Identification for Decentralized Model Predictive Control

Ravindra D. Gudi
Dept. of Chemical Engineering, IIT Bombay, Powai, Mumbai, India

James B. Rawlings
Dept. of Chemical and Biological Engineering, University of Wisconsin, Madison, WI 53706

DOI 10.1002/aic.10781
Published online April 12, 2006 in Wiley InterScience (www.interscience.wiley.com).

*The problem of identifying interaction dynamics that exist between units operating in a decentralized control scheme is addressed. Identification of such interaction relationships is crucial to the deployment of coordinated decentralized control. The proposed methodology is based on a variant of the two-step, closed-loop identification methods proposed earlier in the literature. Alternative identification schemes relevant for this scenario are theoretically analyzed and are also evaluated based on the criteria of a priori knowledge necessary about the controller and the plant, as well as the applicability of the methods for the constrained controller case. Validation studies on representative systems taken from literature are presented to demonstrate the efficacy of the proposed schemes.**AIChE J,* 52: 2198–2210, 2006
Keywords: decentralized identification, closed-loop identification

Introduction

Centralized control schemes that are based on a complete description of the cause and effect relationships are known to yield optimized control performance for multivariable systems. However, for optimization and control of large-scale systems, partitioning and decentralized control schemes have been eminently recommended over centralized approaches.[1] This choice is due to a number of reasons. First, the computational intensity for control is known to grow faster than the size of the plant to be controlled. Therefore, it is practical to design and implement control schemes for smaller subsystems by looking at their local cause and effect relationships and the effect of interactions from other subsystems. The second and perhaps more important limitation for centralized control of large-scale systems is from a modeling and identification perspective. It is widely recognized that identification of cause-effect relationships in large-scale systems is a relatively difficult proposition. While such relationships are identified easily at a local level, for example, at a unit level in a chemical or power generation plant, the interaction between such levels is usually associated with a lot of uncertainty. Oftentimes, such interactions are also not perceivable during direct modeling, but manifest themselves when the local control loops are closed. Identification of such interaction dynamics is a critical requirement for implementing coordinated decentralized schemes, whose closed-loop performance can approach that of a centralized control scheme.

As indicated earlier, the interacting dynamics are often not easily characterized unless all available control inputs for the large-scale system are perturbed for identification. Therefore, such dynamics are best identified under closed-loop conditions, wherein the large-scale system is partitioned into smaller subsystems with local controllers. Having identified the interactions, coordinated and decentralized control can then be achieved with these local controllers by implementing a higher-level coordinator or a peer-level communication.[2-4] These coordination mechanisms thus require a knowledge of the interaction to achieve the desired level of coordination. This

Correspondence concerning this article should be addressed to J. B. Rawlings at jbraw@bevo.che.wisc.edu.

interaction thus needs to be identified under controlled conditions using closed-loop identification methodologies.

Closed-loop identification strategies have been extensively proposed in the literature, and an excellent review of the state of the art can be found in Forssell and Ljung.[5] The primary motivation in these strategies has been towards identifying direct models between the inputs and outputs using closed-loop data, when, for example, the plant is open loop unstable or there are inherent feedback mechanisms implemented from a safety viewpoint. Another important reason for identifying such direct models has been towards obtaining reduced order models with a view to achieving better control.[6,7] Broadly, three different approaches to closed-loop identification of the direct dynamics, that is, (i) the direct, (ii) the indirect, and (iii) the joint input-output methods, have been proposed. These methods differ in terms of a priori knowledge assumed about the nature of the controller and the assumptions made regarding the noise models. The relative merits of these strategies, in terms of consistency of estimates and the applicability of these methods (depending on the accuracy of the noise models), are discussed in Ljung[8] and Forssell and Ljung.[9] Goodwin et al.[10] discuss the duality issues that exist between constrained estimation and control.

The identification problem considered in this article involves the characterization of the interaction dynamics between decentralized control loops. The challenges encountered are somewhat similar to what one would expect for identification of the direct plant in closed loop discussed earlier. However, some key differences exist. First, as is the case with closed-loop identification, lack of informative data for identification is a key problem. This problem is overcome in regular closed-loop identification via the use of a dither signal applied either at the controller output or at the setpoint. However, for the case of identification of interaction involving several decentralized loops, dithering of each of the controller outputs is not practical. A more systematic method to prune down this set of variables that needs dithering is necessary. Second, in regular closed-loop identification, the direct method (that is, ignoring the presence of feedback) has been recommended in view of its optimal statistical properties. For the case of interaction identification, the direct method is not applicable because (as will be shown later) the inter-relationships between any two interacting decentralized loops also involve the individual loop sensitivities. Thus, one has to resort to other closed-loop identification methods to first explicitly estimate the individual loop properties and then factor them out. Finally, a recognized drawback of the indirect method of closed-loop identification is that it requires knowledge of the feedback controller mechanisms. In the identification problem considered in this article, each of the individual decentralized controllers is multivariable and perhaps constrained in nature, and therefore the use of the indirect method is precluded. Also, in view of the multivariable nature of the concerned loops and the interacting dynamics, the use of subspace based state-space identification methods[11] greatly facilitates the identification procedure.

This article analyzes the above-mentioned problems encountered in identifying the interaction dynamics in large-scale systems, with a view to achieving coordinated decentralized control. The approach proposed here to identify the interaction is based on using closed-loop data. As mentioned before, dithering all the control inputs regardless of whether they contribute to interaction is not practical. The interaction between loops in large-scale systems is generally sparse. Therefore, we first isolate the channels that contribute to interaction and minimize the number of control inputs that need to be dithered, by using partial correlation analysis. Having dithered a smaller set of relevant manipulated variables and generated sufficiently rich closed-loop data, the problem of identifying the interacting dynamics is then addressed. Towards this end, we present and analyze three methods for identification of the interacting dynamics, from the viewpoints of a priori knowledge necessary as well as the applicability in the constrained controller case. The proposed identification strategies are then validated on representative examples taken from the literature.

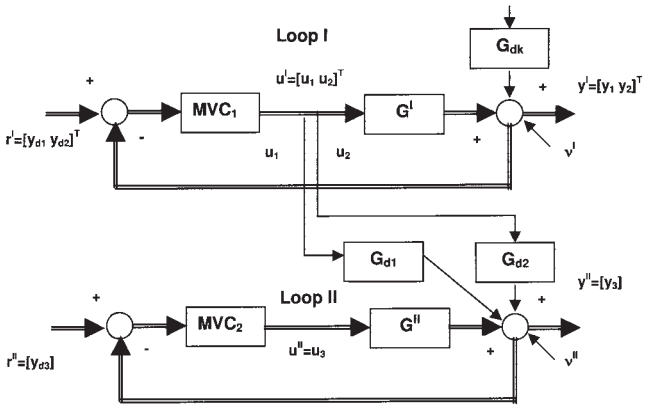

Figure 1. Interaction identification in decentralized control.

Problem Definition and Preliminaries

The identification problem that we seek to address in the article is shown in Figure 1. For simplicity of explanation, we consider the case of two decentralized loops and seek to identify the interacting dynamics $G_{dj}(q)$ between them where q is the backward shift operator ($q^{-1}\ y(t) = y(t-1)$ for any instant t).

Each of the individual loop outputs is assumed to be affected by noise and unmeasured disturbances ν. Each of the individual controllers is assumed to be multivariable in nature. For simplicity of explanation again, we assume that the controllers involved with Loop I and Loop II are multivariable and of size $n_u \times n_u$. These controllers are designed based on the local dynamics $G^I(q)$ and $G^{II}(q)$, which are assumed to be known. We further assume that the interaction dynamics G_d are sparse. This assumption is realistic because if the structure were full, one would deploy a centralized (rather than decentralized) scheme that is based on the complete enumeration of all the cause and effect relationships. In general, no other a priori knowledge is assumed about the interactions and the channels in which they could exist. We seek to estimate the interacting dynamics $G_d(q)$ under closed-loop conditions. The intent is to then use knowledge of the interacting dynamics in coordinated control schemes so as to achieve centralized performance but using decentralized control structures as shown above.

In general, closed-loop data are known to be less informative for identification from an excitation viewpoint. Hence, a dither signal d at the controller output is commonly employed. Figure

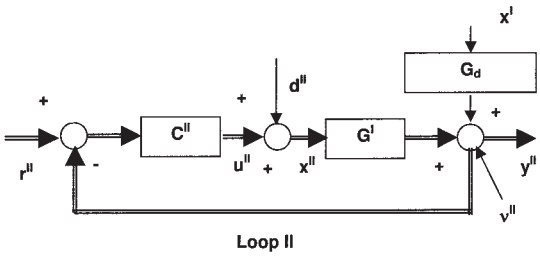

Figure 2. Block diagram of the decentralized loop.

2 shows the block diagram of a single decentralized loop with the introduction of the dither signal.

Definitions. Under the assumption of linearity for any control loop with plant G and controller C, we define the input and output loop sensitivities S_0^i and S_0 as follows:

$$S_0(q) = (I + G(q)C(q))^{-1} \tag{1}$$

$$S_0^i(q) = (I + C(q)G(q))^{-1} \tag{2}$$

These sensitivities are related to each other via the relationship

$$S_0^i(q)C(q) = C(q)S_0(q) \tag{3}$$

Also, we require the following loop property defined as

$$R_0^i(q) = (I + C(q)G(q))^{-1}C(q) \tag{4}$$

With the above definitions, the relationships between all the external inputs to the loop and dependent signals can be written as follows (Eq. 5, Figure 2). In the sequel, we will use the superscripts "I" and "II" to describe the relationship of the concerned variables to each loop. We assume that all the terms below are discrete transfer functions and drop the associated argument q for convenience.

$$\begin{bmatrix} y^{II} \\ x^{II} \\ u^{II} \\ \varepsilon^{II} \end{bmatrix} = \begin{bmatrix} S_0 G^{II} C^{II} & S_0^{II} G^{II} & S_0^{II} G_d & S_0^{II} \\ R_0^i & S_0^{i,II} & -R_0^{i,II} G_d & -R_0^{i,II} \\ R_0^i & -R_0^{i,II} G^{II} & -R_0^{i,II} G_d & -R_0^{i,II} \\ S_0^{II} G^{II} C^{II} - G_m^{II} R_0^i & S_0^{II} G^{II} - G_m^{II} S_0^{i,II} & (S_0^{II} + G_m^{II} R_0^{i,II}) G_d & S_0^{II} + G_m^{II} R_0^{i,II} \end{bmatrix} \begin{bmatrix} r^{II} \\ d^{II} \\ x^I \\ \nu^{II} \end{bmatrix} \tag{5}$$

In the above expressions, $\boldsymbol{\varepsilon}$ is the prediction error obtained by a comparison of the measured and the model predicted outputs, that is, $\varepsilon^I = (G^I - G_m{}^I)\, x^I$, where $G_m{}^I$ is the model of the plant G^I. The relevance of ε^I and $G_m{}^I$ will be shown in a subsequent section. Also, ν^I is representative of unmeasured disturbances or noise associated at the output.

It is important now to reiterate some of the challenges posed in the identification of the interacting dynamics G_d. In traditional closed-loop identification strategies for estimating the plant dynamics G^I using the direct method, the signals u^I and y^I are considered (regardless of the presence of feedback) and the model is identified using the prediction error method (PEM). This is applicable because the noise-free relationship between the signals u^I and y^I is the plant G^I and the presence of feedback does not alter this fact. On the other hand, the identification of G_d involves the use of the signals x^I and y^{II} (or some other dependent variables, such as x^{II}). For this identification, the signal u^{II} acts as a known disturbance signal that passes through the known "disturbance" dynamics G^{II}. In the presence of the unmeasured disturbance ν^I, the PEM could still be considered for identification. However, as can be seen from Eq. 5 above, the closed-loop relationship between these two signals x^I and y^{II} involves the loop sensitivity $S_0{}^i$ of Loop II. This term thus needs to be estimated explicitly and factored out of the relationship between y^{II} and x^I to generate an estimate of G_d.

The next important aspect is the application of a dither signal at the controller output(s) of Loop I, that is, the elements of the vector x^I need to be perturbed in an uncorrelated fashion to estimate the elements of G_d. Typically, the multivariable controller C^I associated with Loop I involves a relatively large number of manipulated variables and it is expected that some of the elements of G_d could be zero, that is, the interaction dynamics G_d are sparse. Therefore, a more practical approach to dithering is to first isolate those channels that show an evidence of interaction and dither only signals associated with these channels. Thus, for isolation of these channels, we require a robust strategy that uses routine closed-loop data.

Correlation analysis is often used to detect the existence of linear relationships between two or more variables. For causal systems, correlation analysis also reveals dependencies between the cause and effect variables. However, when the causal variables themselves exhibit dependence among each other, a regular correlation analysis may be misleading; it could show the existence of a relationship even though there is none. Such a scenario easily occurs when we seek to find which of the signals in x^I indeed influence Loop II. This occurrence is because each of the manipulated variables in the vector x^I emerges out of the controller and is based on the same error vector. Depending on the controller design, these manipulated variables are often correlated to some extent; this dependence in the causal variables confounds the regular correlation analysis between the cause and effect variables. Thus, a more robust strategy of assessing the existence of dependence, which takes into account the correlation in the causal variable set, is necessary. In the next section, we briefly review the partial correlation analysis as an alternative to assessing the dependencies, with an overall objective of isolating the interacting channels.

Finally, it must be mentioned that the controller C, in a model predictive control framework, is a constrained multivariable controller. Depending on the region of closed-loop operation, it is possible that the presence of constraints could bring in an overall non-linearity in the closed-loop relationships.[12] The relationships as indicated in Eq. 5 are valid for the

linear case, that is, as long as the manipulated variable constraints are not active. This nonlinear behavior also precludes the application of indirect linear closed-loop identification methodologies to estimate the interaction dynamics G_d. This issue is addressed later in the article.

The remainder of the article is organized as follows: We first discuss the partial correlation analysis that is used to isolate the interacting channels. Closed-loop identification methods for estimating the interacting dynamics are then presented. Validation studies demonstrating the practicality of the approaches are then presented, followed by concluding remarks.

Isolating Interacting Channels

Consider the block diagram shown in Figure 2. During normal plant operation, the dither signal d^{II} is zero and, therefore, the signals in the vectors x^{II} and u^{II} are the same. Routine operating data are collected on each of the variables in u^{II} and y^{II}. We seek to use this routine operating data to assess the existence of interaction, that is, we test the correlation between any signal *l* in x^{II} and a set of signals *p*, *s*, and *t* (say) in x^{I}. Since we would like to test this correlation in a dynamic sense, (i.e., the cause and effect relationships are dynamic in nature, rather than static, with at least a single lag due to zero order hold), we consider a dynamic map of these signals. Let Y denote the vector of measured observations of *l*, and X denote the matrix of measured and lagged observations of signals *p, s*, and *t,* with n being the number of lags in each of the latter signals. In other words, the k^{th} row of Y would then consist of *l(k)* and that of X would contain [p(*k-1) . . . p(k-n), s(k-1), . . . s(k-n), t(k-1). . . t(k-n)*]. The value of n can be chosen through a priori knowledge to be sufficiently large to account for the existence of the time delay and dynamics in the channel. A regular correlation analysis based on the evaluation of the Spearman correlation coefficient[13] would essentially look at a matrix Z = [Y X] and evaluate the correlation coefficients in terms of the covariance of Z. Due to finite data lengths and random variations, it is possible that a correlation coefficient may be non-zero even when no correlation exists. Therefore, these coefficients would then be tested for significance using the t-test, which essentially evaluates the probability of getting a correlation as large as the observed value by random chance when the actual correlation is zero. A small value of the probability indicates that the correlation is indeed significant. Also, due to the possibility of the variables in X being correlated, we propose to use the partial correlation analysis.[14] The partial correlation coefficient analysis can be used to negate and clearly discriminate the influence of correlated causal variables in X on the effect variable in Y.

Let us assume that the X matrix has dependent columns and, therefore, a reduced rank. Consider the problem of assessing the correlation between any column variable x_i with the variable Y. Let matrix X_R contain all variable x_j ($j \neq i$). We first find a vector e_{x_i} that contains the negation of the effect of all other variables on x_i. To do this, we first consider the regression problem between x_i and X_R. Assuming a linear relationship, the regression can be written as

$$x_i = X_R\theta_1 + E_1 \tag{6}$$

This regression problem can be solved for θ_1 in the simplest case, using the Moore-Penrose inverse as

$$\hat{\theta}_1 = (X_R^T X_R)^{-1} X_R^T x_i \tag{7}$$

We now evaluate the prediction errors for the above model as

$$e_{x_i} = x_i - X_R(X_R^T X_R)^{-1} X_R^T x_i \tag{8}$$

In the next step, the influence of X_R on Y is again estimated using a regression model. The relevant equations can be written as

$$Y = X_R\theta_2 + E_2 \tag{9}$$

from which θ_2 can be estimated as

$$\hat{\theta}_2 = (X_R^T X_R)^{-1} X_R^T Y \tag{10}$$

and the prediction errors for y can be estimated as

$$e_y = y - X_R(X_R^T X_R)^{-1} X_R^T Y \tag{11}$$

In the above equations, e_{x_i} and e_y can be assumed to consist of those components of x_1 and Y that are independent of X_R. These quantities can be used to check for the partial correlation of x_1 with y. This partial correlation can, therefore, be written as a regular correlation between e_{x_i} and e_y as

$$pr_{x_i,y} = \frac{e_{x_i}^T e_y}{(N-1)\sigma_{e_{x_i}}\sigma_{e_y}} \tag{12}$$

where N is the number of data points. It can be seen that, if X is full rank and for sufficiently large n, that is, if x_i is completely uncorrelated with all columns in X_R, then the predictions in Eq. 8 will be white and so e_{x_i} will be equal to x_1 itself. Similarly, e_y will contain only the effect of x_1 on y (that is, the effect of X_R on y will be negated). The partial correlation, along with the significance test discussed earlier, can thus give an assessment of the existence of a linear dependence between the cause and effect variables in the presence of correlated causal variables.

Remark 1. It must be mentioned that, since we use lagged variables in X, the resulting indices would indicate the existence of a linear, dynamic relationship between the signals in x^{I} (that is, *p, s*, and *t*) and *l* in x^{II}. For the variables among *p, s*, and *t* that are correlated with *l*, one would obtain symmetric blocks of non-zero correlation indices of size n, which would indicate the evidence of a non-zero interacting dynamics in G_d for the assumed pair of signals.

Remark 2. It is possible that the reduced signals e_{x_i} and e_y might also reflect the effect of unmeasured disturbances that routinely affect the loop. These effects would corrupt the interpretation from the above analysis. However, if there is any evidence of such disturbances affecting the loop, then the regression problems posed in Eqs, 6 and 9 may be solved using more sophisticated methods, such as the PEM or the instrumental variables (IV) method, rather than the simple Moore-Penrose method indicated above.

Remark 3. If the matrix X contains perfectly correlated

variables (that is, if the signals *p, s*, and *t* are perfectly correlated), the above method would still not be able to discriminate between signals that have a dependent relationship from those that do not. However, perfect correlation is difficult to obtain in practice. Even an uncorrelatedness in the measurement noise associated with the signals *p, s*, and *t* could be adequate to discriminate between these signals.

Remark 4. It must also be noted that the information content in routine closed-loop data is not persistently exciting. The steps proposed in Eqs. 8 and 11 have a further "deflating effect" on the information content of the resulting prediction error sequences. Thus, care must be taken during data collection to choose relatively rich data, for example, data during set-point changes, that reflect the interaction between the concerned loops.

Estimating Interactions: Closed-Loop Identification Methods

Having isolated the channels in which interaction exists, we now focus on the problem of identifying the interacting dynamics. For purposes of explanation, we consider that only one signal p in x^I affects Loop II (see Figure 2) through some disturbance dynamics, which need to be estimated using closed-loop identification methods. It is now assumed that the signal p can be perturbed as a dither signal at the output of controller I (MVC_1) (see Figure 1), in a way that is uncorrelated with set-point changes of Loop II. As well, either through regular target specifications by the upper LP layer in a typical MPC setting,[15] or through deliberate perturbation, it is assumed that sufficient excitation exists at the set-point for Loop II.

Using Eq. 5 and assuming the dither signal d^{II} for Loop II to be zero, the relationship between the two independent, deterministic signals p and r^{II} and the dependent signals x^{II} and y^{II} is

$$x^{II} = R_0^{iII} r^{II} - R_0^{iII} G_d p - R_0^{iII} \nu^{II} \tag{13}$$

$$y^{II} = S_0^{II} G^{II} C^{II} r^{II} + S_0^{II} G_d p + S_0^{II} \nu^{II} \tag{14}$$

As mentioned before, in regular closed-loop identification using the direct method, the effect of feedback is ignored and the signals y^{II} and u^{II} are used to estimate the transfer function G^{II}. In the above problem, we seek to estimate G_d and, in this case, the direct method of closed-loop identification is not really an option. This is because the direct method based on the signals p and y^{II} (or x^{II}) will only yield the product of G_d and one of the loop sensitivities; it is, therefore, necessary to explicitly estimate and factor out the sensitivity dynamics to generate an estimate of G_d. In earlier approaches,[16,17] an alternate two step, indirect method of closed-loop identification of G^{II} was proposed, with some of its theoretical properties being reviewed subsequently by Gevers et al.[18] Here we propose to evaluate both single step as well as two step methods for the identification of the interacting dynamics G_d. In the sequel, except in Method 3, no assumptions are made in relation to the knowledge of the plant dynamics G^{II} or the controller mechanism C^{II}. The methods for closed-loop identification of G_d can be presented as follows:

Method 1: Single step method

Consider that the closed-loop system is simultaneously perturbed both at the set-point r^{II} and at the output of C^I for signal p in an uncorrelated fashion. The dither input d^{II} is assumed to be zero in Figure 2. Using Eq. 13, it can be seen that this results in a multi-input, single output, open-loop identification problem in the variables r^{II} and p as inputs and x^{II} as the output. This identification problem can be solved in the presence of stochastic disturbances affecting Loop II, using the prediction error method (PEM). However, to achieve this solution, the stochastic component ν^{II} has to be characterized more generally. As is usually done for unknown stochastic disturbances, the sequence ν^{II} can be assumed to be an output of an unknown dynamic block H^{II} that is driven by white noise. Thus, at any instant k, we can write $\nu^{II}(k) = H^{II} e(k)$, with e being a white noise sequence. Denoting $\Gamma_1 = R_0^{iII}$, $\Gamma_2 = R_0^{iII} G_d$, and $H = R_0^{iII} H^{II}$, the model for Eq. 13 can be written as

$$x^{II}(k) = \hat{\Gamma}_1 r^{II}(k) + \hat{\Gamma}_2 p(k) + \hat{H} e(k) \tag{15}$$

From Eqs. 13 and 15, the one step ahead prediction error can be written as[8]

$$\varepsilon_x(k) = \hat{H}^{-1}[(\hat{\Gamma}_1 - \Gamma_1) r^{II}(k) + (\hat{\Gamma}_2 - \Gamma_2) p(k) + He(k)] \tag{16}$$

where $\hat{\Gamma}_1$, $\hat{\Gamma}_2$ and $\hat{H}$, are the model estimates of Γ_1, Γ_2, and H, respectively, and are assumed to be parameterized into a single parameter vector θ. Given the measured data set Z^N consisting of the N data points of the variables x^{II}, r^{II}, and p, the prediction error estimate can be written as

$$\hat{\theta}_N = \text{argmin } V_N(\theta, Z^N) \tag{17}$$

with

$$V_N(\theta, Z^N) = \frac{1}{N} \sum_{k=1}^{N} \varepsilon(k, \theta)^T \Lambda^{-1} \varepsilon(k, \theta) \tag{18}$$

where Λ is a positive definite weighting matrix. For data collected in closed loop, data availability is oftentimes not a problem and, therefore, considering in the limiting case that $N \to \infty$ (although if a dither signal is employed, N may be finite), one can write Eq. 18 in terms of the expectation as

$$V(\theta) = E\{\varepsilon(k, \theta)^T \Lambda^{-1} \varepsilon(k, \theta)\} \tag{19}$$

As shown in Ljung,[8] $V(\theta)$ can also be written in the frequency domain in terms of the spectrum of ε Φ_ε as

$$V(\theta) = E\{tr[\Lambda^{-1} \varepsilon(k, \theta) \varepsilon(k, \theta)^T]\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} tr[\Lambda^{-1} \Phi_\varepsilon] dw \tag{20}$$

Lemma 1 thus follows:

Lemma 1. Assuming that the true system parameters are contained in the model set and for the parameter identification defined by Eqs. 17-20, the parameter vector $\hat{\theta}_N \rightarrow D_c = \arg\min V(\theta)$, where D_c is given by

$$D_c = \underset{\theta}{\operatorname{argmin}} \int_{-\pi}^{\pi} tr\left\{ [\Gamma_1 - \hat{\Gamma}_1 \quad \Gamma_2 - \hat{\Gamma}_2 \quad H - \hat{H}] \times \Phi_X \begin{bmatrix} (\Gamma_1 - \hat{\Gamma}_1)^* \\ (\Gamma_2 - \hat{\Gamma}_2)^* \\ (H - \hat{H})^* \end{bmatrix} (\hat{H} \quad \Lambda H)^{-1} \right\} dw \quad (21)$$

where

$$\Phi_X = \begin{bmatrix} \Phi_{r^{II}} & \Phi_{r^{II}p} & \Phi_{r^{II}e} \\ \Phi_{pr^{II}} & \Phi_p & \Phi_{pe} \\ \Phi_{er^{II}} & \Phi_{ep} & \Phi_e \end{bmatrix}$$

is the matrix of (cross) spectrums and the * denotes the complex conjugation.

Proof. The proof is a straightforward extension of the SISO case considered in Forssell and Ljung.[50] Rewriting Eq. 16 as

$$\varepsilon_x(k) = \hat{H}^{-1}[(\hat{\Gamma}_1 - \Gamma_1)r^{II}(k) + (\hat{\Gamma}_2 - \Gamma_2)p(k) + (H - \hat{H})e(k)] + e(k) \quad (22)$$

the spectrum of ε_x can then be written as

$$\Phi_\varepsilon = \hat{H}^{-1}[\Gamma_1 - \hat{\Gamma}_1 \quad \Gamma_2 - \hat{\Gamma}_2 \quad H - \hat{H}]\, \Phi_X \begin{bmatrix} (\Gamma_1 - \hat{\Gamma}_1)^* \\ (\Gamma_2 - \hat{\Gamma}_2)^* \\ (H - \hat{H})^* \end{bmatrix} \hat{H}^{-*} + \Lambda_0 \quad (23)$$

It can be seen that the last term is independent of θ, and the result follows from substituting Eq. 23 into Eq. 20 and using the matrix property that trace(AB) = trace(BA).

Thus, as long as the models $\hat{\Gamma}_1$, $\hat{\Gamma}_2$, and $\hat{H}$ are chosen flexibly so that the true system is contained in the model set, a positive definite Φ_X ensures consistent model estimates. By design, if p and r^{II} signals are chosen to be uncorrelated and to have a high signal to noise ratio relative to the white noise sequence w, Φ_X is diagonal and positive definite and will satisfy the requirements as outlined in Eqs. 20-23. The signals p and r^{II} themselves need to be designed based on a knowledge of the dominant time constants of the plant. This aspect is more clearly discussed in the case study outlined below. Having obtained consistent estimates of Γ_1 and Γ_2, one can then estimate G_d from the definition as,

$$G_d = \Gamma_1^{-1}\Gamma_2 \quad (24)$$

An important aspect that must be recognized towards realization of G_d as above is that Loops I and II are multivariable in nature and could involve a large number of manipulated and controlled variables. Towards ease of parameterization of the dynamics in Γ_1 and Γ_2 during identification, it is more suitable to use state space models of appropriate order, identify them using suitable subspace methods, and then realize the transfer function representations of G_d, if necessary, as indicated by Eq. 24 above. However, as will be seen in the next section, inverses are also easily realizable in state space forms and the entire model identification steps can be performed in the state space domain.

Method 2: Two step method

The single step method requires simultaneous dither perturbation of the set-point and the manipulated variable. This could be a deterrent in many closed-loop process identification applications. However, one could take advantage of the fact that the set-point is normally perturbed during closed-loop operation by an upper optimization layer.[15] In such routine closed-loop operation involving relatively weaker set-point perturbation, availability of the data is not really a constraint.

In the first step of the proposed two step method, one could therefore seek to estimate the dynamic term Γ_1 using routine closed-loop data. Towards this identification, Eq. 13 can be rewritten as

$$x^{II}(k) = \hat{\Gamma}_1 r^{II}(k) + \underbrace{\hat{\Gamma}_2 p(k) + \hat{H}e(k)}_{H_1 e(k)} \quad (25)$$

In the above equation, the effect of interactions through the input p(k) is lumped with the stochastic components and is modeled together in terms of the dynamics H_1 that is driven by a white noise sequence e.

The one step ahead prediction error for the identification of $\hat{\Gamma}_1$ can then be written as

$$\varepsilon_x(k) = \hat{H}_1^{-1}[(\hat{\Gamma}_1 - \Gamma_1)r^{II}(k) + H_1 e(k)] \quad (26)$$

As before, the objective function for identification can be written in the frequency domain in terms of the power spectrum of ε. At this stage, it is important to recognize two different kinds of noise parameterizations that could be used for the above estimation. In the identification literature, the prediction error method (PEM) has been proposed to estimate the model parameters (including those of the stochastic components) and it has been shown to generate statistically efficient estimates of both the deterministic and stochastic components of the process.[8] The PEM therefore uses a flexible noise parameterization. The output error method (OEM) is a special case of the PEM where the noise or the stochastic component model $\hat{H}_1$ is parameterized as unity (that is, fixed noise parameterization). In other words, the noise model is not estimated in the OEM. To solve the above identification problem in terms of minimizing the prediction errors defined in Eq. 26, we propose to use the OEM to estimate the dynamics Γ_1. The objective function for identification can then be written as shown in Eq. 20.

The following Lemma can then be written:

Lemma 2. Assuming that the true dynamics of G_1 are contained in the model set and for the parameter identification

defined by Eqs. 20 and 26, the parameter vector $\hat{\theta}_N \rightarrow D_c = \arg\min V(\theta)$, where D_c is given by

$$D_c = \underset{\theta}{\operatorname{argmin}} \int_{-\pi}^{\pi} tr\{[\Gamma_1 - \hat{\Gamma}_1]\Phi_{r^{II}}[(\Gamma_1 - \hat{\Gamma}_1)^*]\}dw \quad (27)$$

Proof. Rewriting Eq. 26 as

$$\varepsilon_x(k) = \hat{H}^{-1}[(\hat{\Gamma}_1 - \Gamma_1)r^{II}(k) + (H_1 - \hat{H}_1)e(k)] + e(k) \quad (28)$$

the power spectrum of ε(k) can be written as

$$\Phi_\varepsilon = \hat{H}_1^{-1}[\Gamma_1 - \hat{\Gamma}]\Phi_{r^{II}}[(\Gamma_1 - \hat{\Gamma}_1)^*] + [(H_1 - \hat{H}_1)]\Phi_e[(H_1 - \hat{H}_1)^*] + \Lambda_0 \quad (29)$$

It can be seen that with a choice of $\hat{H}_1 = 1$, the last two terms are independent of the parameter vector θ. The result follows after substituting Eq. 29 into 20.

Therefore, so long as the model $\hat{\Gamma}_1$ is chosen flexibly such that the true system is contained in the model set and Φ_r is diagonal, the minimization of V ensures consistent estimates of the Γ_1.

In practical terms, the following comments can be made about the optimality of the estimation of Γ_1. The first aspect is related to the use of the OEM over the PEM for the identification. As stated earlier, the PEM is a statistically efficient method to estimate the deterministic and stochastic components; it can be shown to lead to maximum likelihood estimates whose variance can asymptotically meet the Cramer-Rao lower bound. On the other hand, the OEM guarantees consistency only in the deterministic part and for fixed data size; the OEM estimates have a larger variance (and, hence, smaller confidence in the parameter estimates) than the PEM estimates. However, in closed-loop identification, data size is not a constraint and, therefore, this advantage offered by the PEM method is not very significant. Also, the OEM involves a simple pseudo linear regression in the minimization and is relatively easier to implement than the PEM. The second aspect is related to the requirement that Φ_r be diagonal. If one uses only the data resulting from the perturbations of the upper optimization layer, this requirement cannot necessarily be met. It may be necessary to augment the perturbations from the optimization layer with small and uncorrelated dither perturbation to ensure consistent identification of Γ_1.

Having obtained consistent estimates of Γ_1 in the first step, the next step would involve the estimation of the interacting dynamics G_d. In this second step, the dither at the set-point r^{II} is set to zero and the signal p is dithered. The signals x^{II} and p can now be used to estimate G_d. Towards this end, a new signal x^{II}_f is generated as a filtered version of x^{II} as given by

$$x_f^{II} = (\Gamma_1)^{-1}x^{II} \quad (30)$$

From Eq. 15, this filtered signal x^{II}_f can then be used along with the measurements of p to generate an estimate of the interacting dynamics G_d using the OEM.

A critical step in the above filtering is the realization of the inverse of Γ_1 as required in Eq. 30. It is well known that multivariable systems are more easily parameterized in state space form, rather than their transfer function counterparts. In the context of realizing the inverse of Γ_1 when the dynamics is parameterized in state space form, the expressions relating the state space matrices A, B, C, and D of the forward model, to the state space matrices Φ, Γ, C_i, and D_i of the inverse model can be written as (see Appendix):

$$\Phi = A - BD^{-1}C; \Gamma = BD^{-1}$$
$$C_i = -D^{-1}C; D_i = D^{-1} \quad (31)$$

Thus, if D^{-1} exists, the inverses are realizable. In the above identification, it can be seen that the dynamics in Γ_1 are bi-proper and so the inverse of D always exists. Therefore, these transformations are realizable.

Method 3: Direct method: Constrained controller case

Model predictive controllers are implemented with various constraints on the inputs and outputs. When such constraints are active, it is well-known[12] that the controller dynamics and, hence, the loop sensitivities are no longer linear. The data collected under closed loop would reflect these nonlinear effects. In such a case, it is inappropriate to use the methods proposed in the earlier subsections, which are based on the assumption of linearity. We, therefore, propose to exploit the fact that the plant dynamics G^{II} is well characterized. This latter assumption is valid considering that cause and effect relationships can be more easily characterized in smaller decomposed subsystems.

From Eq. 5, the relationship between the prediction error ε and the external signals affecting Loop II can be written as

$$\varepsilon = (S_0^{II}G^{II}C^{II} - G_m^{II}R_0^{iII})r^{II} + (S_0^{II}G^{II} - G_m^{II}S_0^{iII})d^{II} + (S_0^{II} + G_m^{II}R_0^{iII})G_d x^I + (S_0^{II} + G_m^{II}R_0^{iII})\nu^{II} \quad (32)$$

In the presence of an accurate model, $G_m{}^{II} = G^{II}$ and for signals r^{II} and d^{II} set to zero, the above relationship can be shown to reduce to (see Appendix)

$$\varepsilon = G_d x^I + \nu^{II} \quad (33)$$

Clearly, the relationship shown in Eq. 33 is independent of the controller and is, therefore, linear if the concerned disturbance dynamics are linear. Thus, in the presence of stochastic disturbances affecting at the output, a prediction error method based on the constructed signal ε and the measured signal p can be used to estimate the dynamics G_d.

A relative assessment of the above three methods of identifying the interacting dynamics can be presented as follows: In the presence of an accurate model $\mathbf{G_m}^{II}$, one would always prefer Method 3 as it is a single step, direct method that accommodates the most general constrained controller case. It does not require any other knowledge of the controller mechanisms or the loop sensitivities and, as such, is a parallel counterpart to the Direct Identification scheme proposed for identifying direct models in closed loop. However, it is based

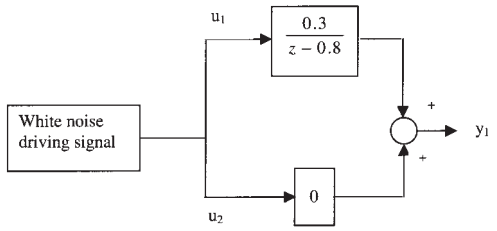

Figure 3. Block diagram for illustrative Study 1.

on the key assumption of an accurate plant model. If this assumption cannot be justified, one would then choose between the first two methods.

The merits of the first method is that it is a one step method that just uses closed-loop data to simultaneously identify the concerned dynamics in an MISO framework. However, the drawback is that sufficiently strong and uncorrelated dither has to be applied simultaneously at both the set-point and at the manipulated variables to generate unbiased models. This drawback may pose problems in terms of the acceptance of this strategy in practice. The second method, on the other hand, is a two step method that involves the identification of the loop sensitivities in the first step. This identification may be done without strong dithering at the set-point. Regular set-point updating by the upper LP layers could be sufficient to provide the necessary excitation. The identification could be performed using data collected over extended periods of time; for the closed-loop case, availability of such data is not a constraint. After this first step, regular dithering of the manipulated variables of the other controllers over a short period of time is adequate to generate the excitation necessary for identification of the interacting dynamics. Thus, the second method is preferable when dithering at the set-point is not permissible.

Validation Studies

Partial correlation analysis: Numerical study

This numerical example is considered to illustrate the discriminating ability of partial correlation analysis and its use in isolating channels in which interaction is present, in the presence of correlated causal variables. Consider the transfer function models shown in the block diagram of Figure 3. The signals u_1 and u_2 are shown to affect the output y through two transfer function blocks. The transfer function between u_1 and y is set to be zero. In order to simulate a correlation between the two causal variables u_1 and u_2 (as would be the case when the control moves are updated by a centralized controller), the limiting case when they are exactly the same is considered. The signals u_1, u_2, and y are assumed to be measured with random noise of intensity 0.1. The white noise driving signal is designed to have an intensity of 3. The correlations between the causal and effect variables were assessed for the dynamic case by considering three past lags of the causal variables.

The results of the study are shown in Table 1. The regular (Spearman) correlation indices show the existence of a linear relationship between both the causal inputs u_1 and u_2 and the output variable y. Regular correlation analysis is thus not able to resolve u_1 and u_2 in terms of their effect on y. This inability arises because u_1 and u_2 are by themselves correlated. The partial correlation analysis, on the other hand, is able to discriminate and indicate the channels in which a dynamic relationship exists. The absence of a correlation is seen by relatively smaller indices for the relationship between y(k) and $u_2(k-1)$, $u_2(k-2)$, and $u_2(k-3)$. As seen in Eqs. 6-12, the partial correlation analysis accommodates the existence of correlation in the causal variables. The correlation in the causal variables is suppressed via the solution of two local regression problems, and the uncorrelated components that report in the residuals are then used to check for the existence of a correlation. Thus, partial correlation analysis is able to clearly test for the existence of a relationship in the presence of correlated causal variables.

Isolating channels in a polymerization reactor: A simulation study

We next consider the solution copolymerization of methylmethacrylate (MMA) and vinyl acetate (VA) application reported in Congalidis et al.[19] The transfer function model between the inputs and the outputs is as shown in Table 2. The system consists of 4 outputs and 5 inputs, as indicated in Table 2. Congalidis et al. suggested a decentralized approach to control based on a decomposition using RGA analysis. Based on their analysis, the overall system can be partitioned into three decentralized loops, as shown in Figure 4.

Direct synthesis controllers that were based on a closed-loop time constant of half the dominant time constant for each loop were designed for the plant. Further, a dynamic decoupler was also implemented in the first decentralized loop, as shown in Figure 4. The decoupler design thus introduced a correlation among the two controller outputs u_2 and u_3. The second and third decentralized loops, involving the variables u_4 and u_5, were single input single output loops. The input variable u_1 was held constant, as recommended by Congalidis et al. The closed-loop outputs were assumed to be affected by i.i.d. noise having intensity 0.1.

As before, to test for the existence of a correlation, both regular and partial correlation analysis were used. Since the variability in the plant outputs gets transferred to the inputs due the presence of the controllers, these correlation tests were based on the current and past measured values of the controller

Table 1. Results of Correlation and Partial Correlation Analysis on the Numerical Study

	Regular (Spearman) Correlation Coefficients							
	$u_1(k)$	$u_1(k-1)$	$u_1(k-2)$	$u_1(k-3)$	$u_2(k)$	$u_2(k-1)$	$u_2(k-2)$	$u_2(k-3)$
$y(k)$	0.0406	0.6009	0.5083	0.4082	−0.0024	0.405	0.3476	0.2927
	Partial Correlation Coefficients							
	$u_1(k)$	$u_1(k-1)$	$u_1(k-2)$	$u_1(k-3)$	$u_2(k)$	$u_2(k-1)$	$u_2(k-2)$	$u_2(k-3)$
$y(k)$	0.0295	0.6179	0.5346	0.4861	−0.0029	−0.0043	−0.0091	−0.0463

Table 2. Transfer Function Matrix for Copolymerization Application (Congalidis et al.[19])

$$G(s) = \begin{bmatrix} \frac{0.34}{0.85s+1} & \frac{0.21}{0.42s+1} & \frac{0.50(0.50s+1)}{0.12s^2+0.40s+1} & 0 & \frac{6.46(0.9s+1)}{0.07s^2+0.30s+1} \\ \frac{-0.41}{2.41s+1} & \frac{0.66}{1.51s+1} & \frac{-0.3}{2.71s+1} & 0 & \frac{-3.72}{0.80s+1} \\ \frac{0.30}{2.54s+1} & \frac{0.49}{1.54s+1} & \frac{-0.71}{1.35s+1} & \frac{-0.20}{2.71s+1} & \frac{-4.71}{0.08s^2+0.41s+1} \\ 0 & 0 & 0 & 0 & \frac{1.03(0.23s+1)}{0.07s^2+0.31s+1} \end{bmatrix}$$

y_1 = polymer production rate
y_2 = mole fraction (MMA) in copolymer
y_3 = weight average mol. wt
y_4 = reactor temperature

u_1 = VA monomer flow rate
u_2 = MMA monomer flow rate
u_3 = flow rate of initiator
u_4 = flow rate of chain transfer agent
u_5 = reactor jacket temperature

outputs u_2, u_3, u_4, and u_5, assuming that the plant outputs were well-regulated. Based on the t-test for significance, a value of 0 or 1 was assigned to each variable pair depending on the existence of a correlation between them.

The results of the correlation analysis are shown for two cases in Table 3. In the first case (Case a), the full plant as shown in Table 2 is considered. The second case (Case b) assumes the transfer function between y_3 and u_2 to be zero. This latter case is considered to again highlight the discriminative ability of partial correlation analysis. In the first case, both regular and partial correlation analysis gave identical results, as seen in Table 3a. The correlation between u_4 as a causal variable and u_2 and u_3 as effect variables is expectedly zero because u_4 does not affect the outputs y_1 and y_2 (see the plant transfer function matrix in Table 2). Similarly, y_3 is affected by all other controller outputs and, therefore, the dynamic correlations between u_4 (which is paired with y_3) and u_2, u_3, and u_5 are all significant. Furthermore, y_4 is not affected by any other input variable except u_5 and, therefore, the correlation between the latter and the other controller outputs is zero. When the correlation analysis is repeated for Case b, it is seen that partial correlation analysis gives an accurate representation of the interactions in the presence of correlated causal variables. These results can be explained as follows: For Case b, the dynamics between y_3 and u_2 are assumed to be zero. However, the pair $\{y_1,y_2\}$ and $\{u_2,u_3\}$ form a centralized controller and, as such, it is expected that u_2 and u_3 would be strongly correlated. When one tests the correlation between u_2 and u_3 with u_4, regular correlation analysis indicates the existence of a correlation between these variables. However, since

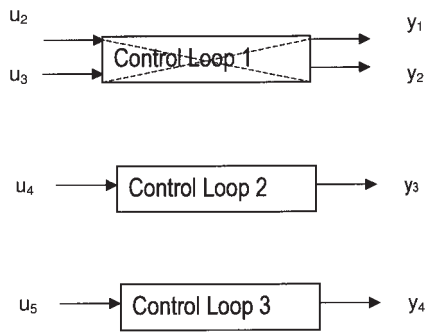

Figure 4. Decentralized control loops for the copolymerization application.[19]

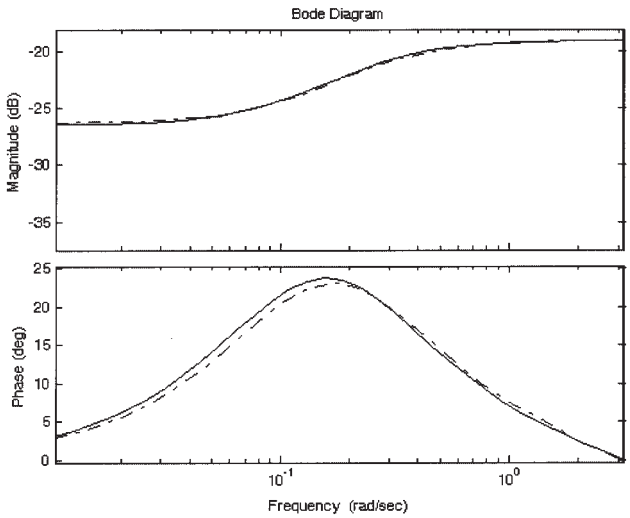

Figure 5. Comparison of the estimated and theoretical R_O^i (theoretical: solid line; estimate: dash-dot).

Table 3. Results of Partial Correlation Analysis on the Polymerization Example

(3a)

	Regular					Partial			
	u_2	u_3	u_4	u_5		u_2	u_3	u_4	u_5
u_2	1	1	0	1	u_2	1	1	0	1
u_3	1	1	0	1	u_3	1	1	0	1
u_4	1	1	1	1	u_4	1	1	1	1
u_5	0	0	0	1	u_5	0	0	0	1

(3b)

	Regular					Partial			
	u_2	u_3	u_4	u_5		u_2	u_3	u_4	u_5
u_2	1	1	0	1	u_2	1	1	0	1
u_3	1	1	0	1	u_3	1	1	0	1
u_4	1	1	1	1	u_4	0	1	1	1
u_5	0	0	0	1	u_5	0	0	0	1

the transfer function between y_3 and u_2 is set to zero (Case b), we know that u_2 does not affect y_3 and therefore u_4 is also unaffected. Regular correlation analysis thus fails to discriminate between the variables u_2 and u_3, and this can be attributed to the fact that the latter are strongly correlated. The results for partial correlation analysis, on the other hand, give a relatively accurate representation of the interaction between the loops, as seen in Table 3b.

The implications of these results for large-scale systems are, therefore, evident. When large-scale systems are partitioned and controlled using decentralized schemes, the existence of interactions between these decentralized controllers can be more accurately verified using partial correlation analysis. Having identified the channels of interaction between these subsystems, the task of dither signal design and identification of the interacting dynamics becomes a relatively easier task. This aspect is considered in the following case study.

Closed-loop identification of interaction: A simulation study

We consider the following 3×3 transfer function and choose to partition the system for decentralized control as shown below:

$$G(s) = \left[\begin{array}{cc:c} \dfrac{12.8}{16.7s+1} & \dfrac{-18.9}{21s+1} & 0 \\ \dfrac{6.6}{10.9s+1} & \dfrac{-19.4}{14.4s+1} & 0 \\ \hdashline 0 & \dfrac{1.5}{4.48s+1} & \dfrac{21}{10s+1} \end{array}\right]$$

The input-output pair $\{y_1, y_2 - u_1, u_2\}$ is treated as a decentralized subsystem that is controlled by two PI controllers ($K_c = 0.47$ and $\tau_I = 3.26$ for the $y_1 - u_1$ pairing and $K_c = -0.09$ and $\tau_I = 9.35$ for the $y_2 - u_2$ pairing) with decouplers. The pair $y_3 - u_3$ forms another decentralized subsystem, which is also regulated by a PI controller with $K_c = 0.09$ and $\tau_I = 10$. The two decentralized loops interact via the dynamics shown in the $(3,2)^{th}$ element of the matrix shown above. The objective of the closed-loop identification was to identify this interacting dynamics. The two loops were assumed to be corrupted at the output with random noise having variance of 0.1. For closed-loop identification, the set-points and the controller outputs were dithered using random signals having variance 0.3. These were designed to be uncorrelated whenever necessary, as, for example, in Method 1. For Method 2, simple step like set-point changes were introduced in the first step to estimate the sensitivities. For Method 3, the controller outputs were clipped using an appropriate saturation block, to simulate constrained controller action. The canonical variate analysis (CVA) based subspace method was used for the identification of the state space models using closed-loop data. The model order for the state space models was selected based on the Akaike Information Criteria, and the models identified were cross-validated using closed-loop data. A sampling time of 1 time unit was assumed in all the simulations.

Table 4. Results of Estimation of the Interacting Dynamics Using Method I

	R_0^i	G_d
Theoretical	$\dfrac{10.45z^3 - 29.40z^2 + 27.55z - 8.6}{100z^3 - 269.6z^2 + 241.1z - 71.5}$	$\dfrac{0.3}{z - 0.8}$
Estimated	$\dfrac{0.1034z^3 - 0.0848z^2 + 0.0407z - 0.043}{1z^3 - 0.6889z^2 + 0.40z - 0.359}$	$\dfrac{0.2991}{z - 0.8003}$

Method 1: Based on simultaneous perturbation of the dither (Loop I) and set-point of Loop II

Assuming that the interacting channel(s) are known ($u_2 - y_3$ in the above case) via the use of partial correlation analysis, two uncorrelated dither signals were designed to be applied at the controller output u_2 and at the set-point of Loop II. These signals were designed as maximum length binary sequences, based on an approximate characterization of the dominant time constants of the dynamics to be identified (that is, the terms R_0^i and its product with G_d), and were designed to be uncorrelated following the methods reported in Godfrey.[20] For example, for Loop II, the expected closed-loop time constant was assumed to be about half of the open-loop dynamics (5 time units) and the perturbation sequence was accordingly designed. To achieve uncorrelatedness of the second signal with respect to this signal, a particular choice, namely, the inverse repeat sequence[20,21] was made. Based on the guidelines for input signal design, the total experiment time should be 5 times the settling time, which for the above system was 130 time units. For the identification, the AIC indicated a model order of 3. The identified state space models were converted into transfer function form and are shown in Table 4 for comparison with the true values. Figure 5 shows a comparison of the estimated R_0^i with the true value in the frequency domain. It can be seen from this figure that the estimated loop properties match quite well with the theoretical values for R_0^i. Also, as seen from Table 4, the estimate of the interacting dynamics agrees quite well with the true value.

It must be mentioned here that in practice, the effect of unmeasured disturbances on the identification of the above loop properties also needs to be considered. These disturbances can be accommodated if a priori knowledge about them is available. This a priori knowledge can be used in appropriate dither signal design to make the latter orthogonal to the disturbances. The PEM can also be used during the identification step to generate unbiased estimates of the loop properties.

One of the main drawbacks of this method lies in the simultaneous excitation of each interacting loop, which would introduce significant variability in the outputs of each loop. As mentioned earlier, natural excitation via set-point perturbations can be used as an alternative to obtain estimates of the loop properties. This approach is discussed in the following subsection.

Method 2: Two step method for estimating the interacting dynamics

As explained before, the central objective in this approach is to achieve less output variability during the identification. Towards this end, in the first step, the loop property R_0^i is estimated first from routine set-point perturbations of Loop II.

Table 5. Results of the Estimation of the Interacting Dynamics Using Method II

State space matrices for the estimation of R_0^i using CVA for filter time constant of 5 time units:

$$\Phi = \begin{bmatrix} 0.44 & -0.73 & 0.085 & -0.085 \\ -0.832 & -0.045 & -0.1077 & -0.0833 \\ -0.84 & -0.634 & 0.5037 & -0.6085 \\ 3.277 & 2.9247 & 0.3463 & 0.6982 \end{bmatrix}; \quad \Gamma = \begin{bmatrix} 1.35 \\ 3.01 \\ 2.18 \\ -8.50 \end{bmatrix}$$

$$H = [0.004 \quad -0.016 \quad 0.0015 \quad -0.0039]; \; A = [0.1033]$$

	R_0^i	G_d
Theoretical	$\frac{10.45z^3 - 29.40z^2 + 27.55z - 8.6}{100z^3 - 269.6z^2 + 241.1z - 71.5}$	$\frac{0.3}{z - 0.8}$
Estimated for a filter time constant of 5 time units	$\frac{0.1033z^4 - 0.1714z^3 + 0.0889z^2 - 0.0091z - 0.005}{z^4 - 1.60z^3 + 0.934z^2 - 0.243z + 0.031}$	$\frac{0.3027}{z - 0.7815}$
Estimated for a filter time constant of 10 time units	$\frac{0.0977z^4 - 0.146z^3 + 0.063z^2 - 0.0031z - 0.0027}{z^4 - 1.54z^3 + 0.95z^2 - 0.38z + 0.117}$	$\frac{0.3194}{z - 0.7344}$

Unlike the earlier approach, no dither signals are applied at the controller output of Loop I during the estimation of R_0^i. In this first step, the output error method is used to estimate the loop property R_0^i (see discussion after Lemma 2) in the presence of noise affecting the output and the interaction related disturbances.

Square wave signals having a unit amplitude and frequency 0.05 hertz were suitably pre-filtered through a first order transfer function and were applied at the set-point to simulate gradual set-point changes by an upper optimizer layer. The time constant of this first order filter was varied to observe the effect of errors in estimation of R_0^i on the estimate of the interacting dynamics. A total of 1000 samples were required to generate approximately the same accuracy of estimates for R_0^i that was reported in Method I. For the estimation of G_d, the controller output u_2 was perturbed in the same fashion as described in Method I.

Table 5 shows the state space matrix for a single case and also the effect of two different filter time constants on the estimation of the interaction. Figure 6 shows a comparison of the estimated R_0^i with the true value in the frequency domain as a function of the filter time constant. It can be seen that the interacting dynamics G_d is accurately identified, although there is deterioration in the accuracy when the filter time constant is increased to 10 time units. This is expected because of the relative decrease in excitation for the latter case. The other important point to note is that the experiment time for Method II was much longer than for Method I.

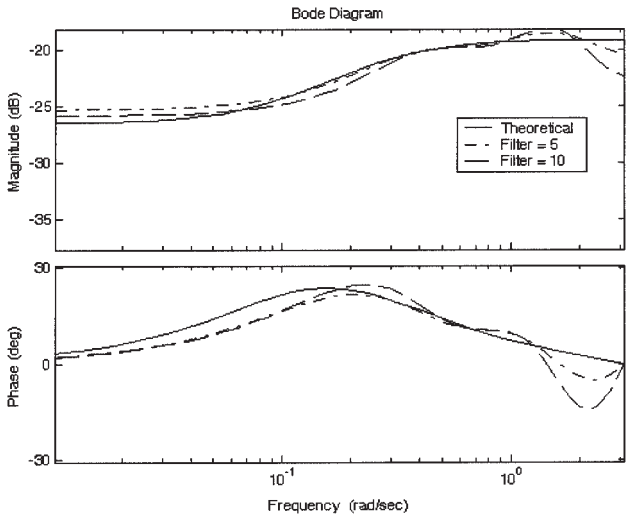

Figure 6. Comparison of the estimated and theoretical R_0^i for Method II.

In practice, both the perturbation at the set-point as well as the experiment times are important considerations. The routine set-point changes made by the optimizer layer may have to be augmented with low intensity changes so that the loop property R_0^i is estimated accurately within a reasonable experiment time. In any case, availability of data in closed-loop operation is not a constraint. If output variability is an important consideration, Method II would be a suitable alternative to Method I.

Method 3: Constrained controller case

As mentioned earlier, the above two methods are applicable when the controller deployed in the loops is linear. However, in the presence of a constrained multivariable controller, the loop behavior may not be linear. To simulate these constraints, the controller output for Loop II was clipped using a saturation block to lie between ± 0.01 for the loop perturbations described above. The data used for the identification are shown in Figure 7.

Under the assumption that the plant for the decentralized loop is accurate, the residuals ε between the plant output and the model predictions essentially consist of the interaction effects, measurement noise, and unmeasured disturbances. The output error method was used to estimate the interaction dynamics between u_2 and ε. The interaction dynamics were estimated accurately, as shown in Table 6. When Methods 1 and 2 were applied to this data, the estimation of the (linear) loop property R_0^i was expectedly not accurate and in some cases did give rise to unstable models due to the linear parameterization of the nonlinear dynamics. As expected, the estimation of the interaction dynamics was also biased.

From a practical viewpoint, this method is the simplest to use but it requires a major assumption of an accurate model. This assumption could be justified by arguing that for large multivariable plants, the major source of uncertainty is in the interactions[1] and the local dynamics (being smaller in size)

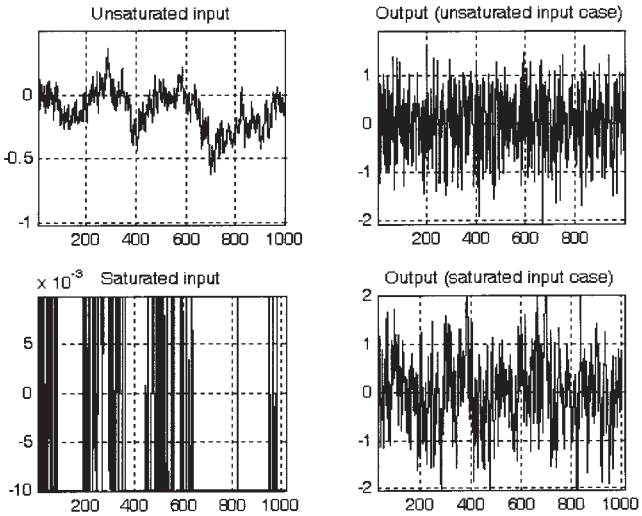

Figure 7. Data used for the validation exercise of Method III.

could be estimated fairly accurately. However, if this assumption is not justifiable, then other methods of characterizing and estimating the loop properties for nonlinear (constrained controller) loops will have to be used. Of particular interest is a recent approach proposed by Forsell and Ljung,[9] which considers the projection method to address nonlinearities in closed-loop data. Another possible approach would be to examine and exploit the decorrelating effect that a nonlinear constrained controller would have on the concerned loop signals. These aspects are potential areas for further study.

Conclusions

In this article, we have proposed three methods for identifying the important, neglected dynamic interaction models in a large, decentralized control system. Because the interactions between units in a large, decentralized plant are likely to be sparse, prior to any of the three identification methods, we first recommend partial correlation analysis to select the important neglected interaction channels. This step provides a large reduction in the amount of identification effort required for the plant. The partial correlation analysis was shown through examples to be effective even when several of the inputs are correlated, which is the case expected in industrial practice.

After the important channels are selected, Methods 1 and 2 assume a linear controller and apply dither signals to the appropriate signal locations. The theory underlying both methods was presented, and both methods are shown via simulation to obtain accurate models from closed-loop operating data. Method 1 requires two uncorrelated dithers (at the set-point and the input) and obtains the model with less data than Method 2, which requires only one dither (at the input) and relies also on the usual set-point variations in the closed-loop operation. Method 1 will produce accurate models with less data than Method 2, but at the cost of introducing more variability in the output during the collection of the data. Because it adds less variability, Method 2 may be preferred when the normal closed-loop operation contains many set-point changes.

If the controller constraints become active often during closed-loop operation, the controller's nonlinearity is significant and neither Method 1 nor 2 provide accurate interaction models. If the local model used in the decentralized controller is accurate, however, then Method 3 can be used to obtain an accurate interaction model. This method, therefore, complements Methods 1 and 2. How to treat the case in which the controller nonlinearity is significant and an accurate local model in the decentralized loop is not available remains a topic of current research in identification.

Table 6. Results of the Estimation of the Interacting Dynamics Using Method III

	True	Estimated
Interaction dynamics G_d	$\frac{0.3}{z-0.8}$	$\frac{0.31}{z-0.7998}$

Acknowledgments

The authors acknowledge the financial support of the Paul A. Elfers Fund, which provided funding for a sabbatical visit of the first author at the University of Wisconsin in 2003. The second author acknowledges the financial support of the NSF through grant CTS–0456694. Extensive research discussions with Dr. P. Vijaysai, Dr. Nabil Jabbar and Aswin Venkat are also acknowledged.

Literature Cited

1. Siljak DD. Decentralized control and computations: status and prospects. *Annual Reviews in Control.* 1996;20:131-142.
2. Mesarovic MD, Macko D, Takahara. Two coordination principles and their application in large scale systems control. *Automatica*. 1970;6: 2647-2473.
3. Katebi MR, Johnson MA. Predictive control design for large scale systems. *Automatica.* 1997;33(3):421-425.
4. Venkat A, Rawlings JB, Wright S. Plant-wide optimal control with decentralized MPC. In: *Proceedings of International Symposium on Dynamics and Control of Process Systems (DYCOPS2004)*. Cambridge, MA: Paper No. 190. Elsevier Press; 2004.
5. Forssell U. Ljung L. Closed-loop identification revisited. *Automatica.* 1999;35(7):1215-1241.
6. Gevers M, Ljung L. Optimal experiment designs with respect to the intended model application. *Automatica.* 1986;22(4):543-555.
7. Zang Z, Bitmead RR, Gevers M. Iterative weighted least squares identification and weighted LQG control design. *Automatica.* 1995; 31(11):1577-1594.
8. Ljung L. *System Identification: Theory for the User*. NJ: Prentice Hall; 1997.
9. Forssell U, Ljung L. A projection method for closed loop identification. *IEEE Trans on AC.* 2000;45(11):2101-2105.
10. Goodwin GC, Graebe SF, Salgado ME. *Control System Design*. NJ: Prentice Hall; 2000.
11. Verhaegen M, Yu X. A class of subspace model identification algorithms to identify periodically and arbitrarily time-varying systems. *Automatica*. 1995;31(2):201-216.
12. Zafiriou E. Robust model predictive control of processes with hard constraints. *Computers and Chemical Engineering.* 1990;14(4/5):359-371.
13. Draper NR, Smith H. *Applied Regression Analysis*. New York: John-Wiley; 1993.
14. Smillie KW. *Introduction to Regression and Correlation*. Toronto: Academic Press; 1966.
15. Shah SL, Huang B. Patwardhan RS. Multivariate controller performance analysis: methods, applications and challenges. In: *CPC–VI, Proceedings of Sixth International Conference on Chemical Process Control*. Rawlings JB, Ogunnaike BA, Eaton JW, eds. AIChE Symposium Series. 2002;98:187-219.
16. Van den Hof PMJ, Schrama RJP. An indirect method for transfer

function estimation using closed loop data. *Automatica.* 1993;29(6): 1523-1527.
17. Huang B, Shah SL. Closed-loop identification: a two step approach. *Journal of Process Control.* 1997;7(6):425-438.
18. Gevers M, Ljung L, Van den Hof P. Asymptotic variance expressions for closed loop identification. *Automatica.* 2001;37:781-786.
19. Congalidis JP, Richards JR, Ray WH. Modeling and control of co-polymerization reactor. *Proceedings of American Control Conference*. 1986;1779-1793.
20. Godfrey K. *Perturbation Signals for System Identification*. New York: Prentice Hall; 1993.
21. Srinivasan R, Raghunathan R. Use of inverse repeat sequence for identification in chemical process systems. *Industrial and Engineering Chemistry Research.* 1999;38(9):3420-3429.

Appendix

Relationship Between forward and inverse state space models

Let the dynamics of a state space model be represented as

$$X(k+1) = AX(k) + Bu(k)$$

$$y(k) = CX(k) + Du(k) \tag{A1}$$

Rewriting the above forward model in terms of u and y,

$$u(k) = -D^{-1}CX(k) + D^{-1}y(k) \tag{A2}$$

Substituting this into the state evolution in Eq. A1,

$$X(k+1) = (A - BD^{-1}C)X(k) + BD^{-1}y(k) \tag{A3}$$

Equations A2 and A3 define the relationship between the forward and inverse models of the state space representation. Thus, the relationship between the state space matrices A, B, C, and D of any forward model are related to the state space matrices Φ, Γ, C_i, and D_i of the inverse model as given by the relationships

$$\Phi = A - BD^{-1}C; \Gamma = BD^{-1}$$

$$C_i = -D^{-1}C; D_i = D^{-1} \tag{A4}$$

Derivation of Eq. 33

In this case, we need to show that the term $(S_O^I + G_m^I R_O^{iI})$ is an identity matrix.

Substituting for S_O^I and R_O^{iI} from Eqs. 1 and 4, we get

$$S_0^I + G_m^I R_0^{iI} = (I + G^I C^I)^{-1} + G_m^I (I + C^I G^I)^{-1} C^I \tag{A5}$$

which can be simplified using Eq. 3 to

$$S_0^I + G_m^I R_0^{iI} = (I + G^I C^I)^{-1} + G_m^I C^I (I + G^I C^I)^{-1}$$
$$= (I + G^I C^I)^{-1}(I + G_m^I C^I), \tag{A6}$$

which is an identity matrix for $G^I_m = G^I$.

Manuscript received Nov. 23, 2004, and revision received Dec. 14, 2005.