

Research Article

Reduced Noise Effect in Nonlinear Model Estimation Using Multiscale Representation

Mohamed N. Nounou¹ and Hazem N. Nounou²

¹ Chemical Engineering Program, Texas A&M University at Qatar, P.O. Box 23874, Doha, Qatar

² Electrical and Computer Engineering Program, Texas A&M University at Qatar, P.O. Box 23874, Doha, Qatar

Correspondence should be addressed to Mohamed N. Nounou, mohamed.nounou@qatar.tamu.edu

Received 24 November 2009; Revised 25 March 2010; Accepted 13 May 2010

Academic Editor: Borut Zupancic

Copyright © 2010 M. N. Nounou and H. N. Nounou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nonlinear process models are widely used in various applications. In the absence of fundamental models, it is usually relied on empirical models, which are estimated from measurements of the process variables. Unfortunately, measured data are usually corrupted with measurement noise that degrades the accuracy of the estimated models. Multiscale wavelet-based representation of data has been shown to be a powerful data analysis and feature extraction tool. In this paper, these characteristics of multiscale representation are utilized to improve the estimation accuracy of the linear-in-the-parameters nonlinear model by developing a multiscale nonlinear (MSNL) modeling algorithm. The main idea in this MSNL modeling algorithm is to decompose the data at multiple scales, construct multiple nonlinear models at multiple scales, and then select among all scales the model which best describes the process. The main advantage of the developed algorithm is that it integrates modeling and feature extraction to improve the robustness of the estimated model to the presence of measurement noise in the data. This advantage of MSNL modeling is demonstrated using a nonlinear reactor model.

1. Introduction

Process models are a core element in many process operations, such as process control and optimization [1, 2], and the accuracy of these models has a direct impact on the quality of these operations and ultimately on the overall performance of the process. Therefore, it is always sought to improve the accuracy of the models used in these operations. Since fundamental models are not always available, in many cases, it is relied on deriving empirical or semiempirical models from measurement of the process variables. However, data-driven approaches for model estimation are associated with many challenges, which include defining the model structure and accounting for the presence of measurement noise in the data. The objective of this work is to improve the prediction accuracy of the well-known class of nonlinear, but linear-in-the-parameters, process models using multiscale representation to account for the presence of measurement noise in the data.

The presence of measurement noise, even in small amounts, can largely affect the estimated model's prediction

accuracy. Therefore, such noise needs to be filtered for improved model's prediction. Modeling of prefiltered measured data does not usually provide satisfactory performance [3]. This is because applying data filtering without taking into account the input-output relationship may result in the removal of certain features from the data which are important for the model. Therefore, filtering and modeling need to be integrated for satisfactory model estimation.

Unfortunately, measured data usually have a multiscale character, which means that they contain features and noise that have varying contributions over both time and frequency [4]. For example, an abrupt change in the data spans a wide range in the frequency domain and a small range in the time domain, while a slow change spans a wide range in the time domain and a small range in the frequency domain. Filtering such data using the conventional low pass filters usually does not result in a good noise-feature separation because these filtering techniques classify noise as high-frequency features and filter the data by removing features with frequency higher than a defined frequency threshold. Thus, modeling multiscale data requires

developing multiscale modeling techniques that account for this multiscale nature of the data.

Many investigators have used multiscale techniques to improve the accuracy of estimated empirical models [3, 5–9]. For example, the authors in [5] showed how to use wavelet representation to design wavelet prefilters for process modeling purposes. In [3], the author discussed some of the advantages of using multiscale representation in empirical modeling, and in [6], he enhanced the noise removal ability of the Principal Component Analysis (PCA) model by constructing multiscale PCA models, which he also used in process monitoring. Also, the authors in [7–9] used multiscale representation to reduce collinearity and shrink the large variations in the Finite Impulse Response (FIR) model parameters. Furthermore, in [10], the authors used wavelets as modulating function for control-related system identification. Finally, the authors in [11, 12] used wavelet-based representation to enhance the accuracy and parsimony of the Autoregressive with exogenous variable (ARX) model and the Takagi-Sugeno fuzzy model.

In this work, multiscale representation of data is utilized to improve the prediction accuracy of nonlinear models by developing a multiscale nonlinear (MSNL) process modeling algorithm that accounts for the presence of noise in the data and improves the model's prediction accuracy. The developed MSNL modeling algorithm integrates modeling and data filtering by constructing multiple nonlinear models at multiple scales using the scaled signal approximations of the input and output data and then selecting, among all scales, the model that provides the optimum prediction and maximum noise-feature separation.

The rest of this paper is organized as follows. In Section 2, the problem statement and the formulation and estimation of the linear-in-the-parameter nonlinear model are introduced, followed by a description of the wavelet-based multiscale representation of data in Section 3. Then, in Section 4, the formulation, estimation, and algorithm for MSNL modeling are described. Then, in Section 5, the performance of the developed MSNL modeling algorithm is illustrated and compared with that of the time-domain method. Finally, the paper is concluded with few remarks in Section 6.

2. Problem Statement

In this work, we address the problem of empirically (from measurement of the input and output variables) estimating linear-in-the-parameters nonlinear models which are less affected by the presence of measurement noise in the data. Given measurements of the input-output data, that is, $\{u(k)\}_{k=1,\dots,n}$ and $\{y(k)\}_{k=1,\dots,n}$, where the output is assumed to be contaminated with additive zero mean Gaussian noise (i.e., $y = \tilde{y} + \varepsilon$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, and the superscript “ \sim ” represents the noise-free variables), it is desired to construct a linear-in-the-parameter nonlinear model of the form

$$y(k+1) = \sum_{i=0}^m \beta_i f_i(z_i(k), \theta_i), \quad (1)$$

where $y(k+1)$ is the process output at time step $(k+1)$, $z_i(k) = \{y(k), \dots, y(k-p_i), u(k), \dots, u(k-q_i)\}$, $f_i(\cdot)$ is the i th nonlinear basis function ($i \in [1, m]$), where these basis functions are assumed to be known, and β_i is the model parameter corresponding to the i th basis function, which is a function of the parameter vector θ_i . Also, p_i and q_i are the number of lagged outputs and inputs used in the model, respectively.

2.1. Nonlinear Model Estimation. The nonlinear model shown in (1) can also be represented in matrix form as

$$Y = X\beta, \quad (2)$$

where

$$Y = [y(k+1) \ y(k) \ y(k-1) \ \dots]^T,$$

$X =$

$$\begin{bmatrix} f_1(z_1(k), \theta_1) & f_2(z_2(k), \theta_2) & \dots & f_m(z_m(k), \theta_m) \\ f_1(z_1(k-1), \theta_1) & f_2(z_2(k-1), \theta_2) & \dots & f_m(z_m(k-1), \theta_m) \\ f_1(z_1(k-2), \theta_1) & f_2(z_2(k-2), \theta_2) & \dots & f_m(z_m(k-2), \theta_m) \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}, \quad (3)$$

and the parameter vector $\beta = [\beta_1 \ \beta_2 \ \dots \ \beta_m]^T$ which can be estimated using ordinary least squares (OLSs) regression as follows:

$$\{\hat{\beta}\}_{\text{OLS}} = (X^T X)^{-1} X^T Y. \quad (4)$$

Note that since OLS minimizes the output prediction mean squares error in the estimation of the model parameters, it implicitly assumes that all variables in the information matrix, X , are noise-free. In the nonlinear model given in (2), however, past outputs are also a part of the information matrix. Therefore, the accuracy of the model prediction can be affected by the presence of measurement noise in the data, especially in large amounts. This is because OLS will provide biased estimate of the parameters, which degrades the accuracy of the model's prediction. More details about bias and its effect on the model's prediction can be found in [13]. In this paper, an alternative modeling approach will be developed to reduce the effect of measurement noise in the data on the prediction accuracy of the estimated model. This approach will utilize multiscale wavelet-based representation of data, which is introduced next.

3. Multiscale Representation of Data

A proper way of analyzing real data requires their representation at multiple scales. This can be achieved by expressing the data as a weighted sum of orthonormal basis functions, which are defined in both time and frequency, such as wavelets. Wavelets are a computationally efficient family of multiscale basis functions. A signal can be represented at multiple resolutions by decomposing the signal on a family

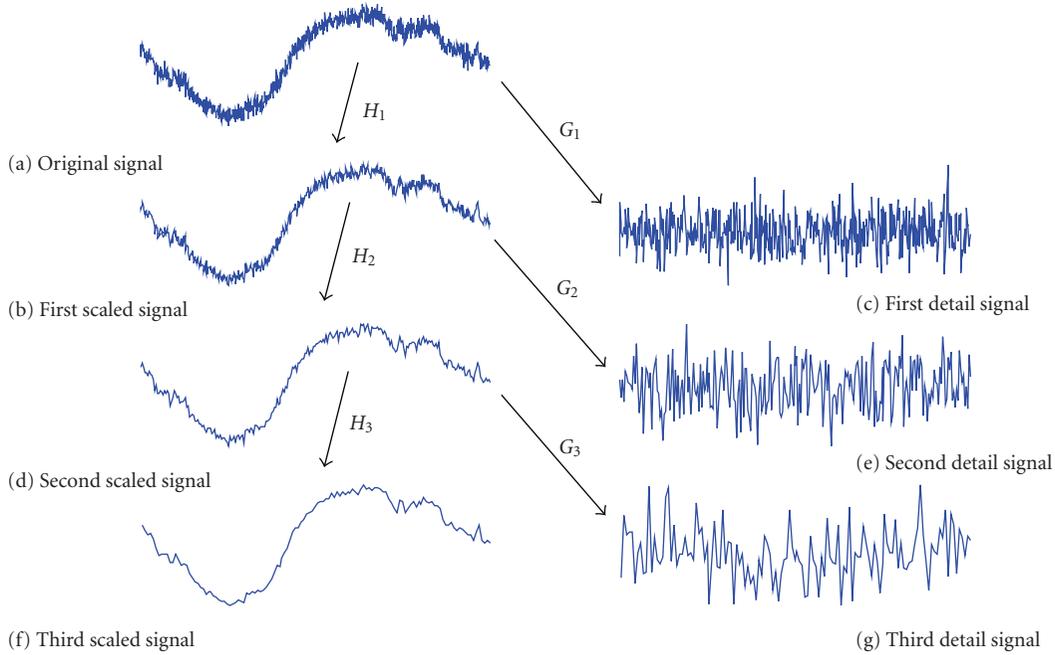


FIGURE 1: A schematic diagram of data representation at multiple scales.

of wavelets and scaling functions. The signals in Figures 1(b), 1(d), and 1(f) are at increasingly coarser scales compared to the original signal in Figure 1(a). These scaled signals are determined by projecting the original signal on a set of orthonormal scaling functions of the form

$$\phi_{jk}(t) = \sqrt{2^{-j}}\phi(2^{-j}t - k) \quad (5)$$

or equivalently by filtering the signal using a low pass filter of length r , $h = [h_1 \ h_2 \ \dots \ h_r]$, derived from the scaling functions [14]. On the other hand, the signals in Figures 1(c), 1(e), and 1(g), which are called the detail signals, capture the differences between any scaled signal and the scaled signal at the finer scale.

These detail signals are determined by projecting the signal on a set of wavelet basis functions of the form

$$\psi_{jk}(t) = \sqrt{2^{-j}}\psi(2^{-j}t - k) \quad (6)$$

or equivalently by filtering the scaled signal at the finer scale using a high pass filter of length r , $g = [g_1 \ g_2 \ \dots \ g_r]$, derived from the wavelet basis functions. Therefore, the original signal can be represented as the sum of all detail signals at all scales and the scaled signal at the coarsest scale as follows:

$$x(t) = \sum_{k=1}^{n2^{-j}} a_{jk}\phi_{jk}(t) + \sum_{j=1}^J \sum_{k=1}^{n2^{-j}} d_{jk}\psi_{jk}(t), \quad (7)$$

where j , k , J , and n are the dilation parameter, translation parameter, maximum number of scales (or decomposition depth), and the length of the original signal, respectively [15, 16].

Fast wavelet transform algorithms of $O(n)$ complexity for a discrete signal of dyadic length (of length 2^j where j is

a positive integer) have been developed [14]. For example, the wavelets and scaling functions coefficients at a particular scale (j), d_j and a_j , can be computed in a compact fashion by multiplying the scaling coefficient vector at the finer scale, a_{j-1} , by the matrices, G_j and H_j , respectively, that is,

$$a_j = H_j a_{j-1}, \quad d_j = G_j a_{j-1}, \quad (8)$$

where

$$H_j = \begin{bmatrix} h_1 & \dots & h_r & 0 & 0 \\ 0 & h_1 & \dots & h_r & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & h_1 & \dots & h_r \end{bmatrix}_{n2^{-j} \times n2^{-j+1}}, \quad (9)$$

$$G_j = \begin{bmatrix} g_1 & \dots & g_r & 0 & 0 \\ 0 & g_1 & \dots & g_r & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & g_1 & \dots & g_r \end{bmatrix}_{n2^{-j} \times n2^{-j+1}}. \quad (10)$$

Note that the length of the scaling and detail signals decreases dyadically at coarser resolutions (higher j). In other words, the length of scaled signal at scale (j) is half the length of scaled signal at the finer scale, ($j - 1$). This is due to downsampling, which is used in discrete wavelet transform. Just as an example to illustrate the multiscale decomposition procedure and to introduce some terminology, consider the

following discrete signal, Y_o , of length (n) in the time domain (i.e., $j = 0$):

$$Y_o = [y_o(1) \ y_o(2) \ \dots \ y_o(k) \ \dots \ y_o(n)]^T. \quad (11)$$

The scaled signal approximation of Y_o at scale (j), which can be written as

$$Y_j = [y_j(1) \ \dots \ y_j(k) \ \dots \ y_j(n2^{-j})]^T \quad (12)$$

can be computed as follows:

$$Y_j = H_j Y_{j-1} = H_j H_{j-1} \dots H_1 Y_o. \quad (13)$$

Note that this decomposition algorithm is batch, that is, it requires the availability of the entire data set beforehand. An online wavelet decomposition algorithm has also been developed and used in data filtering [17].

4. Multiscale Nonlinear Process Modeling

In this section, the feature extraction abilities of multiscale representation of data are utilized to construct multiscale nonlinear models that are less affected by the presence of noise in the data. The main idea is to decompose the input-output data at multiple scales and construct a nonlinear model at each scale using the scaled signal approximations of the data. Then, among all scales, select the optimum nonlinear model that provides the best prediction and noise-feature separation.

The ability of multiscale representation to extract important features from data can be verified by computing the signal-to-noise ratio (SNR) of the scaled signal at multiple scales. Theoretically, the SNR at any scale can be computed as follows:

$$\text{SNR}(j) = \frac{\text{var}(\tilde{y}_j)}{\text{var}(\tilde{y}_j - y_j)}, \quad (14)$$

where \tilde{y}_j is the noise-free scaled signal representation of the data at scale (j). It can be empirically illustrated that the SNR of the scaled signals peaks at some intermediate scale, which can be explained as follows. At very fine scales, high-frequency noise gets filtered out, which decreases the noise content and increases the SNR. However, at very coarse scales, important features start getting removed, which decreases the signal content and decreases the SNR. Therefore, there is an intermediate scale at which the SNR peaks. This observation will be used later to estimate the optimum modeling scale. Another characteristic of multiscale representation is that correlated noise gets decorrelated at multiple scales [4]. This gives another advantage to multiscale models over conventional ones.

4.1. MSNL Model Representation. Having computed the scaled signal approximations of the input-output data at

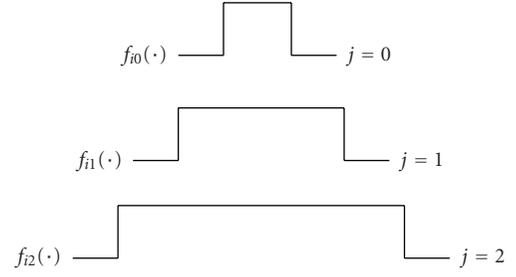


FIGURE 2: Dilation of basis functions at multiple scales.

multiple scales as described in Section 3, nonlinear models of the form

$$y_j(k+1) = \sum_{i=0}^m \beta_{ij} f_{ij}(z_{ij}(k), \theta_{ij}) \quad (15)$$

can be constructed at each scale (j) using these scaled signals of the input and output data, u_j and y_j . Note here that the basis functions used at scale (j), that is, $f_{ij}(z_{ij}(k), \theta_{ij})$, are not the same basis function used in the time domain, $f_i(z_i(k), \theta_i)$. This is because of the different transformations used to compute the scaled signals at different scales. Note, however, that the form of the model's basis functions at any scale can be defined in a similar fashion to that used to dilate the wavelet and scaling functions shown in (5) and (6). For example, for the basis function $f_i(z_i(k), \theta_i)$ in the time domain, the equivalent basis function at any scale (j) is

$$f_{ij}(z_{ij}(k), \theta_{ij}) = \sqrt{2^{-j}} f_i(2^{-j} z_i(k), \theta_i). \quad (16)$$

This effect of basis function dilation at multiple scales is illustrated in Figure 2, which shows that for a given time-domain ($j = 0$) nonlinear basis function, $f_{i0}(\cdot)$, the dilated basis function, $f_{ij}(z_{ij}(k), \theta_{ij})$, is stretched at coarser scales (larger j) to account for the dyadic downsampling used in multiscale representation.

Now, the nonlinear model at scale (j) can be written by combining (15) and (16) as follows:

$$y_j(k+1) = \sum_{i=0}^m \beta_{ij} \sqrt{2^{-j}} f_i(2^{-j} z_i(k), \theta_i). \quad (17)$$

4.2. MSNL Model Estimation. The nonlinear model at scale (j), shown in (17), can also be written in matrix form as

$$Y_j = X_j \beta_j, \quad (18)$$

where

$$Y_j = [y_j(k+1) \ y_j(k) \ y_j(k-1) \ \dots]^T \quad \beta_j = [\beta_{1j} \ \beta_{2j} \ \dots \ \beta_{mj}]^T,$$

$$X_j = \begin{bmatrix} f_{1j}(z_{1j}(k), \theta_{1j}) & f_{2j}(z_{2j}(k), \theta_{2j}) & \dots & f_{mj}(z_{mj}(k), \theta_{mj}) \\ f_{1j}(z_{1j}(k-1), \theta_{1j}) & f_{2j}(z_{2j}(k-1), \theta_{2j}) & \dots & f_{mj}(z_{mj}(k-1), \theta_{mj}) \\ f_{1j}(z_{1j}(k-2), \theta_{1j}) & f_{2j}(z_{2j}(k-2), \theta_{2j}) & \dots & f_{mj}(z_{mj}(k-2), \theta_{mj}) \\ \vdots & \vdots & \dots & \vdots \end{bmatrix}. \quad (19)$$

Therefore, the model parameters at scale (j) can be estimated using OLS as shown in (4) for the time domain model.

4.3. MSNL Modeling Algorithm. Based on the above discussion on the nonlinear model representation and estimation at multiple scales, the following algorithm is proposed for nonlinear multiscale (MSNL) identification:

- (1) decompose the input/output data using the wavelet-filter of choice at multiple scales;
- (2) at each scale and using the input/output scaled signals,
 - (a) define the structure of the model's basis functions at each scale given the basis functions in the time domain using (16),
 - (b) express the model in matrix form as shown in (18),
 - (c) estimate the nonlinear model using OLS and predict the output at each scale,
 - (d) compute the output prediction SNR as follows: $\text{SNR}(j) = \text{var}(\hat{y}_j)/\text{var}(\hat{y}_j - y_j)$, where \hat{y}_j is the predicted output at scale (j),
- (3) select the optimum model among all scales by choosing the one with the maximum output prediction SNR;
- (4) for the optimum model, reconstruct the predicted output back to the time domain.

5. Illustrative Example

In this section, the advantages of the MSNL modeling algorithm are illustrated through a simulated example. The model used in this example relates the concentration of the inlet stream to a stirred tank reactor (input) to the exit stream concentration (output). Let the inlet and exit stream concentrations of a species "A" be C_{Ai} and C_A , respectively, and let the flow rates in and out of the reactor be q (see

Figure 3). If species "A" is converted into "B" in the reactor according to the following reaction, $2A \rightarrow B$, with a reaction rate of $r = k_r C_A^2$, and the reactor volume is constant and equals V , the rate of change of species "A" can be expressed using mass balance as follows:

$$\frac{dC_A}{dt} = \frac{q}{V} C_{Ai} - \frac{q}{V} C_A - k_r C_A^2. \quad (20)$$

Discretizing the model shown in (20) with a sampling time interval of Δt , we obtain the following nonlinear discrete model:

$$C_A(k+1) = \left(\frac{q\Delta t}{V}\right) C_{Ai}(k) + \left(1 - \frac{q\Delta t}{V}\right) C_A(k) - k_r \Delta t C_A^2, \quad (21)$$

where C_{Ai} and C_A are the model input and output, respectively. Assuming the following values for the parameters: $q = 1$ L/s, $k_r = 5 \times 10^{-5}$ mole/(L.s), $V = 100$ L, and $\Delta t = 1$ s, the above model is used to generate data by applying a PRBS input changing between 0 and 2, and the output, which is assumed to be noise-free, is then contaminated with zero mean Gaussian noise. Different levels of noise (standard deviation $\sigma_\varepsilon = 0.05, 0.1, \text{ and } 0.15$, which approximately correspond to output signal to noise ratios of 50, 10, and 5) are used to test the robustness of the MSNL modeling algorithm. A 500-sample input-output data set, where $\sigma_\varepsilon = 0.15$, is shown in Figure 4.

The performance of the MSNL algorithm is compared to the time-domain counterpart by comparing their prediction mean square errors with respect to the noise-free output, that is,

$$\text{MSE} = \frac{1}{n} \sum_i^n (\hat{y}_i - \tilde{y}_i)^2, \quad (22)$$

where \hat{y} and \tilde{y} are the predicted and noise-free outputs, respectively. Note that such comparison is possible in this simulated example because the noise-free output is known. Also, in this example, the Haar wavelet and scaling functions are used in multiscale representation of the data. Please note that smoother wavelet filters (e.g., Daubechies) may provide better performance in the case of smoother data.

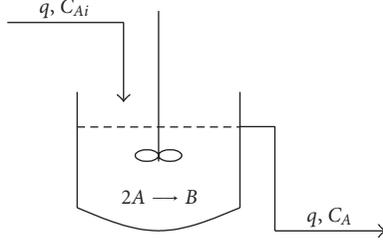


FIGURE 3: A schematic diagram of a stirred tank reactor.

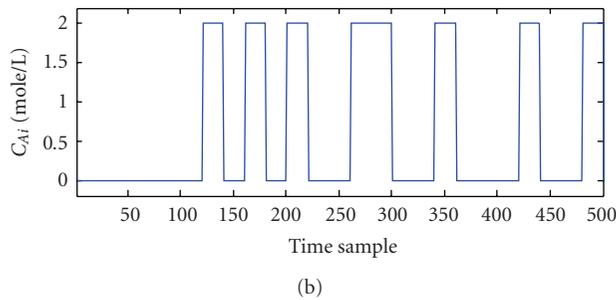
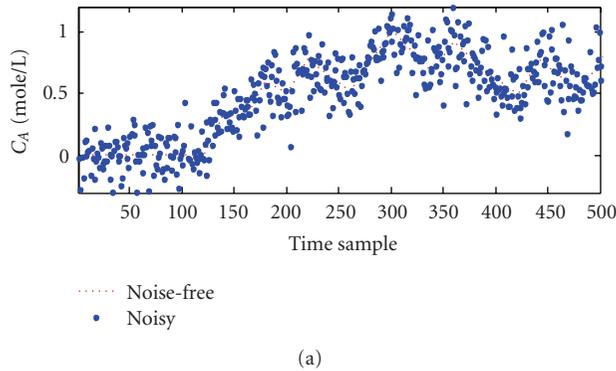


FIGURE 4: A 500-sample input-output data set.

To make statistically valid conclusions about the performances of the various modeling methods, a Monte Carlo simulation of 1000 realizations is performed, and the results are presented in Tables 1 and 2. Table 1 shows that MSNL models achieve a noticeable improvement over those estimated using the raw data (at the time domain) and that this improvement increases for larger noise contents. This improvement can also be visually seen from Figure 5, which shows the advantage of constructing nonlinear models at multiple scales.

Figure 5 shows the change in the models' accuracies at different scales. It can be seen that prediction accuracy improves at coarser scales but up to a certain level beyond which the quality of estimated model deteriorates. In other words, the accuracy improves from the time domain up to scale 3, but starting at scale 4, the quality of the model

TABLE 1: Comparison between the prediction means square errors of the multiscale and time-domain modeling methods.

Modeling Method	Prediction MSE ($\times 10^3$)		
	$\sigma_\epsilon = 0.05$	$\sigma_\epsilon = 0.1$	$\sigma_\epsilon = 0.15$
NL modeling	2.3	10.8	20.5
MSNL modeling	1.2	3.9	6.6

TABLE 2: Comparison between the prediction means square errors at multiple scales.

Scale	Prediction MSE ($\times 10^3$)		
	$\sigma_\epsilon = 0.05$	$\sigma_\epsilon = 0.1$	$\sigma_\epsilon = 0.15$
$j = 0$	2.3 (0)	10.8 (0)	20.5 (0)
$j = 1$	1.3 (24)	5.9 (1)	11.2 (0)
$j = 2$	1.2 (76)	3.6 (61)	6.5 (0)
$j = 3$	3.6 (0)	4.7 (38)	6.3 (79)
$j = 4$	8.6 (0)	9.2 (0)	10.0 (21)
$j = 5$	17.5 (0)	17.8 (0)	18.2 (0)

deteriorates. This can also be noted from Table 2, which reports the MSE of the estimated models at different scales and shows that there is an intermediate scale at which MSNL modeling is the best. This is because at very coarse scales, features in the data (which are important to the model) get eliminated and thus affect the model's quality. That is why it is very important to select the optimum scale for modeling. Table 2 also presents (in parentheses) the percentages at which each scale is selected as optimum using the SNR criterion and shows that the optimum scale increases or gets coarser (higher j) for higher noise levels. This makes sense because, for higher noise levels, more filtering is needed for good noise-feature separation. Also, note that a multiscale model estimated at a particular scale (j) will update its prediction every 2^j samples. This is because of the dyadic upsampling performed during the reconstruction of the model prediction, that is, one time sample prediction at scale (j) corresponds to 2^j samples in the time domain.

6. Conclusions

In this paper, the noise-feature separation capabilities of multiscale representation of data are exploited to improve the estimation and prediction accuracy of the linear-in-the-parameters nonlinear model, by presenting a multiscale nonlinear (MSNL) modeling algorithm with enhanced robustness to measurement noise in the data. The MSNL modeling algorithm integrates modeling and filtering by decomposing the input-output data and using the scaled signal approximations to construct different nonlinear models at different scales. Then, among all scales, the model with the largest output prediction SNR is selected as the optimum MSNL model. Finally, the performance of the MSNL algorithm is illustrated

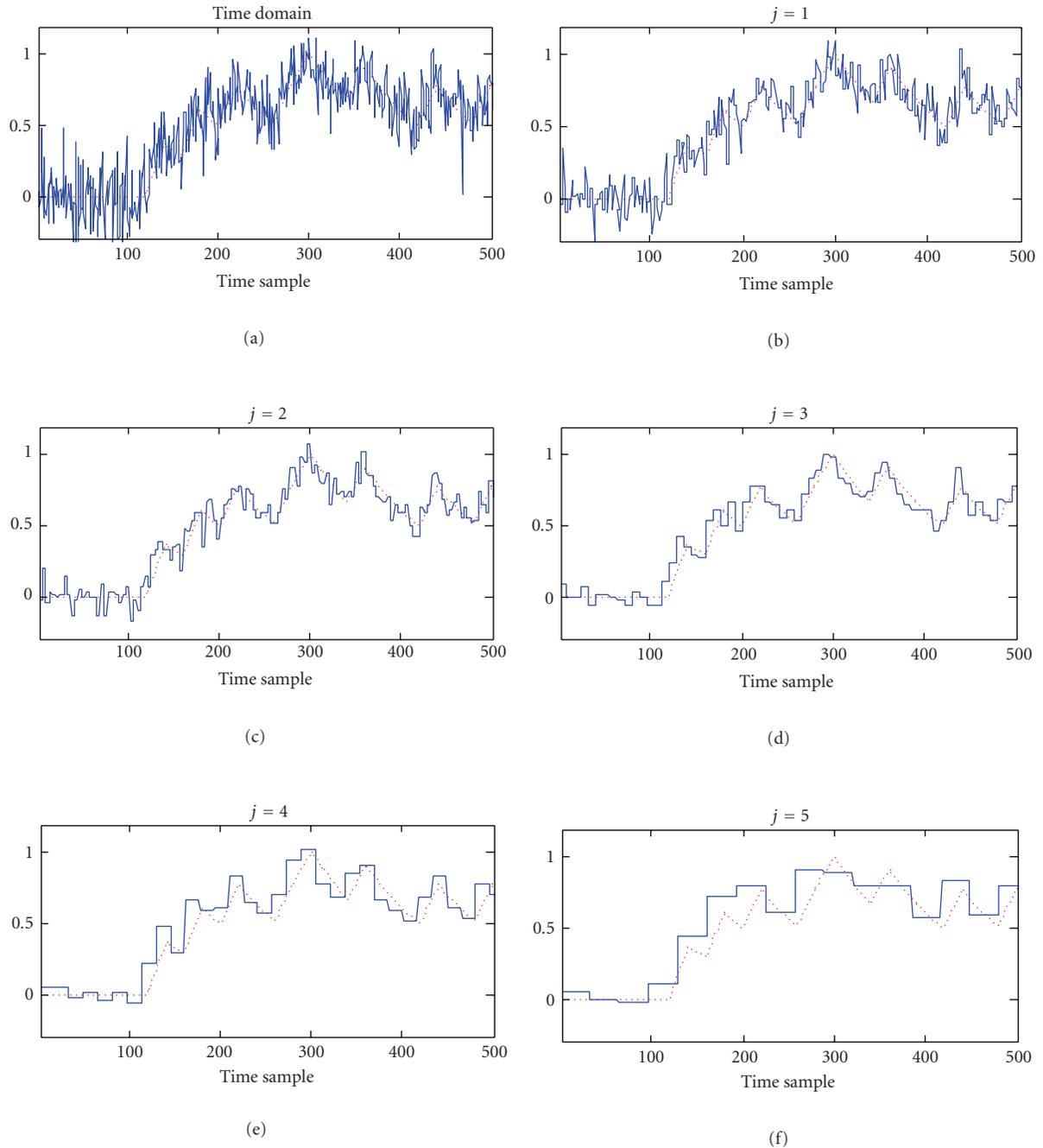


FIGURE 5: MSNL modeling performance at multiple scales, where $\sigma_\varepsilon = 0.15$.

through a simulated example, which clearly shows the improvement in output prediction, especially for larger noise contents.

Acknowledgment

The authors would like to gratefully acknowledge the financial support of Qatar National Research Fund (QNRF).

References

- [1] S. J. Qin and T. Badgwell, "An overview of industrial model predictive control technology," *Aiche Symposium Series*, vol. 93, no. 316, pp. 232–256, 1997.
- [2] R. D. Braatz and G. Mijares, *Control Relevant Identification and Estimation*, AICHE Meeting, Miami Beach, Fla, USA, 1995.
- [3] B. R. Bakshi, "Multiscale analysis and modeling using wavelets," *Journal of Chemometrics*, vol. 13, no. 3-4, pp. 415–434, 1999.

- [4] B. R. Bakshi and G. Stephanopoulos, "Representation of process trends-IV. Induction of real-time patterns from operating data for diagnosis and supervisory control," *Computers and Chemical Engineering*, vol. 18, no. 4, pp. 303–332, 1994.
- [5] S. Palavajhala, R. L. Motard, and B. Joseph, "Process identification using discrete wavelet transforms: design of prefilters," *AIChE Journal*, vol. 42, no. 3, pp. 777–790, 1996.
- [6] B. R. Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," *AIChE Journal*, vol. 44, no. 7, pp. 1596–1610, 1998.
- [7] A. N. Robertson, K. C. Park, and K. F. Alvin, "Extraction of impulse response data via wavelet transform for structural system identification," *Journal of Vibration and Acoustics*, vol. 120, no. 1, pp. 252–260, 1998.
- [8] M. Nikolaou and P. Vuthandam, "FIR model identification: parsimony through kernel compression with wavelets," *AIChE Journal*, vol. 44, no. 1, pp. 141–150, 1998.
- [9] M. N. Nounou, "Multiscale finite impulse response modeling," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 3, pp. 289–304, 2006.
- [10] J. F. Carrier and G. Stephanopoulos, "Wavelet-based modulation in control-relevant process identification," *AIChE Journal*, vol. 44, no. 2, pp. 341–360, 1998.
- [11] M. N. Nounou and H. N. Nounou, "Improving the prediction and parsimony of ARX models using multiscale estimation," *Applied Soft Computing Journal*, vol. 7, no. 3, pp. 711–721, 2007.
- [12] M. N. Nounou and H. N. Nounou, "Multiscale fuzzy system identification," *Journal of Process Control*, vol. 15, no. 7, pp. 763–770, 2005.
- [13] L. Ljung, *System Identification: Theory for Users*, Addison-Wesley, Reading, Mass, USA, 1987.
- [14] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [15] I. Daubechies, "Orthonormal bases for compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.
- [16] G. Strang, "Wavelets and dilation equations. A brief introduction," *SIAM Review*, vol. 31, no. 4, pp. 614–627, 1989.
- [17] M. N. Nounou and B. R. Bakshi, "On-line multiscale filtering of random and gross errors without process models," *AIChE Journal*, vol. 45, no. 5, pp. 1041–1058, 1999.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

