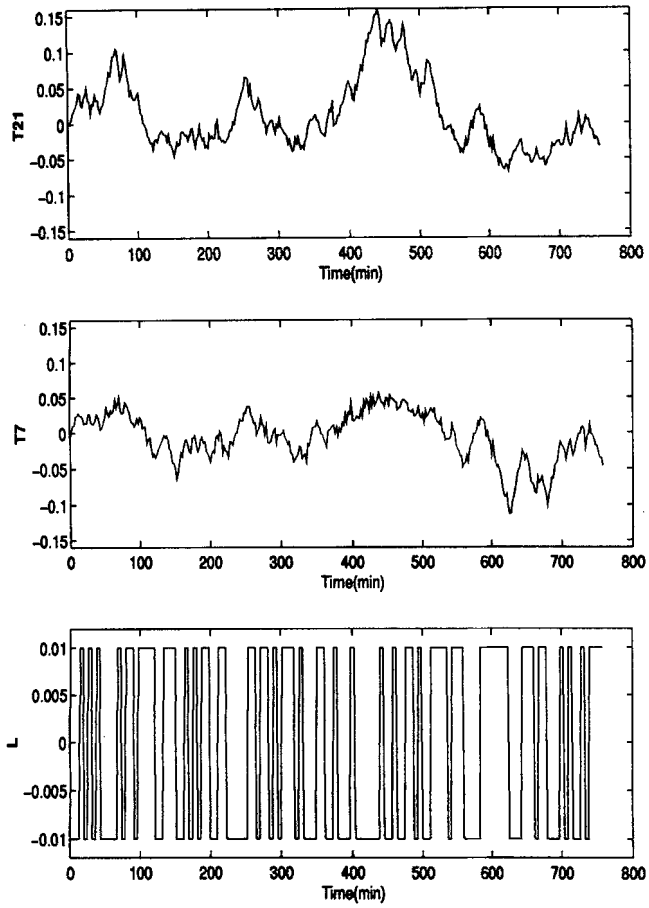

Figure 5. Sample data obtained from an open-loop experiment.

L was perturbed and V was kept constant.

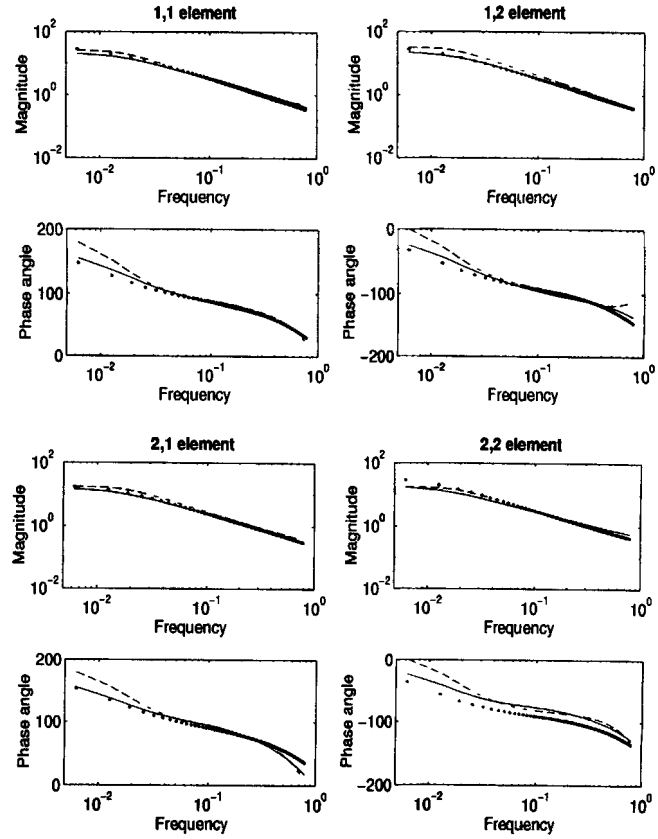

Figure 6. Frequency-response data: open-loop experiment with frequency-response plots of the linearized theoretical model vs. Chien and Ogunnaike ($\tilde{G}_1$) model.

Solid—theoretical model; dash—data obtained from the open-loop experiment; dot—$\tilde{G}_1$.

Note that these two models are similar in terms of individual transfer-function elements; however, their RGAs differ dramatically (for instance, $\lambda_{11}(\tilde{G}_1(0)) = 2.09$ and $\lambda_{11}(\tilde{G}_2(0)) = -0.008$). Indeed, the theoretically linearized model gives λ_{11} of 38.4. Hence, the (1,1) element of the RGA for $\tilde{G}_2$ has a wrong sign, and this would lead to instability when the controller is designed to have an integral action. The discrepancy among the RGA elements is caused by the poor fit of the weak direction, which can be improved by using our proposed method. Next we will apply the proposed method to identify a control-relevant MIMO model for the process.

Identification of a nonparametric transfer matrix

Obtaining Frequency Responses from Open-Loop Experiments. In obtaining the experimental data, we used as input perturbations PRBSs consisting of seven shift registers with a switching time of 6 min. We first perturbed the input L (reflux flow) with the PRBS of magnitude 0.01 and recorded the outputs T_{21} and T_7. The identification experiment lasted for 762 min. With the 2-min sampling time ($T_s = 2$), 381 samples of each output were collected. We then repeated the same process with the other input V (vapor boilup). To make the problem realistic, we added measurement noise (white noise) of standard deviation ± 0.005 to each of the output data. From the two open-loop experiments, four sets of SISO data were generated. Data for T_{21} and T_7 obtained by perturbing L are plotted in Figure 5. The data are then used to fit fourth-order ARX models using the PEM. We then converted the models into frequency responses at selected frequency points between 0 and π/T_s. The open-loop frequency responses obtained from the procedure are plotted in Figure 6 and compared with those of $\tilde{G}_1$ and the theoretically linearized model. The fit is quite reasonable for both models.

Obtaining Frequency Response from Closed-Loop Experiments. In obtaining the closed-loop frequency response, we first generated the data by performing the closed-loop experiment described in the section on obtaining the frequency-response data. Input–output data for L and T_{21} were obtained by perturbing L with the same PRBS signal, but of magnitude 0.05 this time. As before, white measurement noise of standard deviation ± 0.005 was added to the output data for T_{21}. A PRBS perturbation of a larger magnitude was needed because the closed-loop testing mainly excites the weak direction of the system of which the gain is considerably smaller. The sampling time and the duration of the experiment were chosen in the same way as in the open-loop experiment. During the testing, a PI controller was used to manipulate V so that T_7 remained at the steady-state value. The

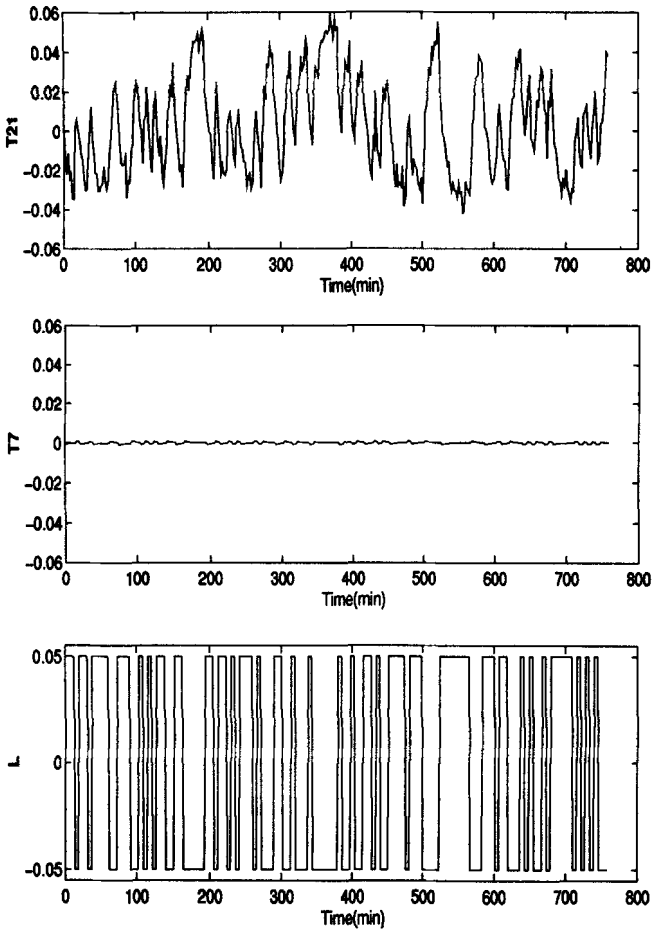

Figure 7. Sample data obtained from a closed-loop experiment.

L was perturbed and V was manipulated such that T_7 remains constant.

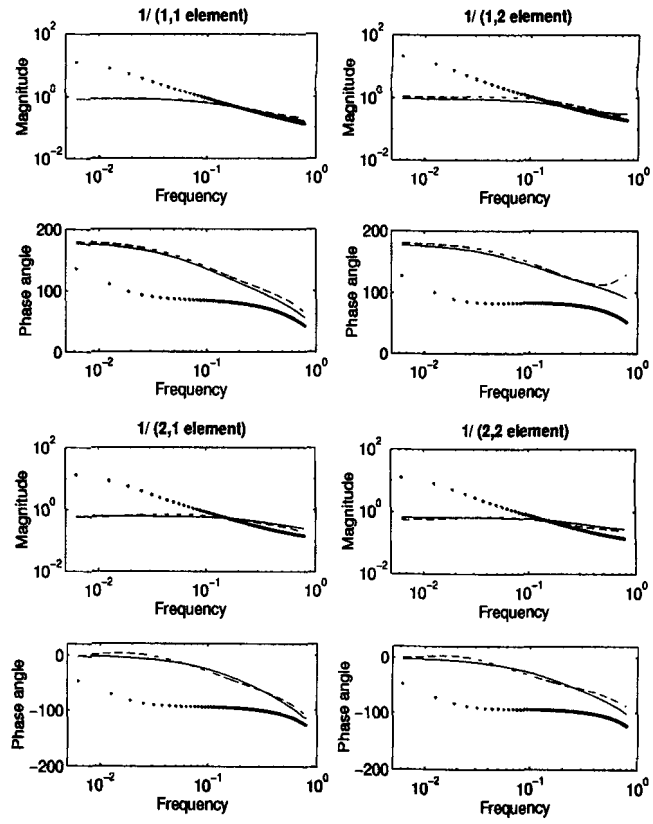

Figure 8. Frequency-response data: closed-loop experiment with the equivalent plots of the linearized theoretical model vs. $\tilde{G}_1$ model.

Solid—reciprocal of the elements of the theoretical model inverse; dash—data obtained from the closed-loop experiment; dot—reciprocal of the elements of $\tilde{G}_1^{-1}$.

loop was tuned tightly so that $\hat{T}_7(j\omega) \approx 0$ for the entire frequency range of interest. This fact can be seen in Figure 7 that plots the input–output data obtained from the closed-loop experiment. Note that T_7 remains very close to the set point at all times. As we have mentioned, for a 2×2 system, one closed-loop experiment is sufficient. Closed-loop data are used to fit a fifth-order ARX model through the PEM. The ARX model was then used to generate $\widehat{k_{11}}(j\omega)$, the data for the reciprocal of the (1,1) element of the frequency response matrix inverse. The rest of the elements are then computed through the RGA relations given in Eq. 15. Figure 8 compares the frequency plots of $\widehat{k_{11}}(j\omega)$, $\widehat{k_{12}}(j\omega)$, $\widehat{k_{21}}(j\omega)$, and $\widehat{k_{22}}(j\omega)$ with the reciprocals of the corresponding elements of $\tilde{G}_1^{-1}$ (the model given in Chien and Ogunnaike (1994)) and the theoretical model. Note the excellent match between the theoretical model and the data. On the other hand, $\tilde{G}_1^{-1}(j\omega)$ shows a large deviation from the other two, especially in the low-frequency region.

Fitting a Nonparametric Transfer Matrix to both Sets of Data. Next we fitted a frequency-response matrix (denoted as $\tilde{G}_3(j\omega)$ from here on). Since exact knowledge of the quality of the individual frequency data points was not available, we assumed that all the four data points for a particular frequency have the same multiplicative errors, and hence chose $\gamma_{ij}(\omega) = \rho_{ij}(\omega) = 1$ for all i, j. The actual computation was carried out using the Matlab Optimization Toolbox. Figure 9 and Figure 10 show the frequency responses of $\tilde{G}_3(j\omega)$ and $\tilde{G}_3^{-1}(j\omega)$ against the frequency-response data to which the model was fitted. The fit is very reasonable for both data. Because the frequency-response data matched the theoretical model very closely, we can conclude that $\tilde{G}_3$ also matches the theoretical model very closely (i.e., in terms of frequency response and matrix inverse). We also plotted the RGA and singular values for the linearized theoretical model, model $\tilde{G}_3(j\omega)$ and model $\tilde{G}_1(j\omega)$ in Figure 11. Both the RGA and singular-value plots show an excellent fit between our model $\tilde{G}_3(j\omega)$ and the theoretical model. On the other hand, Chien and Ogunnaike's model $\tilde{G}_1(j\omega)$ has a large mismatch in the low- and intermediate-frequency region that is caused by poor estimation of the low-gain direction.

Closed-Loop Simulation. To compare the quality of the model obtained from the proposed method with $\tilde{G}_1$ obtained from open-loop identification, closed-loop simulations were performed using the dynamic matrix controllers (formulated as in Lee et al., 1994). The tuning parameters used for both DMC controllers (see Lee et al. (1994) for notation) are

$$p = 100 \qquad m = 25 \qquad \tilde{\Lambda}^y = I_2 \qquad \tilde{\Lambda}^u = 0.3I_2.$$

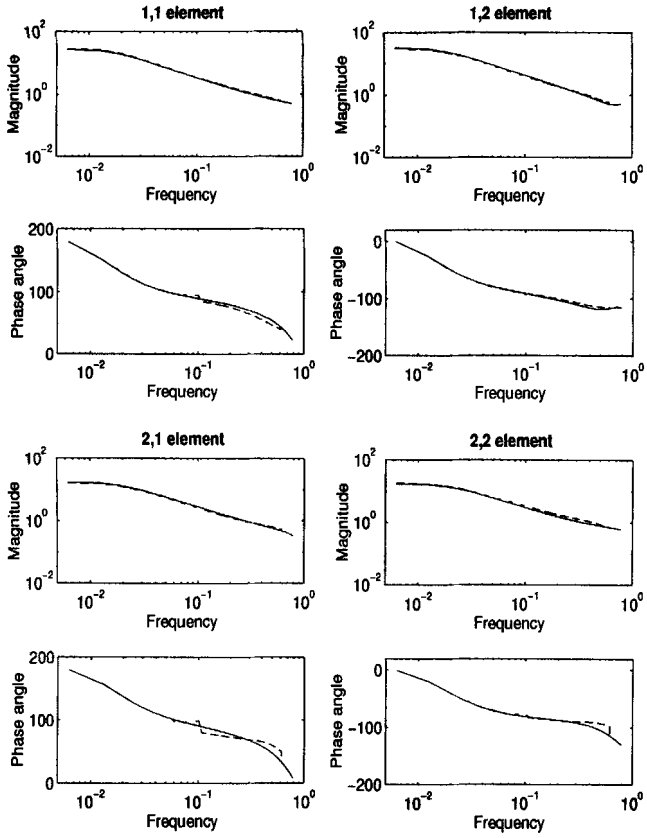

Figure 9. Fit between the model $\tilde{G}_3$ obtained through the least-squares optimization and the open-loop frequency-response data used in the optimization.

Solid—data from the open-loop experiment; dash—$\tilde{G}_3$.

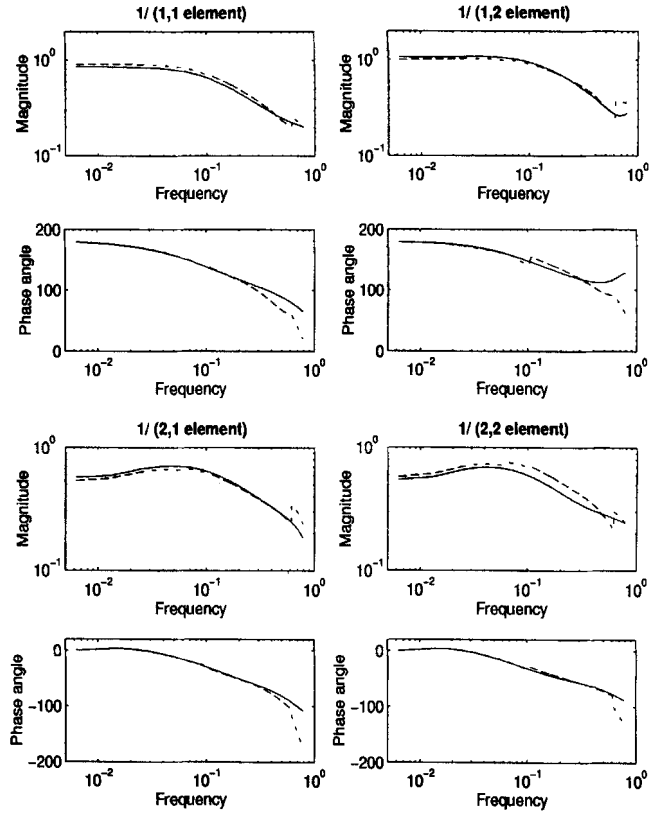

Figure 10. Fit between the model $\tilde{G}_3$ obtained through the least-squares optimization and the closed-loop frequency-response data used in the optimization.

Solid—data from the closed-loop experiment; dash—reciprocal of the elements of $\tilde{G}_3^{-1}$.

Impulse-response models tested were obtained from $\tilde{G}_3$ (through the proposed method) and $\tilde{G}_1$ (provided by Chien and Ogunnaike (1994)). Sixty impulse-response coefficients were used for both models. In all simulations, the nonlinear model by Weishedel and McAvoy (1980) was used to generate the input–output data for the plant.

The closed-loop responses of T_{21} and T_7 for a step change of 0.05 in the set point for T_{21} are plotted in Figure 12. Similarly, the closed-loop responses for a step change of 0.05 in the set point for T_7 are given in Figure 13. Recall that model $\tilde{G}_1$ had a problem matching the plant frequency-response matrix inverse in the low- and intermediate-frequency region. This error manifests as the very sluggish settling of the output variables in the closed-loop simulation. On the other hand, the DMC controller based on $\tilde{G}_3$ results in much quicker settling and lower overshoots.

Fitting a low-order parametric transfer matrix

In the previous section we fitted a nonparametric model to the open-loop and closed-loop frequency-response data through a frequency-by-frequency least-squares optimization. As mentioned earlier (in the section titled "Fitting a model to the open-loop and closed-loop data"), it is also possible to fit a low-order parametric transfer matrix. In this section we present the results we obtained from doing this. We chose the same model structure as those used in Chien and Ogunnaike (1994) in order to make the direct comparison possible. The structure is given below:

$$\tilde{G}_4(s,\theta)=\begin{bmatrix}\tilde{g}_{11}(s,\theta) & \tilde{g}_{12}(s,\theta)\\ \tilde{g}_{21}(s,\theta) & \tilde{g}_{22}(s,\theta)\end{bmatrix}$$

$$=\begin{bmatrix}\dfrac{b_{11}}{(\tau_{11}^1 s+1)(\tau_{11}^2 s+1)} & \dfrac{b_{12}}{(\tau_{12}^1 s+1)(\tau_{12}^2 s+1)}\\ \dfrac{b_{21}}{(\tau_{21}^1 s+1)(\tau_{21}^2 s+1)} & \dfrac{b_{22}}{(\tau_{22}^1 s+1)(\tau_{22}^2 s+1)}\end{bmatrix}. \quad (26)$$

The parameter vector to be determined is

$$\theta=[b_{11},\tau_{11}^1,\tau_{11}^2,b_{12},\tau_{12}^1,\tau_{12}^2,b_{21},\tau_{21}^1,\tau_{21}^2,b_{22},\tau_{22}^1,\tau_{22}^2]^T. \quad (27)$$

We fitted this parameter vector based on the objective function, Eq. 22 and Eq. 23. The integral was approximated with the sum of the integrand evaluated at evenly spaced frequency points. We note that, because we are fitting a low-order parametric model, the model bias can be quite signifi-

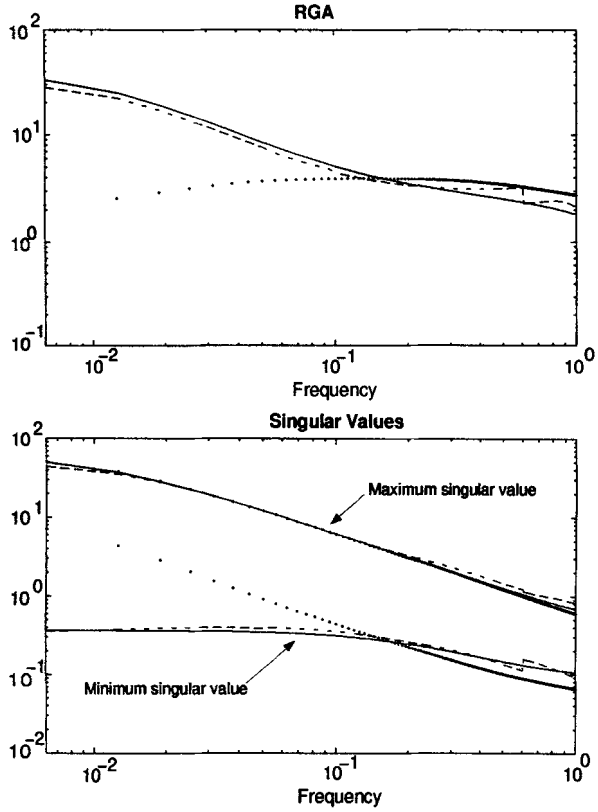

Figure 11. Comparison of RGA and singular values of $\tilde{G}_3$ with those of the theoretical model and $\tilde{G}_1$.

Solid—theoretical model; dash—$\tilde{G}_3$; dot—$\tilde{G}_1$.

cant. Considering the fact that the model can only fit the data for a limited frequency region accurately, one should emphasize the frequency range that is important for control in evaluating the integral. Assuming we want the desired

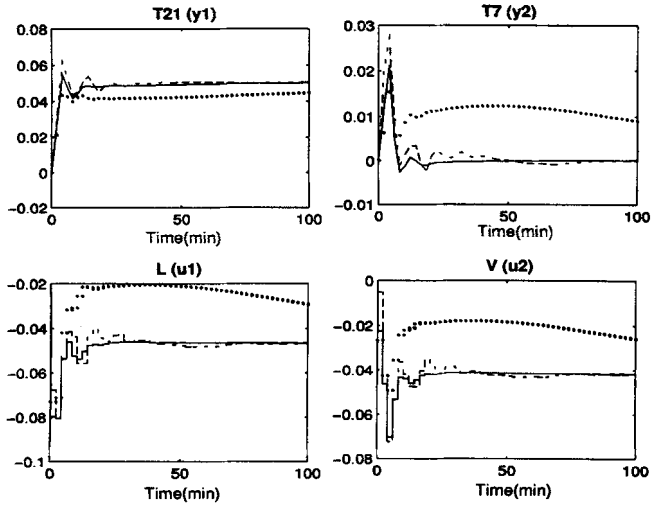

Figure 12. Closed-loop simulation of the nonlinear plant for a setpoint change of $[0.05, 0]^T$ with a DMC controller based on the theoretical model, open-loop/closed-loop data-fitted nonparametric model $\tilde{G}_3$ and Chien and Ogunnaike's model $\tilde{G}_1$.

Solid—G; dash—$\tilde{G}_3$; dot—$\tilde{G}_1$.

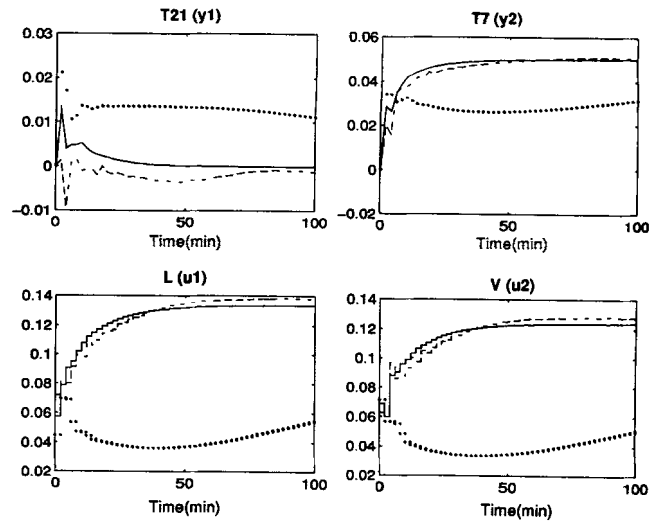

Figure 13. Closed-loop simulation of the nonlinear plant for a setpoint change of $[0, 0.05]^T$ with a DMC controller based on the theoretical model, open-loop/closed-loop data-fitted nonparametric model $\tilde{G}_3$ and Chien and Ogunnaike's model $\tilde{G}_1$.

Solid—theoretical model; dash—$\tilde{G}_3$; dot—$\tilde{G}_1$.

closed-loop time constant to be around 5 min, we need a model accurate up to a frequency of about 0.5 rad/min. Hence, we selected 30 frequency points within this frequency range and used the results for these frequencies to evaluate the objective function. The optimization in Eqs. 22 and 23 is nonconvex and requires a good starting point for convergence to the true optimum. We used the parameters in Ogunnaike and Arkun's model $\tilde{G}_1(s)$ as the initial guess values. The weighting factors $\gamma_{ij}(\omega)$ and $\rho_{ij}(\omega)$ were all chosen as 1, since both output directions were considered to be

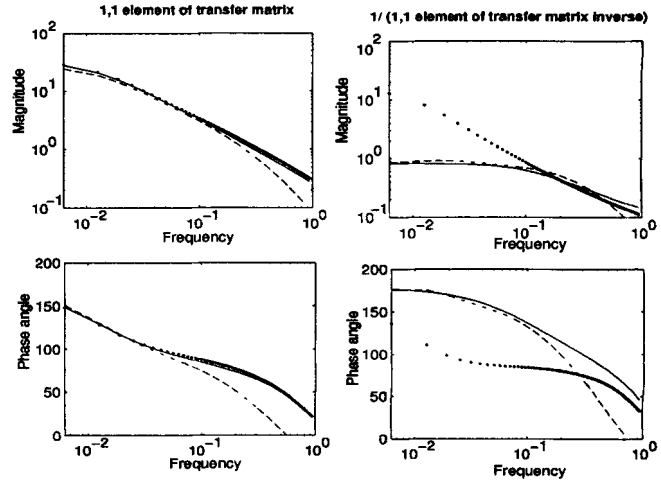

Figure 14. 1,1 element of the frequency-response matrix and the reciprocal of the 1,1 element of the frequency-response matrix inverse for the theoretical model, open-loop/closed-loop data-fitted model $\tilde{G}_4$ and Chien and Ogunnaike's model $\tilde{G}_1$.

Solid—theoretical model; dash—$\tilde{G}_4$; dot—$\tilde{G}_1$.

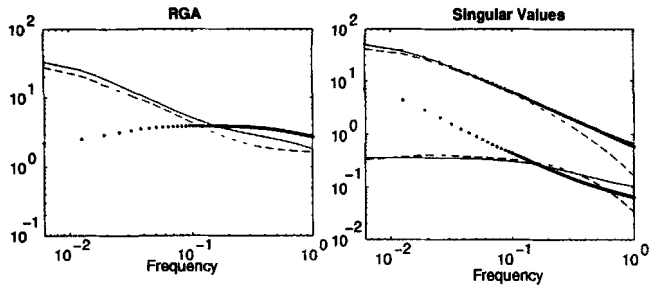

Figure 15. Comparison of the RGA and the singular values of $\tilde{G}_4$ with those of the theoretical model and Chien and Ogunnaike's model $\tilde{G}_1$.

Solid—theoretical model; dash—$\tilde{G}_4$; dot—$\tilde{G}_1$.

equally important for control. The model we obtained from the optimization is

$$\tilde{G}_4(s,\theta) = \begin{bmatrix} \dfrac{-27.24}{(84.99s+1)(2.99s+1)} & \dfrac{28.76}{(80.62s+1)(3.53s+1)} \\ \dfrac{-15.82}{(65.14s+1)(3.42s+1)} & \dfrac{17.29}{(60.85s+1)(1.66s+1)} \end{bmatrix} \quad (28)$$

Figure 14 shows the comparison of the 1,1 element of $\tilde{G}_4(j\omega)$ with that of $\tilde{G}_1(j\omega)$ and the linearized theoretical model. Also compared are the closed-loop frequency responses (i.e., the reciprocal of the 1,1 element of the transfer matrix inverse) for the three models. The good directionality fit between our model and the theoretical model is also seen in the close match of the RGA($j\omega$) and the singular values, shown in Figure 15. Note that there is a slight mismatch between $\tilde{G}_4(j\omega)$ and the theoretical model both in the RGA and the maximum singular value. This may be due to the model bias or the presence of a local minima. When we compare $\tilde{G}_4(j\omega)$ with the open-loop model $\tilde{G}_1(j\omega)$, however, we see a clear improvement of the model quality in terms of directionality fit. This is also verified through closed-loop simulations that are presented in Figures 16 and 17.

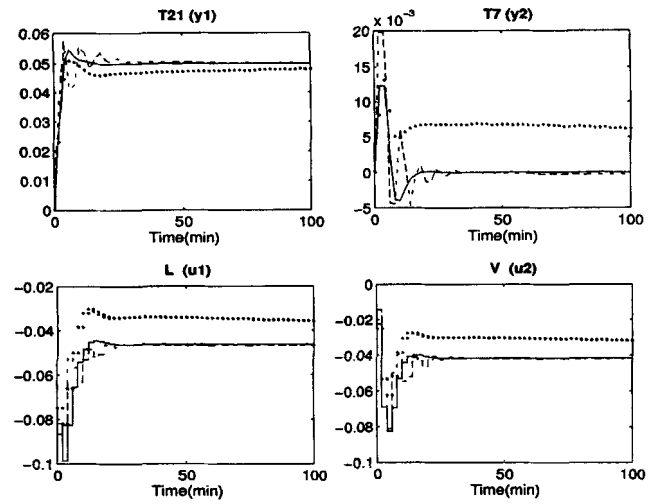

Figure 16. Closed-loop simulation of the nonlinear plant for a setpoint change of $[0.05, 0]^T$ with a DMC controller based on the theoretical model, open-loop/closed-loop data-fitted parametric model $\tilde{G}_4$ and Chien and Ogunnaike's model $\tilde{G}_1$.

Solid—theoretical model; dash—$\tilde{G}_4$; dot—$\tilde{G}_1$.

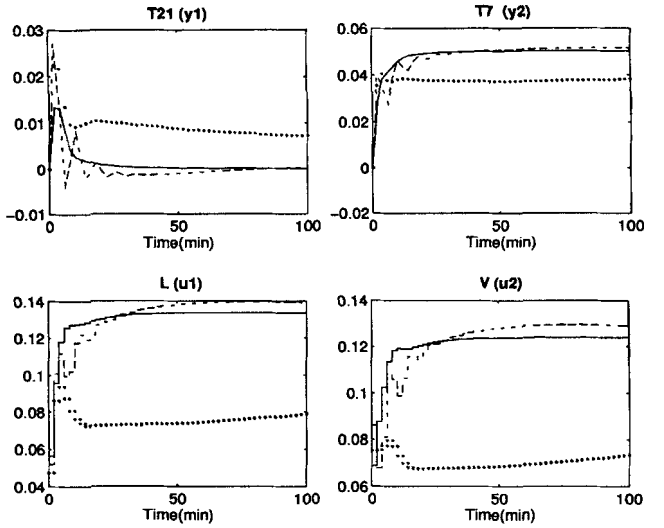

Figure 17. Closed-loop simulation of the nonlinear plant for a setpoint change of $[0, 0.05]^T$ with a DMC controller based on the theoretical model, open-loop/closed-loop data-fitted parametric model $\tilde{G}_4$ and Chien and Ogunnaike's model $\tilde{G}_1$.

Solid—theoretical model; dash—$\tilde{G}_4$; dot—$\tilde{G}_1$.

Conclusions

This article addressed the issue of identifying a multivariable system model for the purpose of feedback control. A requirement for such a model is that it correctly captures the system's gain directionality over the frequency range of interest. It was shown that this requirement is satisfied if the model accurately matches the system's frequency responses as well as the elements of the inverse of the frequency-response matrix. We proposed a combined open-loop and closed-loop testing and model-fitting procedure for obtaining such a model.

The RGA once again proved to be useful in both analyses and reducing the number of experiments needed. In addition, we mentioned that, in large-scale system identification, RGA can help identify subsystems that may require such an elaborate identification. The article also demonstrated that an RGA computed purely based on the identified open-loop gains can have significant errors (e.g., sign errors) leading to erroneous conclusions.

Acknowledgments

The financial support from the National Science Foundation (CTS No. 9209808) is gratefully acknowledged. We would also like to acknowledge the contribution by Dr. Babatunde Ogunnaike, who shared many discussions with us and provided us with the simulation code for the distillation column studied in this article.

Literature Cited

Andersen, H. W., M. Kummel, and S. B. Jorgensen, "Dynamics and Identification of a Binary Distillation Column," *Chem. Eng. Sci.*, **44**(11), 2571 (1989).

Andersen, H. W., and M. Kummel, "Evaluating Estimation of Gain Directionality, Part 1: Methodology," *J. Process Contr.*, **2**(2), 59 (1992a).

Andersen, H. W., and M. Kummel, "Evaluating Estimation of Gain Directionality, Part 2: A Case Study of Binary Distillation," *J. Process Contr.*, **2**(2), 67 (1992b).

Arifin, N., L. Wang, E. Goderdhansingh, and W. R. Cluett, "Identification of the Shell Distillation Column Using the Frequency Sampling Filter Model," *J. Process Contr.*, **5**, 71 (1995).

Bristol, E. H., "On a New Measure of Interaction for Multivariable Process Control," *IEEE Trans. Automat. Control*, **AC-11**, 133 (1966).

Chien, I. L., and B. A. Ogunnaike, "Modeling and Control of High-Purity Distillation Columns," AIChE Meeting, Miami Beach, FL (1994).

Gaikwad, S. V., and D. E. Rivera, "Integared Identification and Control for Model Predictive Controllers," AIChE Meeting, San Francisco, CA (1994).

Garcia, C. E., D. M. Prett, and M. Morari, "Model Predictive Control: Theory and Practice—A Survey," *Automatica*, **25**, 335 (1989).

Hovd, M., and S. Skogestad, "Simple Frequency-Dependent Tools for Control System Analysis, Structure Selection and Design," *Automatica*, **28**(5), 989 (1992).

Jacobsen, E. W., "Identification for Control of Strongly Interactive Plants," AIChE Meeting, San Francisco, CA (1994).

Jacobsen, E. W., and S. Skogestad, "Inconsistencies in Dynamic Models for Ill-Conditioned Plants: Application to Low-Order Models of Distillation Column," *Ind. Eng. Chem. Res.*, **33**, 631 (1994).

Lee, J. H., M. Morari, and C. E. Garcia, "State-Space Interpretation of Model Predictive Control," *Automatica*, **30**(4), 707 (1994).

Li, W., and J. H. Lee, "Control-Relevant Identification of Ill-Conditioned Multivariable Systems: Estimation of Gain Directionality," *Comp. Chem. Eng.*, **20**(8), 1023 (1996).

Ljung, L., *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, NJ (1987).

Koung, C. W., and J. F. MacGregor, "Design of Identification Experiments for Robust Control. A Geometric Approach for Bivariate Processes," *Ind. Eng. Chem. Res.*, **32**, 1658 (1993).

Ogunnaike, B. A., and Y. Arkun, "Nonlinear Model Predictive Control of High-Purity Distillation Columns Using Polynomial ARMA Models," European Control Conf. (1993).

Skogestad, S., M. Morari, and J. C. Doyle, "Robust Control Ill-Conditioned Plants: High-Purity Distillation," *IEEE Trans. Automat. Contr.*, **AC-33**(12), 1092 (1988).

Van Overschee, P., and B. De Moor, "N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems," *Automatica*, **30**(1), 75 (1994).

Varga, E. I., and S. B. Jorgensen, "Multivariable Process Identification: Estimating Gain Directions," AIChE Meeting, San Francisco, CA (1994).

Weischedel, K., and T. J. McAvoy, "Feasibility of Decoupling in Conventionally Controlled Distillation Columns," *Ind. Eng. Chem. Fundam.*, **19**, 379 (1980).

Manuscript received May 17, 1995, and revision received Feb. 29, 1996.