MODIFIED SUBSPACE IDENTIFICATION METHOD FOR BUILDING A LONG-RANGE PREDICTION MODEL FOR INFERENTIAL CONTROL

Yangdong Pan, Jay H. Lee [1]

School of Chemical Engineering, Georgia Institute of Technology, Atlanta, GA

Abstract: In a chemical plant involving a series of processing units, it is beneficial to have a model that can accurately forecast the behavior of downstream variables based on upstream measurements. Such a model can be useful in feedforward and inferential control of the downstream variables to compensate for various upstream disturbances. However, creating such a dynamic model can be very difficult. The conventional multivariable identification approach based on minimizing single-step-ahead prediction error, can result in models leading to poor prediction and control in the described context. To alleviate this difficulty, we propose a modification to the conventional subspace identification method geared towards accurate k-step-ahead prediction, where k is a number chosen according to the estimated dead time. It is shown that the modified subspace identification method can be used in conjunction with the k-step prediction error minimization (PEM). Using an illustrative examples involving six mixing units with a recycle loop, we demonstrate the improvement that is possible from adopting the suggested modification.

1. INTRODUCTION

Most modern plants involve a large number of interconnected processing units, thus raising the need to consider the interactions and information flows among them. A typical plant setup involves measurements and manipulated variables located at the upstream and downstream property variables that need to be controlled. For disturbances occurring in the feed or upstream units, the upstream variables show more immediate responses. Their quick responses, if measured, can be used to manipulate upstream processing conditions in order to keep the downstream properties in control – as in feedforward control or inferential control. To realize this, the upstream measured process variables must be accurately related to the downstream property variables in a dynamic manner. The same situation appears in distributed parameter systems with a large residence time, such as a continuous pulp digester.

Developing a model that accurately captures the dynamic correlation between upstream and downstream variables presents a major challenge. Such models are likely to involve large time delays and dynamics of high order and possibly multiple time scales (due to recycle loops commonly found in industrial plants). Any one of the above features can pose difficulties for the existing system identification approaches. Furthermore, inferential control puts a higher demand on the model accuracy.

In the described problem's context, it is obvious that long-range prediction performance of the model is what ultimately matters. Since a large dead time is involved typically, the short-term predictions, however accurate they may be, are not useful. The importance of emphasizing the long-range prediction over the short-term prediction becomes more clear when one considers the significant model bias typical in most system identification carried out in practice. In the literature, the minimization of k-step-ahead prediction error in the prediction error minimization (PEM) method has been suggested and discussed [8][10] . In addition to the time-domain interpretation, Wahlberg and Ljung [6] formally showed that the

[1] To whom all correspondence should be addressed. phone (404)385-2148, fax (404)894-2866, e-mail:Jay.Lee@che.gatech.edu

use of k-step-ahead prediction methods amounts to emphasizing the accuracy of low-frequency dynamics more in distributing the bias, compared to the conventional one-step-ahead error minimization, which tends to put higher emphasis on the high frequency behavior.

In spite of these developments, understanding of where and how to use the more general k-step PEM in process control's context has been fairly limited. The few exceptions include papers by Shook *et al* [2], and Huang *et al* [3]. Still, a clear link between the method and situations or types of process applications, from which substantial benefits of the method are likely to be realized, is not there. Another reason for the lack of its use in practice is the numerical difficulty associated with using k-PEM for multivariable systems. In addition to the usual complexities (*e.g.*, local minima) associated with the standard PEM, the design of the prefilter necessary to turn the multi-step-ahead prediction error minimization into the one-step-ahead prediction error minimization requires the noise model, which is usually not known *a priori*. In many works, such as the long-range predictive identification (LRPI) approach advocated by Shook *et al* [2], the noise model is assumed to be fixed *a priori*. In this case, the quality of the identified model as well as the performance of the final predictive controller can be strongly influenced by the choice of the noise model.

For multivariable identification problems, the subspace identification method has many attractive features, including the numerical robustness and non-iterative nature of the algorithm [9]. However, the conventional subspace identification method is geared implicitly towards providing accurate one-step-ahead predictions. It is shown in this paper that, for those applications requiring accurate long-range predictions, the conventional method can perform poorly. Given the above-mentioned merits of the subspace method, however, it is useful to consider how the method can be extended to give higher emphasis on the long-range prediction performance.

The contribution of this paper can be two-fold. First, we bring to attention a situation ubiquitous in the process industries, for which the importance of fitting a model to optimize its long-range prediction performance is very high. Second, we present a modified version of subspace identification, in which the emphasis is given to the k-step-ahead prediction performance, where k is a general number chosen according to the process dead-time. We also show how a model obtained from the modified subspace method can be further improved through the k-step-ahead prediction error minimization (k-PEM). An example involving 6 mixing units with a recycle loop is chosen to show the importance of emphasizing the long-range performance through the proposed method.

2. PROPOSED MODIFICATIONS FOR EMPHASIZING THE K-STEP-AHEAD PREDICTION PERFORMANCE

Here we propose a modification to the conventional identification method with the aim of obtaining more accurate k-step-ahead predictions. We first show the modifications for the subspace identification method. After that, we discuss how the resulting model can be improved through the PEM method.

2.1 *Subspace Identification Based on Minimizing the k-Step-Ahead Prediction Error*

The conventional subspace ID approach, such as the N4SID method described in [9], implicitly assumes that the purpose of the model is to provide accurate one-step-ahead prediction. This is seen in the step where state space matrices A, B, C, D are estimated through least squares. In N4SID, data bank for one-step ahead Kalman state estimate $x_{t+1|t}$ is first created from the input/output data based on the following multi-step prediction equation:

$$\begin{bmatrix} y_{t+1} \\ y_{t+2} \\ \vdots \\ y_{t+\bar{n}} \end{bmatrix} = L_1 \begin{bmatrix} y_{t-\bar{n}+1} \\ y_{t-\bar{n}+2} \\ \vdots \\ y_t \end{bmatrix} + L_2 \begin{bmatrix} u_{t-\bar{n}+1} \\ u_{t-\bar{n}+2} \\ \vdots \\ u_t \end{bmatrix} + L_3 \begin{bmatrix} u_{t+1} \\ u_{t+2} \\ \vdots \\ u_{t+\bar{n}-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1|t} \\ \varepsilon_{t+2|t} \\ \vdots \\ \varepsilon_{t+\bar{n}+1|t} \end{bmatrix} \quad (1)$$

Since we can write the optimal predictions in terms of the Kalman state estimate (i.e., the estimate by the nonstationary Kalman Filterinitialized at $t - \bar{n} + 1$ as

$$\begin{bmatrix} y_{t+1|t} \\ y_{t+2|t} \\ \vdots \\ y_{t+\bar{n}|t} \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\bar{n}-1} \end{bmatrix} x_{t+1|t} + L_3 \begin{bmatrix} u_{t+1} \\ u_{t+2} \\ \vdots \\ u_{t+\bar{n}+1} \end{bmatrix}, \quad (2)$$

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\bar{n}-1} \end{bmatrix} x_{t+1|t} = \begin{bmatrix} L_1 & L_2 \end{bmatrix} \begin{bmatrix} y_{t-\bar{n}+1} \\ \vdots \\ y_t \\ u_{t-\bar{n}+1} \\ \vdots \\ u_t \end{bmatrix} \quad (3)$$

Because state coordinates are not fixed *a priori*, one-step ahead state estimate $x_{t+1|t}$ can be created by estimating $[L_1 \quad L_2]$ through least squares

and then finding a set of basis that spans its range space. In N4SID, this is done through a series of oblique matrix projections [9]. Once data for $x_{t+1|t}$ and $x_{t+2|t+1}$ are created, the state space matrices are obtained by solving the linear least squares problem

$$\begin{aligned} x_{t+2|t+1} &= Ax_{t+1|t} + Bu_{t+1} + w_{t+1|t} \\ y_{t+1} &= Cx_{t+1|t} + \varepsilon_{t+1|t} \end{aligned} \tag{4}$$

where the residuals w and ε are minimized. Hence, in this step of the subspace method, one-step-ahead prediction error is minimized. The covariance matrix for w and ε, $\begin{pmatrix} R_w & R_{w,\varepsilon} \\ R_{w,\varepsilon}^T & R_\varepsilon \end{pmatrix}$, is estimated from the residuals of the least squares and the Kalman filter is designed with the calculated system and covariance matrices to obtain the following innovation form of the model.

$$\begin{aligned} x_{t+2|t+1} &= Ax_{t+1|t} + Bu_{t+1} + K\varepsilon_{t+1|t} \\ y_{t+1} &= Cx_{t+1|t} + \varepsilon_{t+1|t} \end{aligned} \tag{5}$$

We may generalize N4SID to emphasize the k-step-ahead prediction in the following manner. To create k-step ahead state estimates, the optimal multi-step prediction equation of (1) can be modified to

$$\begin{aligned} \begin{bmatrix} y_{t+k} \\ y_{t+k+1} \\ \vdots \\ y_{t+k+\bar{n}} \end{bmatrix} &= L_1 \begin{bmatrix} y_{t-\bar{n}+1} \\ y_{t-\bar{n}+2} \\ \vdots \\ y_t \end{bmatrix} + L_2 \begin{bmatrix} u_{t-\bar{n}+1} \\ u_{t-\bar{n}+2} \\ \vdots \\ u_{t+k-1} \end{bmatrix} \\ &\quad + L_3 \begin{bmatrix} u_{t+k} \\ u_{t+k+2} \\ \vdots \\ u_{t+k+\bar{n}-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+k|t} \\ \varepsilon_{t+k+1|t} \\ \vdots \\ \varepsilon_{t+k+\bar{n}|t} \end{bmatrix} \end{aligned} \tag{6}$$

As before, it follows that

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\bar{n}-1} \end{bmatrix} x_{t+k|t} = \begin{bmatrix} L_1 & L_2 \end{bmatrix} \begin{bmatrix} y_{t-\bar{n}+1} \\ \vdots \\ y_t \\ u_{t-\bar{n}+1} \\ \vdots \\ u_{t+k-1} \end{bmatrix} \tag{7}$$

Following the same procedure as before, data bank for k-step-ahead state estimates $x_{t+k|t}$ and $x_{t+k+1|t+1}$ can be obtained. Then, a state space model can be obtained by performing least squares on the following equations:

$$\begin{aligned} x_{t+k+1|t+1} &= Ax_{t+k|t} + Bu_{t+k} + w_{t+k|t} \\ y_{t+k} &= Cx_{t+k|t} + \varepsilon_{t+k|t} \end{aligned} \tag{8}$$

The residual $\varepsilon_{t+k|t}$ represents the k-step-ahead prediction error, which is minimized. Note that, if the data-based Kalman estimates were perfect, then

$$w_{t+k|t} = \underbrace{A^{k-1}K}_{\bar{K}} \varepsilon_{t+1|t} \tag{9}$$

Also,

$$\varepsilon_{t+k|t} = \underbrace{\sum_{i=0}^{k-1} q^{-i} \mathcal{H}_i}_{\bar{F}_k(q)} \varepsilon_{t+k|t+k-1} \tag{10}$$

where $\mathcal{H}_i$ is the i^{th} Markov parameter of the noise model (A, K, C, I).

Based on these, the procedure for extracting $(A, B, C,)$ and K are as follows:

(1) Solve the least squares problem for the output equation to find C that minimizes $y_{t+k} - Cx_{t+k|t}$ in the 2-norm sense. The residuals represent the data for $\varepsilon_{t+k|t}$.
(2) Solve the least squares for the state equation to find A, B. The residual can be viewed as $w_{t+k|t}$.
(3) On the generated residual of $\varepsilon_{t+k|t}$, use a whitening filter to obtain one-step-ahead prediction error $\varepsilon_{t+k|t+k-1}$. A convenient way to do this is to apply subspace identification to the data. The ouput residual from this will be $\varepsilon_{t+1|t}$.
(4) Calculate the covariance matrix for $w(t+k|t)$ and using the whitened residual $\varepsilon(t+1|t)$. According to (9), the covariance matrix for the residual $w_{t+k|t}$ and $\varepsilon_{t+1|t)}$ has the form of

$$\begin{bmatrix} A^{k-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} R_w & R_{w,\varepsilon} \\ R_{w,\varepsilon}^T & R_\varepsilon \end{bmatrix} \begin{bmatrix} A^{k-1} & 0 \\ 0 & I \end{bmatrix}^T \tag{11}$$

where $\begin{pmatrix} R_w & R_{w,\varepsilon} \\ R_{w,\varepsilon}^T & R_\varepsilon \end{pmatrix}$ represents the covariance for $w_{t+1|t}$ and $\varepsilon_{t+1|t}$. With the calculated system matrices and the extracted covariance matrix for w and ε, one can proceed to design the Kalman filter to put the model in the innovation form. The k-step ahead predictor can be easily derived from it.

It should be obvious to those familiar with the subspace identification method that the asymptotic properties of N4SID such as unbiasedness and consistency remain intact with the above modifications.

2.2 *k-Step Prediction Error Minimization*

Although the modified subspace ID method puts higher emphasis on the accuracy of the k-step ahead prediction in obtaining state space matrices, it does not directly minimize k-step-ahead

prediction error for a finite data set. It has been suggested in Ljung [1] that the subspace method be used to initialize PEM, which generally requires a special parameterization and a good initial guess to be successful. Here we propose to use the model from the proposed k-step subspace ID method to start the k-step PEM.

A MIMO state space model,

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + Ke_t \\ y_t &= Cx_t + Du_t + e_t \end{aligned} \tag{12}$$

can be represented in the following input/output form:

$$y_t = G(q)u_t + H(q)e_t \tag{13}$$

where

$$\begin{aligned} G &= C(qI - A)^{-1}B + D \\ H &= C(qI - A)^{-1}K + I \end{aligned} \tag{14}$$

The optimal one-step ahead predictor is given by Ljung [1]

$$\hat{y}_{t|t-1} = H^{-1}Gu_t + (1 - H^{-1})y_t \tag{15}$$

If parameterized models G_θ and H_θ are used, then the optimal one-step-ahead predictor can be written as

$$\hat{y}_{t|t-1} = H_\theta^{-1}G_\theta u_t + (1 - H_\theta^{-1})y_t \tag{16}$$

Optimal k-step-ahead predictor is

$$\hat{y}_{t|t-k} = W_k G_\theta u_t + (1 - W_k)y_t \tag{17}$$

where

$$W_k = F_k H_\theta^{-1} \tag{18}$$

and

$$F_k = \sum_{i=0}^{k-1} \mathcal{H}_i q^{-i} \tag{19}$$

Here, $\mathcal{H}_i$ is a $n_y \times n_y$ matrix representing the i^{th} impulse response coefficient matrix of $H(q)$. The optimal k-step ahead predictor can also be viewed as the optimal one-step ahead predictor associated with the model

$$y_t = Gu_t + HF_k^{-1}\varepsilon_t \tag{20}$$

where ε_t is a white noise.

For a SISO system, F_k, if known, can be regarded as a prefilter and the k-step prediction error minimization is the same as the one-step prediction error minimization with the filtered I/O data. However, for a MIMO system, because matrices do not commute in multiplication, prefiltering the data before applying the one-step ahead PEM does not work. Therefore, F_k has to be embedded into the model structure when applying the PEM, resulting in a structured identification problem. Let us use the state-space representation of

$$\begin{aligned} F_k &= (\tilde{A}_F, \tilde{B}_G, \tilde{C}_G, \tilde{D}_G) \\ G &= (A, B, C, D) \\ H &= (A, K, C, I) \end{aligned} \tag{21}$$

First, the inverse system F_k^{-1} is,

$$\begin{aligned} F_k^{-1} = (&\tilde{A}_F - \tilde{B}_F\tilde{D}_F^{-1}\tilde{C}_F, \tilde{B}_F\tilde{D}_F^{-1}, \\ &-\tilde{D}_F^{-1}\tilde{C}_F, \tilde{D}_F^{-1}) \end{aligned} \tag{22}$$

Let us denote

$$F_k^{-1} = (\tilde{A}_{F^{-1}}, \tilde{B}_{F^{-1}}, \tilde{C}_{F^{-1}}, \tilde{D}_{F^{-1}}) \tag{23}$$

where $\tilde{D}_{F^{-1}} = I$. Then, the combined model structure HF_k^{-1} is,

$$\begin{aligned} &HF_k^{-1} = \\ &\left(\begin{bmatrix} A & K\tilde{C}_{F^{-1}} \\ 0 & \tilde{A}_{F^{-1}} \end{bmatrix}, \begin{bmatrix} K\tilde{D}_{F^{-1}} \\ \tilde{B}_{F^{-1}} \end{bmatrix}, \begin{bmatrix} C & \tilde{C}_{F^{-1}} \end{bmatrix}, \tilde{D}_{F^{-1}} \right) \end{aligned} \tag{24}$$

Now, the final combined model structure of both G and H is adopted as,

$$\begin{aligned} [\, G \;\; HF_k^{-1} \,] = \Bigg(&\begin{bmatrix} A & K\tilde{C}_{F^{-1}} \\ 0 & \tilde{A}_{F^{-1}} \end{bmatrix}, \\ &\begin{bmatrix} B & K \\ 0 & \tilde{B}_{F^{-1}} \end{bmatrix}, \begin{bmatrix} C & \tilde{C}_{F^{-1}} \end{bmatrix}, \begin{bmatrix} D & I \end{bmatrix} \Bigg) \end{aligned} \tag{25}$$

To solve this structured system identification problem, a grey box identification method, for example 'idgrey' in Matlab, can be used.

The overall iterative procedure can be described as follows.

(1) Use the proposed k-step-ahead subspace identification method to obtain the initial state space model (A, B, C, D, K).
(2) Obtain F_k from the noise model $H = C(qI - A)^{-1}K + I$.
(3) Apply the structured identification approach to minimize the prediction error for (25) in order to obtain new (A, B, C, D, K).
(4) Obtain a new prefilter F_k from the new noise model H.
(5) Go back to step 3. Continue until the model converges.

3. CASE STUDY

3.1 *CST Tanks in Series with A Recycle Loop*

The example chosen for illustrative purposes involves a 6 CST mixers and 1 plug flow pipe connected in series, as shown in Fig 1. In addition,

there is a recycle flow, from mixer 6 back to mixer 1. The flowrate of the secondary inlet, represented by F_u, is assumed to be the manipulated input. The concentration of the main inlet flow C_{Ad}, is treated as an unknown disturbance variable. Outputs are C_{A1} and C_{A6}. The steady state condition is $F_u = 20$, $F_d = 100$, $F_r = 200$, $C_{Ad} = 2$, and $C_u = 20$. The volume of each mixer is 1000. The dynamics of the plug flow pipe between mixer 3 and mixer 4 are represented as a pure delay of 10 time units. We assume that C_{A1} is measured and we are interested in using this measurement to inferentially control the downstream concentration C_{A6}.

First identification data are generated by performing simulations with random input variables. Both the manipulated input and the unknown disturbance variable are drawn from uniform distributions with standard deviations of 15 and 0.5 respectively and switching probability of 0.2. 50 data sets are generated for identification, each with 4000 data points. First, to test the quality of the deterministic part of the identified models, two data sets are generated with manipulated input movement only, one with a step input change and the other with random input changes. Next, to test the model-based inferential prediction and control performance, additional 1000 data points are generated with the same type of input and disturbance variations as those used to generate the 50 modeling data sets.

3.2 *Simulation Results*

The conventional subspace ID method (N4SID) and the modified subspace ID method (k-N4SID) are applied to each of the 50 data sets, which resulted in 50 pairs of state-space models. For the both identification approaches, models with 8 states are identified, and k is chosen to be 50 in applying the k-N4SID algorithm.

The resulting 50 pairs of models are first tested on the two data sets with MV movement only. After that, the identified models are tested for their final purpose, inferential prediction and control. These are done with the validation data set involving the stochastic disturbances. The control objective is to regulate the concentration of the last mixer at the steady-state value. For this, model predictive controllers are designed based on the identified models. The controllers decide the adjustments in the MV based on the inferentially predicted values of the concentration of the last mixer. For every MPC controller, the prediction horizon is chosen to be 200 time units and the control horizon is chosen to be 10 time units. Also, the input and output weighting parameters are chosen to be 10^{-7} and 1, respectively.

Table 1. Comparison of inferential prediction performances of the k-N4SID and N4SID models obtained from the 50 modeling data sets

	1-step inference			k-step inference		
	mean	min	max	mean	min	max
N4SID	0.6259	0.1732	3.8915	0.7835	0.1593	6.6758
k-N4SID	0.4188	0.1812	0.9344	0.4183	0.1457	0.9691

The benefits of the proposed modification to the subspace identification method are clearly seen in the statistical comparison involving the 50 pairs models obtained with N4SID and k-N4SID. First, N4SID resulted in more unstable models, 28 compared to 23 by the k-N4SID. Unstable models for a stable system do not necessarily lead to bad prediction and control performance as long as a stable predictor is formed. However, depending on the location of the unstable eigenvalues, extremely poor prediction and control performance can result, even though the predictor may be stable. It was observed that none of the unstable models obtained by k-N4SID resulted in bad inferential prediction and control, whereas many unstable models obtained by N4SID led to very poor inferential prediction and control results, implying the unstable modes for the N4SID models were much faster growing than those found in the k-N4SID models. Table 1 shows the better performance by the k-N4SID models over the N4SID models, in terms of both one-step inferential prediction and k-step inferential prediction. The subsequent inferential control tests also confirmed the superior quality of the models by k-N4SID over those by N4SID.

To further scrutinize the differences, the identified models were grouped in four categories according to whether both or one of the N4SID and k-N4SID methods resulted in an unstable model. For all four categories, models obtained by k-N4SID method showed better overall inferential prediction and control performance than the corresponding models by the N4SID method. This includes the cases, where N4SID gave a stable model but k-N4SID gave an unstable model. Due to space limit, only the result from the first category, for which the data sets resulted in stable models with k-N4SID but unstable models with N4SID, is shown here. The unstable nature of the models from conventional N4SID can clearly be seen from Figure 2, which shows for one of the data sets the *open-loop* predictions of the two models for a step change in the MV. Figures 3 and 4 display the corresponding differences in the inferential prediction and control performances. We can see that significant improvements in inferential prediction and control performances could be achieved by using k-N4SID instead of N4SID.

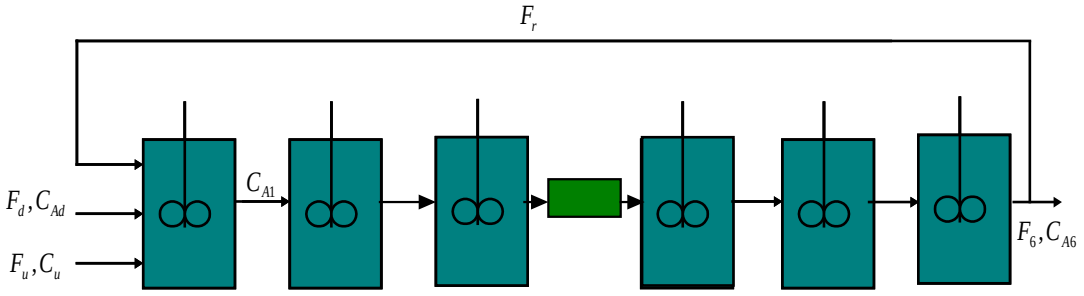

Fig. 1. The schematic for the example of 6 CST mixers in series with a recycle loop

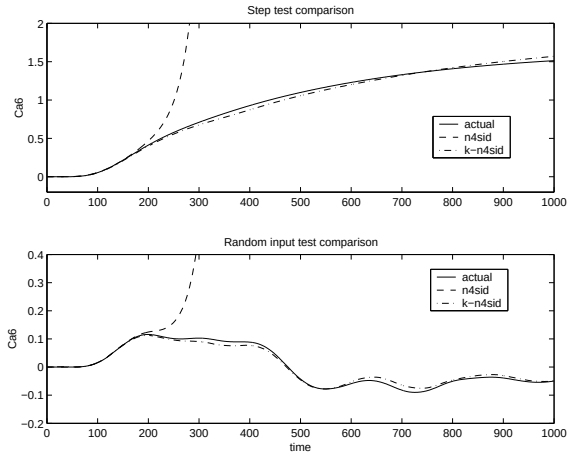

Fig. 2. Deterministic model comparison based on the MV movement data for case 1

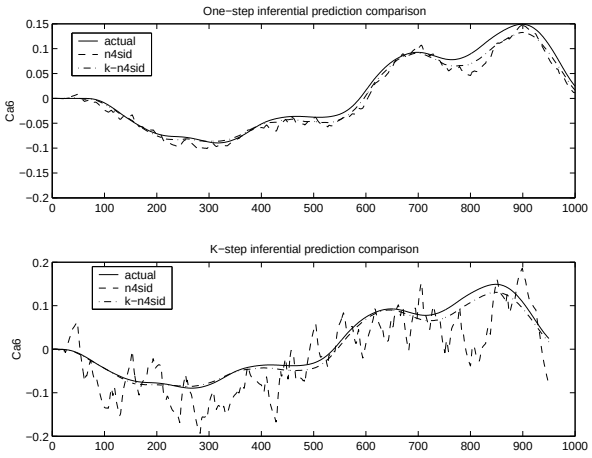

Fig. 3. Inferential prediction performance comparison for case 1

Acknowledgement: The Authors gratefully acknowledge the financial support from the National Science Foundation (CTS-#0096326).

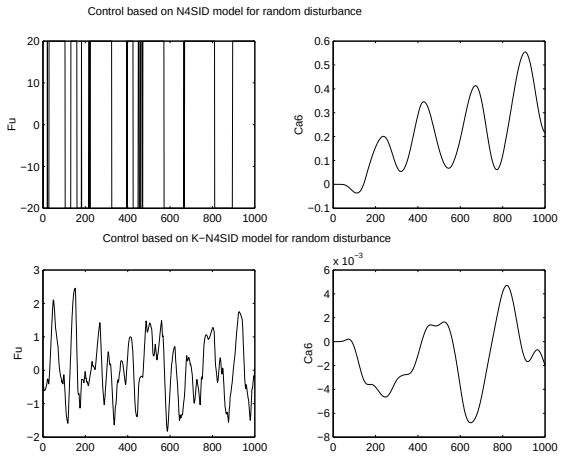

Fig. 4. Inferential control performance comparison for case 1

4. REFERENCES

[1] Ljung, L., *System Identification: Theory For the User*, Prentice-Hall, 2nd edition, 1999.

[2] Shook, D., Mohtadi, S., C., and Shah, S. L., Identification for long-range predictive control, *IEE-Proceedings-*, 138, 75-84, 1991.

[3] Huang, B., Malhotra, A., and Tamayo, E.C., Model predictive control relevant identification and validation, submitted to *Chemical Engineering and Science*, 2001.

[4] Ljung, L., Estimation focus in system identification: prefiltering, noise models, and prediction, *Proceedings of the 38th Conference on Decision and Control*, Phoenix, Arizona, USA, 2810-2815, 1999.

[5] Chaplais, F., and Alaou, K., Two time scaled parameter identification by coordination of local identifiers, *Automatica*, Vol. 32, No. 9, 1303-1309, 1996.

[6] Wahlberg, B., and Ljung, L., Design variables for bias distribution in transfer function estimation, *IEEE Trans. Automatic Control*, AC-31, 134-144, 1986.

[7] Luse, D. W., and Khalil, H., Frequency domain results for systems with slow and fast dynamics,*IEEE Trans. Automatic Control*, AC-30, 1171-1179, 1985.

[8] Van Zee, G. A., and Bosgra, O. H., Validation of prediction error identification results by multivariable control implementation, in *Proc. 6th IFAC Symp. Ident. Syst. Parameter Estimation*, Washington, DC, 560-565, 1982.

[9] Van Overschee, P. and DeMoor, B., *Subspace Identification for Linear Systems: Theory, Implementation, Applications.*, Kluwer Academic, Dordrecht, The Netherlands, 1996.

[10] Åström, K. J., Maximum likelihood and prediction error methods, *Automatica*, Vol 16, 551-574, 1980.