FIR Model Identification: Parsimony Through Kernel Compression with Wavelets

Michael Nikolaou and Premkiran Vuthandam
Dept. of Chemical Engineering, Texas A&M University, College Station, TX 77843

Although finite impulse response (FIR) models are nonparsimonious, they are frequently used in model predictive control systems because they can fit arbitrarily complex stable linear dynamics. However, identification of FIR models from experimental data may result in data overfitting and high modeling uncertainty. To overcome this, FIR models may be determined by regularization-based least squares and indirectly after prior identification of other parametric models such as ARX. In both cases, some prior knowledge about the model is essentially assumed to be known. ARX models, although parsimonious in terms of the number of identified parameters, perform poorly for bad choices of model structure and order. A methodology for the direct identification of parsimonious FIR models is proposed. In this way, most advantages of the FIR structure are retained without its disadvantages. The idea relies on the effective use of some prior information about the model through wavelet-based signal compression. The proposed methodology is compared with other FIR identification methodologies, on the basis of the closeness of the identified FIR to the true FIR, steady-state gain estimation, and analysis of the prediction residuals on a cross-validation set of fresh data. Simulation studies on a single-input-single-output process show that this method performs very well in all tests considered. Certain industrial practices are shown to be special cases of the proposed formalism.

Introduction

At the heart of any model predictive control (MPC) scheme lies a process model. Popular model structures used in applied MPC schemes are finite impulse-response (FIR) models, or step-response (SR) models. FIR and SR models can be estimated directly from process input–output data. The main reasons for the popularity of FIR models are:

- They can fit any complex dynamic system.
- No model structure needs to be selected, provided a sufficiently long model kernel is chosen.

However, FIR and SR models are nonparsimonious, requiring a large number of parameters (typically, 30–100) to be identified. Therefore, large amounts of data are needed for the identification of FIR or SR model parameters with small error margins and without data-overfitting problems.

The most straightforward method for the identification of FIR models is the ordinary least-squares (OLS) method, relying on minimization of some sort of square error between measurement and model prediction. To avoid overfitting with OLS, one can use the method of regularization [ridge regression (RR)], where a penalty on either the size or the change of the model parameters (FIR model kernel) is imposed (Wise and Ricker, 1992; MacGregor et al., 1991). Kozub (personal communication, 1994) proposed a refinement of the regularization method, by suggesting weighting matrices in the least-squares objective function that penalize changes in FIR coefficients toward the tail end of the FIR coefficient sequence. Ricker (1988) studied the use of partial least squares (PLS) and singular value decomposition (SVD) for estimating FIR models. MacGregor et al. (1991) pointed out that the SVD method applied by Ricker (1988) was the same as principal-component regression (PCR). MacGregor et al. (1991) and Dayal and MacGregor (1996) provide extensive discussions of the preceding methods.

To overcome the problems associated with the nonparsimony of FIR models, one can use parsimonious models of

Correspondence concerning this article should be addressed to M. Nikolaou.
Current addresses of: M. Nikolaou, Chemical Engineering Dept., University of Houston, TX 77204; P. Vuthandam, Aspen Tech, 9896 Bissonet, Houston, TX 77036.

low order, such as transfer-function models or ARX models (Box and Jenkins, 1976; Ljung, 1987; Söderström and Stoica, 1989). FIR models can then be obtained from the ARX models through mathematical manipulations. Although parsimonious ARX models can be identified more accurately than mathematically equivalent FIR models, the choice of appropriate model order and structure for ARX models becomes a crucial issue. In essence, some prior knowledge about the model must be used for a reasonable model structure to be selected. However, prior knowledge about a model could also be used for the development of *parsimonious* FIR models. The main thesis of this article is that parsimonious FIR model structures can indeed be constructed and corresponding models can be identified from experimental data, provided certain prior knowledge about the model is available. For most chemical processes that knowledge is often available and, thus, can be easily used.

The main idea of the article can be explained through Figure 1. While virtually all coefficients of the FIR model kernel $\{h_j\}_{j=1,2\ldots}$ need to be identified for small values of j, only a few coefficients need to be identified for large j. Values of h_j not explicitly identified can be constructed through appropriate interpolation. One question, then, is how to select which terms of the sequence $\{h_j\}$ to identify explicitly and how to construct terms of $\{h_j\}$ not explicitly identified. In fact, as we show in the next section, the true problem in FIR model identification is that of estimating the kernel $\{h_j\}_{j=1,2\ldots}$ through reconstruction of a corresponding *continuous function* from *sampled* values of that function. Clearly, techniques more sophisticated than uniform sampling can be used for the sampling of that continuous function. We propose a solution based on the discrete wavelet transform (DWT). Using the compression capabilities of wavelets, we present a methodology to identify parsimonious FIR models. We show that certain simple methodologies used in industrial practice generate wavelet-compressed FIR models using specific wavelets (such as Haar's) and we propose nontrivial improvements on these methodologies.

The main advantage of the discrete wavelet transform (DWT) of a sequence is that it is localized in both frequency and time. By truncating *small* DWT coefficients, the original sequence may be sparsely represented in terms of a smaller number of coefficients. From prior partial knowledge of the model and the nature of the FIR coefficients, we can truncate a large number of wavelet coefficients *a priori*, and thus require fewer coefficients to be identified. This is the basic idea upon which we base our methodology of identifying FIR coefficients parsimoniously.

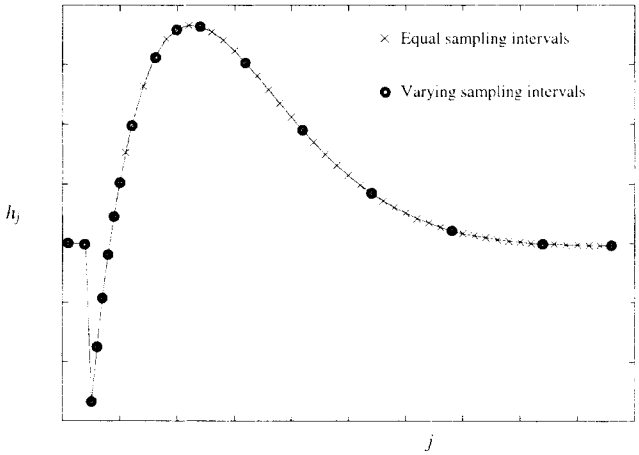

Figure 1. Typical impulse response with dead time and inverse response, with equal sampling intervals (×) and nonequal sampling intervals (○).

Methodologies using wavelets for identification have been proposed in the literature in various contexts. Tsatsanis and Giannakis (1993) propose an algorithm for the identification of time-varying autoregressive (AR) and autoregressive moving-average (ARMA) models using wavelets. Pati et al. (1993) examine their method for model-order reductions of linear stable systems using wavelet approximations. Sureshbabu and Farrell (1995) study a wavelet-based system-identification method for nonlinear systems. A method for the denoising of input–output identification data using wavelet-based prefiltering was presented by Palavajjhala et al. (1996). Wavelets were applied by Tsatsanis and Giannakis (1994) in detecting transient plant disturbances and jumps.

The rest of the article is organized as follows. First, we discuss the FIR model structure for discrete-time systems, and explain how a continuous-time function underlies that structure. Next, we briefly present some of the prevailing identification schemes for the identification of FIR models, namely OLS, regularization, and ARX identification (along the lines presented in Dayal and MacGregor, 1996). Following this, we present the DWT, explain how it can be used in the development of parsimonious FIR model structures, and present the resulting identification methodology. Next, we illustrate the proposed method through an example, where we consider an SISO process with dead time and inverse response, and compare our methodology to other FIR identification methodologies. Finally, we summarize our results and present directions for future research.

FIR Model Identification

Continuous- and discrete-time FIR models

Consider a stable, causal, *continuous-time* SISO process, modeled by the following convolution model:

$$y(t) = (g * u)(t) \triangleq \int_{-\infty}^{t} g(t-r)u(r)\,dr, \qquad (1)$$

where $*$ is the convolution operator; $y(t)$ is the output at time t; $u(t)$ is the input at time t; $g(t)$ is the model kernel, equal to the impulse response of the system. For a process uniformly sampled every T time units, with $u(r) = u(t_{j-1})$ constant for $r \in [t_{j-1}, t_j)$, as is the case with digital-controlled processes, the output at time $t_k = kT$ can be written as

$$y(t_k) = \sum_{j=1}^{\infty} h_j u(t_{k-j}) \approx \sum_{j=1}^{n} h_j u(t_{k-j}), \qquad (2)$$

where h_j, defined as

$$h_j \triangleq \int_{t_j}^{t_{j+1}} g(\tau)\,d\tau = \int_{jT}^{(j+1)T} g(\tau)\,d\tau, \qquad j = 1,2,\ldots \qquad (3)$$

are the coefficients of the causal *discrete-time* convolution model corresponding to the continuous-time model of Eq. 1. At this point one might conclude that all coefficients h_j need to be identified, in order for the FIR model of Eq. 2 to be identified. This is not necessarily true. The identification problem consists of identifying the corresponding coefficients c_i, which parametrize the FIR kernels h_j as

$$\boldsymbol{h} \triangleq \begin{bmatrix} h_1 \\ \vdots \\ \\ \vdots \\ h_n \end{bmatrix} = \begin{bmatrix} w_{11} & \cdots & w_{1m} \\ \vdots & \ddots & \vdots \\ \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nm} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} = \sum_{i=1}^{m} c_i \boldsymbol{w}_i. \quad (4)$$

The advantage of this transformation is that the basis vectors $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$ can be virtually always chosen in a way such that $m \ll n$. Indeed, while the value of n is determined by the continuous-time system sampling period T (which, in turn, should be small enough to prevent aliasing), the value of m is not. The preceding discussion naturally leads to the problem of how to select the vectors $\{\boldsymbol{w}_i \triangleq [w_{1i}, \cdots w_{ni}]^T, i = 1, \ldots, m\}$. Certainly, there are several alternatives. In the sequel we propose an approach relying on the DWT and explain its advantages.

FIR model identification

For notational simplicity, in the sequel we will assume that $x(kT) \equiv x(k)$ for any variable x. Assume the following FIR model structure for a stable system:

$$y(k) = \sum_{j=1}^{n} h_j u(k-j) + e(k) = \boldsymbol{h}^T \boldsymbol{u}(k) + e(k), \quad (5)$$

where $e(k)$ is output additive noise that is identically, independently distributed (i.i.d.) with 0 mean and variance σ^2, and

$$\boldsymbol{u}(k) = [u(k-1) \quad u(k-2) \cdots u(k-n)]^T \quad (6)$$

is the vector of lagged inputs.

FIR Identification Using OLS. With N observations, input–output data can be written in matrix form as

$$\boldsymbol{y} = \boldsymbol{\Phi h} + \boldsymbol{e} \quad (7)$$

where

$$\boldsymbol{y} \triangleq [y(k-N+1) \quad y(k-N+2) \cdots y(k)]^T \quad (8)$$

$$\boldsymbol{\Phi} \triangleq [\boldsymbol{u}(k-N+1) \quad \boldsymbol{u}(k-N+2) \cdots \boldsymbol{u}(k)]^T. \quad (9)$$

$$\boldsymbol{e} \triangleq [e(k-N+1) \quad e(k-N+2) \cdots e(k)]^T. \quad (10)$$

The OLS solution to the preceding problem is

$$\hat{\boldsymbol{h}} = (\boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^T \boldsymbol{y} \triangleq \boldsymbol{R}^{-1} \boldsymbol{\Phi}^T y, \quad (11)$$

where $\boldsymbol{R} \triangleq \boldsymbol{\Phi}^T \boldsymbol{\Phi}$, with

$$E\left[(\boldsymbol{h} - \hat{\boldsymbol{h}})(\boldsymbol{h} - \hat{\boldsymbol{h}})^T\right] = \sigma^2 \boldsymbol{R}^{-1}. \quad (12)$$

Regularization. In regularization, the objective function of the parameter estimation scheme imposes a penalty on a function of $\boldsymbol{h}$. For example, $\boldsymbol{h}$ can be estimated from

$$\min_{\boldsymbol{h}} (\boldsymbol{y} - \boldsymbol{\Phi h})^T (\boldsymbol{y} - \boldsymbol{\Phi h}) + \alpha \boldsymbol{H}^T \boldsymbol{Q h}, \quad (13)$$

where $\alpha \geq 0$ and $\boldsymbol{Q}$ is a positive definite matrix. The solution for this minimization is

$$\hat{\boldsymbol{h}} = [\boldsymbol{R} + \alpha \boldsymbol{Q}]^{-1} \boldsymbol{\Phi}^T \boldsymbol{y}. \quad (14)$$

Several options for $\boldsymbol{Q}$ are available (Dayal and MacGregor, 1996). Kozub (personal communication, 1994) suggested the choice

$$\boldsymbol{Q} = \boldsymbol{A}^T \boldsymbol{L} \boldsymbol{A}, \quad (15)$$

where $\boldsymbol{A}$ and $\boldsymbol{L}$ are the following $n \times n$ matrices

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}, \quad \text{and}$$

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 2 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & n-1 & 0 \\ 0 & 0 & \cdots & 0 & n \end{bmatrix}. \quad (16)$$

This choice for $\boldsymbol{Q}$ places an increasing penalty on the FIR coefficient differences $(h_i - h_{i-1})$ as i increases from 2 to n.

FIR models through ARX model identification

To determine the FIR coefficients, one could first identify a structured model, like an autoregressive with exogenous inputs (ARX) model, from which one could subsequently determine the FIR coefficients. The ARX model structure is of the following form:

$$y(k) = -\sum_{j=1}^{n_a} a_j y(k-j) + \sum_{j=1}^{n_b} b_j u(k-j) + e(k). \quad (17)$$

ARX models can be identified using standard linear least-squares techniques. If the noise e is not white, then a noise model can be added to the preceding model.

Discrete Wavelet Transform

The DWT of a (finite- or infinite-dimensional) vector is the result of a linear transformation that generates a new vector

of dimension equal to that of the original vector. Figure 2 shows an example of how Mallat's multiresolution analysis algorithm (Mallat, 1989) transforms a vector $f=[f_1 \cdots f_8]^T$ in $\Re^8$ to its DWT $\tilde{f}=[\tilde{f}_1 \cdots \tilde{f}_8]^T$ in $\Re^8$. At each step, the weighted averages and differences of corresponding vector segments are computed. The decomposition of a signal shown in Figure 2 can also be seen as passing the signal through a filter bank (Strang and Nguyen, 1996). At each resolution level i ($1 \le i \le 3$) the signal is passed through a low-pass ("averaging") filter C, and a high-pass ("differencing") filter D. C and D are digital filters with kernel coefficients c_i and d_i, respectively, according to the kind of wavelet chosen. The net effect of the calculations in Figure 2 is multiplication of the original vector by an invertible matrix W, or

$$\tilde{f}=\begin{bmatrix} f^{(0)} \\ g^{(0)} \\ g^{(1)} \\ \vdots \\ g^{(n_r-1)} \end{bmatrix} = (W_1 W_2 \cdots W_{n_r-1} W_{nr}) f = Wf, \qquad (18)$$

where the matrices W_i ($i=1, \ldots, n_r$) are defined as follows:

$$W_i=\left[\begin{array}{c|c} H_{0,i} & \\ \cdots\cdots & \mathbf{0} \\ H_{1,i} & \\ \hline \mathbf{0} & I_{n-n_i} \end{array}\right] \qquad (19)$$

and $n_r \ge 1$ is the finest resolution level (original signal level) with the coarsest resolution level being 0. However, as Figure 2 shows, Mallat's algorithm carries out the matrix-vector multiplication of Eq. 18 in an extremely efficient manner [$O(n)$ computations].

The DWT of a signal depends on the particular kind of wavelet (filters C and D) chosen. In this study, we use a quadratic spline (QS) wavelet basis, because it results in smooth approximations of the original function. The coefficients for the filters C and D corresponding to the QS wavelet are (Strang and Nguyen, 1996)

$$(c_0 \quad c_1 \quad c_2 \quad c_3)=\frac{1}{4}(-1 \quad 3 \quad 3 \quad -1) \qquad (20)$$

$$(d_0 \quad d_1 \quad d_2 \quad d_3)=\frac{1}{4}(-1 \quad 3 \quad -3 \quad 1), \qquad (21)$$

resulting in

$$H_{0,i}=\begin{bmatrix} c_0+c_1 & c_2 & c_3 & & & & & & \\ & c_0 & c_1 & c_2 & c_3 & & & & \\ & & & c_0 & c_1 & c_2 & c_3 & & \\ & & & & & \ddots & \ddots & \ddots & \\ & & & & & & c_0 & c_1 & c_2+c_3 \end{bmatrix} \qquad (22)$$

and

$$H_{1,i}=\begin{bmatrix} d_0+d_1 & d_2 & d_3 & & & & & & \\ & d_0 & d_1 & d_2 & d_3 & & & & \\ & & & d_0 & d_1 & d_2 & d_3 & & \\ & & & & & \ddots & \ddots & \ddots & \\ & & & & & & d_0 & d_1 & d_2+d_3 \end{bmatrix} \qquad (23)$$

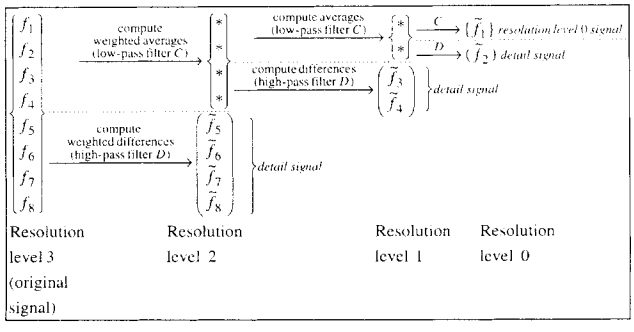

Figure 2. Example of discrete wavelet transform.

For comparison purposes, we also use the Haar wavelet, with

$$(c_0 \quad c_1)=\frac{1}{\sqrt{2}}(1 \quad 1) \qquad (24)$$

$$(d_0 \quad d_1)=\frac{1}{\sqrt{2}}(1 \quad -1). \qquad (25)$$

The Haar wavelet is simple, but has poor filtering characteristics.

From a practical viewpoint, we use the QS wavelet transform to resolve signals of length

$$n = 3 \times 2^{n_r} \geq 6 \tag{26}$$

and the Haar wavelet for signals of length

$$n = 2^{n_r} \geq 2. \tag{27}$$

Finally, the inverse DWT (IDWT) can be easily found. Given the DWT $\tilde{f}$, the IDWT can be found as

$$f = W^{-1}\tilde{f}. \tag{28}$$

For several classes of wavelets Mallat's algorithm can be used to perform the preceding matrix/vector multiplication in $O(n)$ multiplications, because the matrix W^{-1} usually has the same sparse structure as W. For orthonormal wavelets, W is also orthonormal, with $W^{-1} = W^T$.

FIR Model Compression Using the DWT

First we explain how wavelets can be used to compress a given FIR model. The discussion then leads to a methodology on how to identify in compressed form an FIR model that is initially only partially known.

Compression of FIR models

FIR model compression (approximation) using wavelets first involves determining the DWT coefficients of the FIR model $\{h_i\}$ using Eq. 18 for a certain wavelet. Of the obtained DWT coefficients, only coefficients with magnitude above a certain threshold value are significant and, thus, retained, with the rest being neglected (set to zero). Good compression can be attained if a large number of the DWT coefficients are neglected. The characteristics of the particular wavelet used (e.g., smoothness) are also important.

The features of the finite impulse response of typical stable process dynamics are dead times, inverse responses, a rise in the process output, and finally a decay. This can be observed in Figure 1, which shows all of these features in an FIR process model. The discrete-time FIR model coefficients are equidistantly sampled points of the continuous-time impulse response of the process. The sampling interval T is chosen according to Shannon's sampling theorem (Benedetto, 1992), to prevent aliasing effects. This choice of T would be capable of capturing virtually all *fast* or *abrupt* changes (high frequency) in the impulse response. From Figure 1 it can be observed that for typical impulse responses, all the *fast* and *abrupt* changes occur during the early part of the impulse response. However, the tail of the impulse response changes (decays) much more slowly, which suggests that we could sample that function at a slower rate there. (It should be stressed that our method does not correspond to variable sampling rate for input and output signals. Incidentally, if a pulse was used as the input signal for noise-free identification, then Figure 1 would show both the response of the process and the FIR kernel, and the variable sampling rate for the process output would make sense. Realistically, however, a pulse is not a good input for identification (it results in poor S/N ratio because it is not persistently exciting of any order (Astrom and Wittenmark, 1994)).) Therefore, a methodology is desired that samples the impulse response at different rates, during different regimes of the response. By doing so, one would need to know fewer parameters by which to characterize the impulse response. This could be achieved in many ways, including wavelet compression techniques. The idea is as follows.

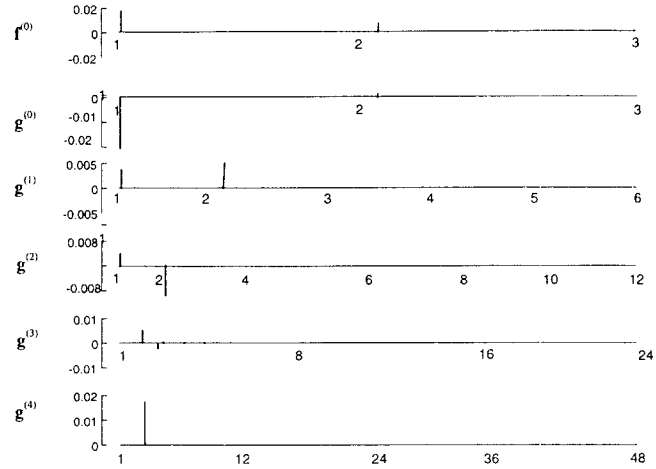

Figure 3. Coarsest resolution and detail signals for FIR in Figure 1.

Consider the DWT

$$\tilde{h} = Wh \tag{29}$$

of the FIR kernel vector h. The DWT coefficients are contained in the vector $\tilde{h}$ shown in Figure 3, and the numbers are given in Table 1. At each resolution level there are only a few significant coefficients, with the rest being negligible. The percentage of insignificant coefficients increases at finer resolution levels. Thus, by retaining only the significant DWT coefficients, and neglecting all the others (setting them to zero) we can achieve compression of the FIR kernel. To explain this procedure of compression, consider a vector of the retained coefficients, denoted by $\tilde{h}_c$, as follows:

$$\tilde{h}_c = P^T\tilde{h}, \tag{30}$$

where $\tilde{h}_c$ is a vector of length n_c; n_c is the number of retained coefficients after compression; and P is a projection matrix of dimensions $n \times n_c$. The projection matrix P is constructed with its columns consisting of unit vectors $v_i \in \Re^n$, whose ith entry is 1 and all others 0. The indices i are chosen

Table 1. Wavelet Coefficients at Various Levels of Resolution*

Resolution Level	Wavelet Coefficients
0 ($f^{(0)}$)	**0.0186, 0.0088, −0.0009**
0 ($g^{(0)}$)	**−0.0257, 0.0005, −0.0006**
1 ($g^{(1)}$)	**0.0037, 0.0047**, 0.0005, 0, −0.0001,0
2 ($g^{(2)}$)	**0.0034, −0.0075**, 0.0005, 0.0002, 0.0001, 0,...,0
3 ($g^{(3)}$)	0, **0.0064, −0.0028**, 0.0001, 0.0001, 0, ..., 0
4 ($g^{(4)}$)	0, 0, 0, **0.0118**, 0, ..., 0

*Significant coefficients are in boldface.

so that $i \in I_c$, where I_c denotes the set of indices of the retained DWT coefficients. The compressed FIR model is then reconstructed using the IDWT as follows

$$\boldsymbol{h} \approx \boldsymbol{h}_c = \boldsymbol{W}^{-1}\boldsymbol{P}\tilde{\boldsymbol{h}}_c = \boldsymbol{X}\tilde{\boldsymbol{h}}_c, \tag{31}$$

where $\boldsymbol{X} \triangleq \boldsymbol{W}^{-1}\boldsymbol{P}$ is a matrix of dimensions $n \times n_c$.

FIR model identification using wavelet compression

In identifying FIR models using wavelet compression, we use the standard OLS approach.

Consider the linear FIR model in Eq. 5. This equation can be written as

$$\begin{aligned} y(k) &= \boldsymbol{h}^T\boldsymbol{u}(k) + e(k) = (\boldsymbol{W}\boldsymbol{h})^T[\boldsymbol{W}^{-T}\boldsymbol{u}(k)] + e(k) \\ &= \tilde{\boldsymbol{h}}^T\boldsymbol{W}^{-T}\boldsymbol{u}(k) + e(k). \end{aligned} \tag{32}$$

By compressing the DWT coefficients as described earlier, we can approximate $y(k)$ as

$$y(k) \approx \tilde{\boldsymbol{h}}_c^T\boldsymbol{P}\boldsymbol{W}^{-T}\boldsymbol{u}(k) + e(k) = \tilde{\boldsymbol{h}}_c^T\boldsymbol{X}^T\boldsymbol{u}(k) + \tilde{\boldsymbol{h}}_c^T\breve{\boldsymbol{u}}(k) + e(k), \tag{33}$$

where $\boldsymbol{u}(k) \triangleq \boldsymbol{X}^T\boldsymbol{u}(k)$.

Remark. If the wavelet matrix $\boldsymbol{W}$ is orthonormal, then $\boldsymbol{W}^{-T} = \boldsymbol{W}$, and $\breve{\boldsymbol{u}}(k) = \boldsymbol{P}^T\tilde{\boldsymbol{u}}(k) = \tilde{\boldsymbol{u}}_c(k)$, where $\tilde{\boldsymbol{u}}(k) = \boldsymbol{W}\boldsymbol{u}(k)$ is the DWT of $\boldsymbol{u}(k)$.

We then identify the retained coefficients $\tilde{\boldsymbol{h}}_c$ using OLS to get the estimate of the compressed DWT coefficients as

$$\hat{\tilde{\boldsymbol{h}}}_c = (\breve{\boldsymbol{\Phi}}^T\breve{\boldsymbol{\Phi}})^{-1}\breve{\boldsymbol{\Phi}}^T\boldsymbol{y}, \tag{34}$$

where

$$\breve{\boldsymbol{\Phi}} \triangleq \begin{bmatrix} \breve{\boldsymbol{u}}^T(k-N+1) \\ \breve{\boldsymbol{u}}^T(k-N+2) \\ \vdots \\ \breve{\boldsymbol{u}}^T(k) \end{bmatrix} = \boldsymbol{\Phi}\boldsymbol{X}. \tag{35}$$

The FIR model is then estimated by reconstruction using the IDWT as follows:

$$\hat{\boldsymbol{h}}_c = \boldsymbol{W}^{-1}(\boldsymbol{P}\hat{\tilde{\boldsymbol{h}}}_c) = \boldsymbol{X}\hat{\tilde{\boldsymbol{h}}}_c. \tag{36}$$

Bias and variance of FIR $\hat{\boldsymbol{h}}_c$

Theorem 1. The bias $\boldsymbol{b}_c \triangleq E[\boldsymbol{h} - \hat{\boldsymbol{h}}_c] = \boldsymbol{h} - E[\hat{\boldsymbol{h}}_c]$ in the estimate of the FIR model kernel $\hat{\boldsymbol{h}}_c$ using wavelet compression is equal to

$$\boldsymbol{b}_c = \boldsymbol{h} - \boldsymbol{h}_c, \tag{37}$$

where $\boldsymbol{h}$ is the true kernel, and $\boldsymbol{h}_c$ is the true DWT-compressed approximation of $\boldsymbol{h}$.

Proof. See Vuthandam (1997).

Theorem 2. The variance of the estimator $\hat{\boldsymbol{h}}_c$ is

$$E\left[(\boldsymbol{h} - \hat{\boldsymbol{h}}_c)(\boldsymbol{h} - \hat{\boldsymbol{h}}_c)^T\right] = \boldsymbol{b}_c\boldsymbol{b}_c^T + \sigma^2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}^T. \tag{38}$$

Proof. See Vuthandam (1997).

Corollary 1. The mean square error of the estimate $\hat{\boldsymbol{h}}_c$ is given by

$$E\left[(\boldsymbol{h} - \hat{\boldsymbol{h}}_c)^T(\boldsymbol{h} - \hat{\boldsymbol{h}}_c)\right] = \frac{1}{n}\left\{\boldsymbol{b}_c^T\boldsymbol{b} + \sigma^2\,\mathbf{trace}[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}^T]\right\}. \tag{39}$$

Choice of retained DWT coefficients for compression

The bias term $\boldsymbol{b}_c^T\boldsymbol{b}_c$ in Eqs. 38 and 39 can be made small by retaining appropriate DWT coefficients, which, however, we seek to determine in the first place. The choice of retained DWT coefficients is made using a number of factors discussed below.

The first step involves a choice of retaining DWT coefficients based on prior knowledge of the process. Let the prior knowledge available to us about the process be a very crude FIR model kernel denoted by $\boldsymbol{h}_0$, with DWT denoted by $\tilde{\boldsymbol{h}}_0$. The initial set of retained coefficients is obtained by discarding DWT coefficients in $\tilde{\boldsymbol{h}}_0$ that are small, that is,

$$I_{0,c} \triangleq \{i : |\tilde{h}_{0,i}| \geq \delta\}, \tag{40}$$

where δ is some small (possibly resolution-level-dependent) positive threshold value (not to be confused with thresholding used in denoising of signals by Donoho and coworkers (1995)), and $I_{0,c}$ refers to the initial set of indices of the retained DWT coefficients. This compression introduces a small error, as the following theorem shows.

Theorem 3. Consider an FIR model $L: u \mapsto y$ with kernel $\boldsymbol{h}$, which is compressed to the model $L_c: u \mapsto y$ using the DWT with a threshold δ. Then the modeling error resulting from the compression is bounded as

$$\| L - L_c \|_{i2} \leq c\delta, \tag{41}$$

where c is a constant that depends on n and the chosen wavelet basis, and $\| \cdot \|_{i2}$ denotes the induced 2-norm of an operator.

Proof. See the Appendix.

Once we have a set of retained DWT coefficients, we identify these coefficients from process input–output data. The FIR coefficients are then reconstructed from these identified DWT coefficients using the IDWT. Using the identified model, we obtain the prediction residuals

$$\hat{e}(k) = y(k) - \hat{\boldsymbol{h}}_c^T\boldsymbol{u}(k). \tag{42}$$

The residuals are computed for a cross-validation data set (not used in the identification) and are tested for

1. Independence of residuals: The autocorrelations of the residuals are tested for independence (whiteness) by using the method outlined in Ljung (1987). The autocorrelation estimate at delay τ of the residuals $\hat{e}(k)$, for $k = 1, \ldots, N_x$, (where N_x is the length of the cross-validation set) is given by

$$\hat{r}_e(\tau) = \frac{1}{\hat{\sigma}^2 N_x} \sum_{k=1}^{N_x - |\tau|} e(k)e(k+\tau), \tag{43}$$

where $\hat{\sigma}^2$ is the variance estimate of the residuals. The residuals can then be checked for independence by checking whether

$$|\hat{r}_e(\tau)| \leq \sqrt{\frac{P}{N_x}}\, K_\alpha, \tag{44}$$

where K_α is the α-level of the normal distribution $N(0,1)$.

$$P = \sum_{\tau=-N_x+1}^{N_x-1} \hat{r}_e^2(\tau). \tag{45}$$

2. Residual sum of squares (R_{ss}): We seek to retain DWT coefficients that would make R_{ss} as small as possible. R_{ss} is defined as

$$R_{ss} = \sum_{k=1}^{N_x} e^2(k). \tag{46}$$

After identifying an FIR model, with the initial set of retained DWT coefficients, we check the residuals using the preceding tests. We then add/drop coefficients and repeat the identification, and residual testing. This procedure is repeated until the residuals are independent (with some confidence), and R_{ss} is small enough.

Remarks

- Our computational experience has shown that all DWT coefficients corresponding to the coarsest resolution level should be retained. This helps proper identification of the steady-state characteristics of the impulse response.
- The indices of the retained DWT that detail coefficients corresponding to coarser resolution levels (0–2 typically) of the prior crude model $\boldsymbol{h}_0$ end up being good choices for the model being identified as well. The reason for this is that the coarser resolution details capture the lower-frequency content of the FIR models. It is the indices of the finer resolution details that are usually different for different models.
- The outlined procedure to select the coefficients suggests that we may have to try a large number of combinations of DWT coefficients to arrive at the best set. This is, however, not the case because

The coarse resolution level indices of retained components usually remain the same for typical processes modeled by FIR models.

Only very few of the finer resolution level coefficients need to be retained, and they tend to be the early ones.

Choosing coefficients at a coarser level is independent of the retained coefficients at finer levels. Therefore, a bottom-up procedure for selection of retained coefficients is suggested, that is, we start choosing coefficients from coarser resolution levels first, and then move upward to finer levels.

Algorithm

The following steps are suggested for choosing the retained DWT coefficients:

Step 1. Obtain the DWT $\tilde{\boldsymbol{h}}_0$ of the prior FIR model $\boldsymbol{h}_0$. Obtain a set of indices for detail coefficients to retain, according to a certain threshold (possibly resolution-level-dependent). Retain all indices corresponding to the coarsest resolution signal. Let this set of retained coefficient indices be denoted by I_c. Set the resolution level $l = 1$.

Step 2. Identify the DWT coefficients corresponding to indices in I_c using experimental input–output data. Construct FIR model.

Step 3. Perform tests on the prediction residuals on cross-validation experimental data.

Step 4. If the residuals test satisfactorily (i.e., independent and small R_{ss}), then go to Step 6, otherwise continue.

Step 5. Add/Drop DWT details corresponding to level l, and update retained indices set I_c, and repeat steps 2–4.

Step 6. If $l < n_r - 1$, increase l and repeat steps 2–5.

Step 7. Stop.

Example

A subsystem of a steam gas reformer is considered (Meziou and Alatiqi, 1992). The reformer can operate at either 58% or 100% capacity. When the reformer moves from one capacity to the other, the transfer function between the steam-to-carbon ratio and temperature changes from

$$H(s) = \frac{0.841(-2.5s+1)e^{-3s}}{(44s^2+11s+1)} \tag{47}$$

at 58% capacity to

$$H(s) = \frac{1.118(-2.5s+1)e^{-1.5s}}{(44s^2+11s+1)}. \tag{48}$$

Alatigi and Meziou (1991) show that the process can be satisfactorily controlled at 58% capacity by a controller designed on the basis of the 58% capacity model. However, if that controller is used to control the plant at 100% capacity, the resulting closed-loop behavior is highly oscillatory, hence unsatisfactory. Based on that, these authors establish the need for the development of an improved process model at 100% capacity. We solve this problem using the approach proposed in this article.

We seek to identify a discrete-time FIR model for the steam reformer at 100% capacity. With sampling performed every 0.5 min, the process has a settling time of around 90 sampling intervals. The number of coefficients we choose for the FIR model is $n = 96$. Three thousand process input–output data points are generated by applying a PRBS input to the process. The time constant of the process is around 30 sampling intervals. A PRBS input with a switching period of 25 sampling periods is used to excite the process. The output is cor-

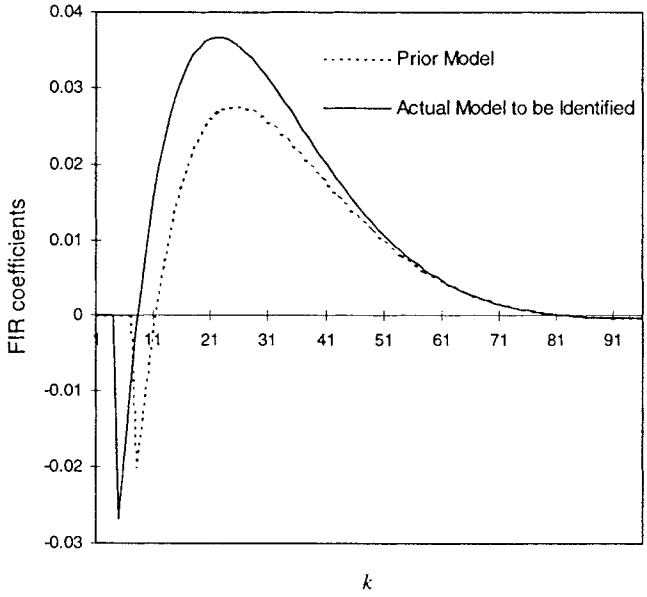

Figure 4. FIR coefficients of prior model and actual model to be identified.

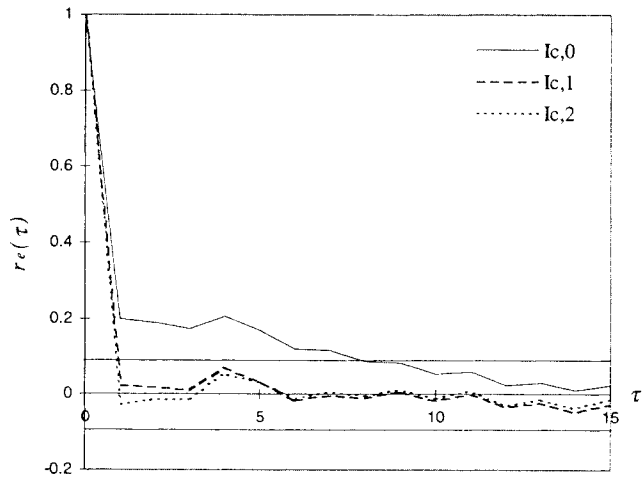

Figure 5. Residual autocorrelations with lag (horizontal bars indication of 95% confidence intervals).

rupted by additive Gaussian noise with zero mean and a standard deviation of 0.05. The first 2,000 points are used for identification, and the next 1,000 points for cross validation of the identified model.

Through this example we perform the following:

- Demonstrate the proposed FIR identification methodology with wavelet compression.
- Compare the proposed method to other existing FIR identification schemes.
- Compare the proposed method using the QS wavelet compression to Haar wavelet compression (corresponding to heuristic methods used in practice).

FIR identification using wavelet compression

In this subsection we identify the FIR model coefficients (contained in $\boldsymbol{h}$) using the proposed wavelet compression scheme with the QS basis. The selection of wavelet coefficients to be retained is demonstrated by first using some prior knowledge of the process, and then by performing tests on a cross-validation data set.

Prior knowledge of the process is available as an FIR model at a different operating point (58% capacity). The prior FIR model, and the FIR model to be identified (unknown) are shown in Figure 4. As an initial guess for the retained wavelet coefficients to be identified, we retain all the coarsest resolution coefficients and the detail at resolution levels 0, 1, and 2. The retained coefficients at the initial step ($I_{0,c}$) are

1 2 3	4 5	7 8	13 14
all coefficients 1,2,3 at lowest resolution level	details 1,2 at resolution level 0	details 1,2 at resolution level 1	details 1,2 at resolution level 2

The FIR model is determined by identifying these coefficients using the first 2,000 input–output data points. This model is used to predict the output in the next 1,000 points, and the residuals are obtained using the 1,000 points from the cross-validation set. The autocorrelation of the residuals is analyzed to examine their whiteness using the method described in the previous section. The residuals are found to be autocorrelated up to a lag of 8. We then add coefficients 25 and 26 (details 1, 2 at resolution level 3) to get the new index set $I_{1,c}$. The residual analysis is repeated, and we find that the residuals appear white (95% confidence level), and also there is a reduction in R_{ss}. We then add coefficient 51 (detail 1 at resolution level 4) to get the index set $I_{2,c}$. Residual analysis now reveals that the residuals are white and there is a further reduction in R_{ss}. The autocorrelations of the residuals (up to a lag of 15) for the 3 set of retained indices and 95% confidence bounds (for whiteness) are plotted in Figure 5. These results are summarized in Table 2.

Comparison with other methods

Through this example we compare the identification of following schemes:

(A) Direct nonparsimonious FIR identification

(B) FIR identification using regularization

(C) FIR model determination from ARX identification with exact model order selection

(D) FIR model determination from ARX identification with inexact model order selection

(E) FIR model identification using wavelet compression with Haar basis

(F) FIR model identification using wavelet compression with QS basis.

We compare these models based on the mean square errors of their FIR coefficients. The errors are measured as deviations of the estimated FIR coefficients from the true

Table 2. Selection of Retained DWT Coefficients for Identification and Residual Tests

Index Set	Whiteness Test	R_{ss}
$I_{0,c}$	Colored	2.5452
$I_{1,c}$	White	2.4135
$I_{2,c}$	White	2.4081

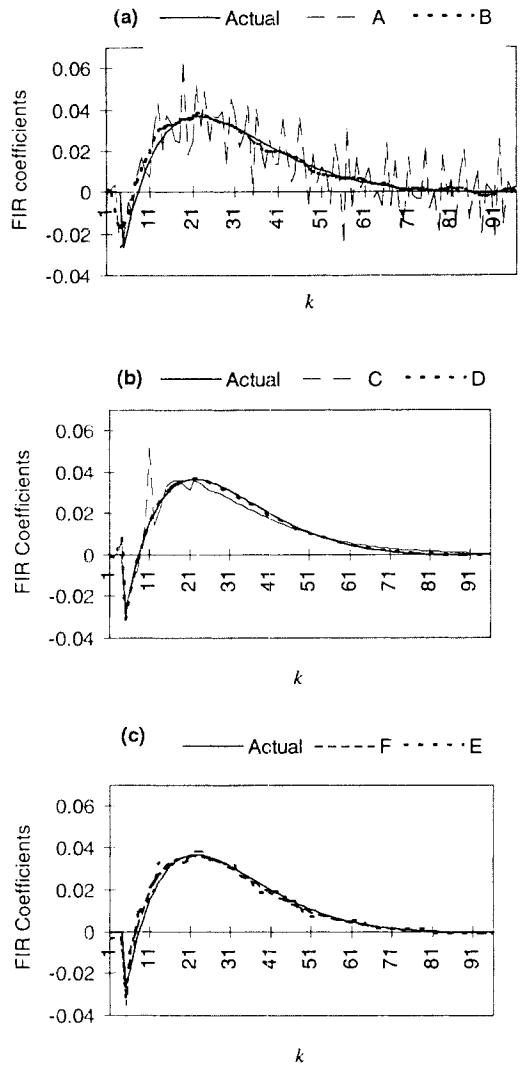

Figure 6. FIR coefficients: (a) actual, A and B; (b) actual, C, and D; (c) actual, E, and F.

(assumed known for the purpose of comparison) FIR coefficients as follows:

$$MSE_h = \frac{1}{n}\sum_{j=1}^{n}\left(h_j - \hat{h}_j\right)^2, \tag{49}$$

where the h_j refer to the true FIR coefficients, and the $\hat{h}_j$ refer to the identified FIR coefficients using the different techniques. The FIR models identified using methods A–F are plotted along with the actual FIR in Figure 6.

Comparisons based on output prediction errors are also made on a cross-validation data set consisting of 1,000 input–output data points. The residual squared sums (R_{ss}) for methods A–F are computed as in Eq. 46.

Finally, we compare the steady-state gains obtained using the different FIR models as

$$\hat{G}_{ss} = \sum_{j=1}^{n}\hat{h}_j, \tag{50}$$

Table 3. Comparison of FIR Identification Methods A–F

Method	MSE_h (10^{-5})	R_{ss}	G_{ss}
A	15.98	3.312	1.1189
B	0.8226	3.165	1.1188
C	0.1436	2.470	1.1245
D	2.0778	4.389	1.1232
E	0.7982	3.059	1.1189
F	0.6393	2.408	1.1187

where $\hat{h}_j$ are the identified FIR coefficients using methods A–F. The true gain of the process is 1.118. All the results of these comparisons are given in Table 3.

Comparison of the wavelet compression method to identify FIR models reveals a number of important characteristics of the proposed scheme.

- As can be seen from the comparison of MSE_h, method F (using wavelet compression with QS basis) is lowest second to method C, which is the ARX model identification with exact choice of model order and with the appropriate noise model being fitted. However the ARX scheme does very poorly when an incorrect model order is selected.
- Comparison of steady-state gains reveals that the wavelet method F is slightly better than methods A, B, and E, and considerably better than the ARX schemes.
- The tests based on MSE_h and G_{ss} cannot be performed in real situations, as the true model and model gain would not be available.
- The more realistic test is cross validation on a fresh set of testing data. The wavelet method F shows the lowest R_{ss}, on this cross-validation test. Again, method C performs almost as well as the wavelet method, but method D performs poorly, again because of incorrect model selection.

We also make comparisons of the wavelet compression method, using the QS wavelet basis, with method E, which uses the Haar wavelet basis. Again, method F performs better than this method in all the tests. Also, the QS basis provides better compression, than compression with the Haar wavelet basis. Compression is determined as

$$\text{Compression} = \frac{\text{total number of FIR kernels}}{\text{number of identified DWT coefficients}}.$$

With the QS wavelet basis we get a compression of 7.4, while with the Haar wavelet basis we get a compression of 5.1.

Conclusions

In this article we have proposed a new methodology for the identification of process FIR models using wavelet compression techniques. We have made comparisons with other existing schemes, on the basis of closeness of fit with true process models, steady-state gains, and tests on prediction residuals. We have shown that our method (with the QS wavelet basis) performs comparably to or better than the existing schemes for FIR identification for the example considered. One of the attractive features of the proposed approach is that qualitative prior knowledge about a process model can be easily translated to quantitative information that can be used in process identification. We have also shown

that existing heuristic approaches for piecewise constant approximation of FIR model kernels are equivalent to compression using Haar wavelets, and smoother wavelets can be used instead with better results.

Literature Cited

Alatiqi, I. M., and A. M. Meziou, "Dynamic Simulation and Adaptive Control of an Industrial Steam Gas Reformer," *Comp. Chem. Eng.*, **15**(3), 147 (1991).

Astrom, K. J., and B. Wittenmark, *Adaptive Control*, Addison-Wesley, Reading, MA (1994).

Bakshi, B. R., A. Koulouris, and G. Stephanopoulos, "Wave-Nets: Novel Learning Techniques, and the Induction of Physically Interpretable Models," *Proc. SPIE—The International Society for Optical Engineering Wavelet Applications*, Orlando, FL, p. 637 (1994).

Bendetto, J. J., "Irregular Sampling and Frames," *Wavelet Analysis and Its Applications*, Vol. 2, C. Chui, ed., Academic Press, New York (1992).

Box, G. E. P., and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Rev. ed., Holden-Day, San Francisco (1976).

Dayal, B. S., and J. F. MacGregor, "Identification of Finite Impulse Response Models: Methods and Robustness Issues," *Ind. Eng. Chem. Res.*, **35**, 4078 (1996).

Donoho, D. L., and I. M. Johnstone, "Adapting to Unknown Smoothness via Wavelet Shrinkage," *J. Amer. Stat. Assoc.*, **90**(432), 1200 (1995).

Ljung, L., *System Identification: Theory for Users*, Addison-Wesley, Reading, MA (1987).

MacGregor, J. F., T. Kourti, and J. V. Kresta, "Multivariate Identification: A Study of Several Methods," *IFAC ADCHEM Proc.*, Toulouse, France (1991).

Mallat, S. G., "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbf{R})$," *Trans. Amer. Math. Soc.*, **315**(1), 69 (1989).

Palavajjhala, S., R. L. Motard, and B. Joseph, "Process Identification using Discrete Wavelet Transforms: Design of Prefilters," *AIChE J.*, **42**(3), 777 (1996).

Pati, Y. C., R. Rezaiifar, P. S. Krishnaprasad, and W. P. Dayawansa, "A Fast Recursive Algorithm for System Identification and Model Reduction Using Rational Wavelets," *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Vol. 1, Pacific Grove, CA, p. 35 (1993).

Ricker, N. L., "The Use of Biased Least-Squares Estimators for Parameters in Discrete-Time Pulse-Response Models," *Ind. Eng. Chem. Res.*, **27**, 343 (1988).

Söderström, T., and P. Stoica, *System Identification*, Prentice Hall, Englewood Cliffs, NJ (1989).

Strang, G., and T. Nguyen, *Wavelets and Filter Banks*, Wellesley, Cambridge, MA (1996).

Sureshbabu, N., and J. A. Farrell, "Wavelet Based System Identification for Nonlinear Control Applications," *Proc. IEEE Int. Symp. on Intelligent Control*, Monterey, CA, p. 236 (1995).

Tsatsanis, M. K., and G. B. Giannakis, "Time-Varying System Identification and Model Reduction Using Wavelets," *IEEE Trans. Signal Process.*, **SP-41**, 3512 (1993).

Tsatsanis, M. K., and G. B. Giannakis, "A Basis Expansion Approach for Detecting Transient Plant Disturbances and Jumps," *Proc. Conf. on Decisions and Control*, Lake Buena Vista, FL, p. 3412 (1994).

Vuthandam, P., *Performance Enhancement of Model-Based Predictive Controllers Through Robust and Adaptive Designs*, PhD Diss., Texas A&M Univ., College Station, TX (1997).

Wise, B. M., and N. L. Ricker, "Identification of Finite Impulse Response Models by Principal Component Regression: Frequency Response Properties," *Process Control Qual.*, **4**, 77 (1992).

Appendix

Proof of Therorem 3. The induced 2-norm of an operator $L: u \mapsto y = Lu$, defined as $y(k) = (Lu)(k) \triangleq \boldsymbol{h}^T \boldsymbol{u}(k)$, is given by

$$\|L\|_{i2} = \sup_{0 \le \omega < 2\pi} \left| \sum_{k=1}^{n} h_k e^{-jk\omega} \right|.$$

Let $L_c: u \mapsto y = L_c u$ be defined as $y(k) = (L_c u)(k) \triangleq \boldsymbol{h}_c^T \boldsymbol{u}(k)$. Then

$$\begin{aligned} \|L - L_c\|_{i2} &= \sup_{0 \le \omega < 2\pi} \left| \sum_{k=1}^{n} (h_k - h_{c,k}) e^{-jk\omega} \right| \\ &\le \sup_{0 \le \omega < 2\pi} \sum_{k=1}^{n} \left| h_k - h_{c,k} \| \overbrace{e^{-jk\omega}}^{1} \right| \\ &= \sup_{0 \le \omega < 2\pi} \sum_{k=1}^{n} \left| h_k - h_{c,k} \right| = \sum_{k=1}^{n} \left| h_k - h_{c,k} \right| \le n\epsilon, \end{aligned} \quad \text{(A1)}$$

where

$$\begin{aligned} \epsilon = \max_k | h_k - h_{c,k} | &= \|\boldsymbol{h} - \boldsymbol{h}_c\|_\infty = \|\boldsymbol{W}^{-1}\left(\tilde{\boldsymbol{h}} - \boldsymbol{P}\tilde{\boldsymbol{h}}_c\right)\|_\infty \\ &\le \|\boldsymbol{W}^{-1}\|_{i\infty} \|\tilde{\boldsymbol{h}} - \boldsymbol{P}\tilde{\boldsymbol{h}}_c\|_\infty = \|\boldsymbol{W}\|_{i\infty} \delta. \end{aligned} \quad \text{(A2)}$$

Therefore, if $c = n \cdot \|\boldsymbol{W}\|_{i\infty}$, we have

$$\|\boldsymbol{h} - \boldsymbol{h}_c\|_{i2} \le c \cdot \epsilon. \quad \text{(A3)}$$

Manuscript received Mar. 20, 1997, and revision received Sept. 22, 1997.