

## Research Article

# Multiscale Latent Variable Regression

Mohamed N. Nounou<sup>1</sup> and Hazem N. Nounou<sup>2</sup>

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

<sup>2</sup> 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 7 November 2009; Accepted 14 January 2010

Academic Editor: Ahmet N. Palazoglu

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.

Multiscale wavelet-based representation of data has been shown to be a powerful tool in feature extraction from practical process data. In this paper, this characteristic of multiscale representation is utilized to improve the prediction accuracy of some of the latent variable regression models, such as Principal Component Regression (PCR) and Partial Least Squares (PLS), by developing a multiscale latent variable regression (MSLVR) modeling algorithm. The idea is to decompose the input-output data at multiple scales using wavelet and scaling functions, construct multiple latent variable regression models at multiple scales using the scaled signal approximations of the data and then using cross-validation, and select among all MSLVR models the model which best describes the process. The main advantage of the MSLVR modeling algorithm is that it inherently accounts for the presence of measurement noise in the data by the application of the low-pass filters used in multiscale decomposition, which in turn improves the model robustness to measurement noise and enhances its prediction accuracy. The advantages of the developed MSLVR modeling algorithm are demonstrated using a simulated inferential model which predicts the distillate composition from measurements of some of the trays' temperatures.

## 1. Introduction

Process models are an essential part of many process operations, such as model-based control [1, 2]. However, constructing empirical models using measurements of the process variables is associated with many difficulties, which include dealing with collinearity or redundancy in the variables and accounting for the presence of measurement noise in the data.

Collinearity is common in models which involve large number of variables, such as Finite Impulse Response (FIR) models [3, 4] and inferential models. Collinearity increases the variance of the estimated model parameters, which degrades their accuracy of estimation. Many modeling techniques have been developed to deal with collinearity, which include Ridge Regression (RR) [5–7] and latent variable regression [3–5]. RR reduces the variations in model parameters by imposing a penalty on the norm of their estimated values. The latent variable regression models, on the other hand, use singular value decomposition to reduce the dimension of the input variables to provide a more

conditioned set of inputs. Some of the popular latent variable regression model estimation techniques include the well-known Principal Component Regression (PCR) and Partial Least Squares (PLS) modeling methods [3–5].

Also, the presence of measurement noise in the data used in empirical modeling, even in small amounts, can largely affect the estimated model's prediction accuracy. Therefore, measurement noise needs to be filtered for improved model prediction. Modeling of prefiltered data does not usually provide satisfactory modeling performance [8]. 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 improved model accuracy.

Unfortunately, measured data usually are multiscale in nature, which means that they contain features and noise that have varying contributions over both time and frequency [9]. 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 [8, 10–17]. For example, the authors in [10] showed how to use wavelet representation to design wavelet prefilters for process modeling purposes. In [9], the author discussed some of the advantages of using multiscale representation in empirical modeling, and in [11], he enhanced the noise removal ability of the Principal Component Analysis model by constructing multiscale PCA models, which he also used in process monitoring. Also, the authors in [12–14] used multiscale representation to reduce collinearity and shrink the large variations in FIR model parameters. Furthermore, the authors in [15, 16] used multiscale representation to enhance the prediction of fuzzy models and the parsimony and accuracy of ARX models. Finally, in [17], the authors used wavelets as modulating functions for control-related system identification.

In this work, multiscale representation of data is utilized to improve the prediction accuracy of some of the common latent variable regression modeling methods, such as PCR and PLS, by developing a multiscale latent variable regression (MSLVR) modeling algorithm that reduces the effect of measurement noise in the data on the accuracy and prediction of these models. The MSLVR algorithm integrates modeling and data filtering by constructing multiple latent variable regression 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 formulation and estimation of latent variable regression models are introduced, followed by a description of the wavelet-based multiscale representation of data in Section 3. Then, in Section 4, the representation and algorithm of MSLVR modeling are presented. Then, in Section 5, the performance of the developed MSLVR modeling algorithm is illustrated and compared to time-domain models through a simulated distillation column example. Finally, the paper is concluded with few remarks in Section 6.

## 2. Latent Variable Model Representation and Estimation

Given measurements of the input and output data, that is,  $\{x_k\}_{k=1,\dots,n}$  where  $x_k = [x_{1k} \cdots x_{ik} \cdots x_{mk}]$  and  $\{y_i\}_{i=1,\dots,n}$ , where all variables are assumed to be contaminated with additive zero mean Gaussian noise (i.e.,  $x = \tilde{x} + \varepsilon_x$  and  $y = \tilde{y} + \varepsilon_y$ , and the superscript “ $\sim$ ” represents the

noise-free variables), it is desired to construct latent variable regression models of the following form:

$$y_k = z_k \beta, \quad (1)$$

where

$$z_k = x_k \alpha^T. \quad (2)$$

In (1) and (2),  $z_k$ , the latent variables vector at time step ( $k$ ),  $\beta$ , the latent variable model parameter vector, and  $\alpha$ , the projection directions matrix are of sizes  $(1 \times p)$ ,  $(p \times 1)$ , and  $(p \times m)$ , respectively, and  $\alpha^T \alpha = I$ .

Note that latent variable regression models reduce the dimension of the input variables from ( $m$ ), which is the length of  $x_k$ , to ( $p$ ), which is the length of  $z_k$ , where  $m > p$ , and the model output is regressed to the latent variables instead of the original input variables.

$$\text{Defining: } Y = \begin{bmatrix} y_1 \\ \cdot \\ y_k \\ \cdot \\ y_n \end{bmatrix}_{n \times 1}, \quad X = \begin{bmatrix} x_1 \\ \cdot \\ x_k \\ \cdot \\ x_n \end{bmatrix}_{n \times m}, \quad Z = \begin{bmatrix} z_1 \\ \cdot \\ z_k \\ \cdot \\ z_n \end{bmatrix}_{n \times p}, \quad (3)$$

the latent variable regression model can be written in matrix form as follows:

$$Y = Z \beta, \quad (4)$$

where

$$Z = X \alpha^T. \quad (5)$$

Common methods of estimating the above latent variable regression model include PCR and PLS, which are described below.

**2.1. Principal Component Regression (PCR).** PCR accounts for collinearity in the input variables by reducing the dimension of these variables using Principal Component Analysis (PCA), which uses Singular Value Decomposition (SVD) to compute the latent variables or principal components,  $z_k$ . Then, it constructs a simple linear model between the latent variables and the output using the well-known Ordinary Least Squares (OLS) regression method [5, 18]. Therefore, PCR can be formulated as two consecutive estimation problems

(I)

$$\{\hat{\alpha}, \hat{z}_i\}_{\text{PCA}} = \arg \min_{\hat{\alpha}, \hat{z}_i} \sum_{i=1}^n (x_i - \hat{x}_i)^T (x_i - \hat{x}_i), \quad (6)$$

$$\text{s.t. } \hat{x}_i = \hat{\alpha} \hat{z}_i, \quad \hat{\alpha}^T \hat{\alpha} = I$$

(II)

$$\{\hat{\beta}\}_{\text{OLS}} = \arg \min_{\hat{\beta}} \sum_{i=1}^n (y_i - \hat{y}_i)^T (y_i - \hat{y}_i), \quad \text{s.t. } \hat{y}_i = \hat{z}_i^T \hat{\beta}. \quad (7)$$

**2.2. Partial Least Squares (PLS) Regression.** PLS regression uses the same model structure used by PCR but extends PCR to consider the output variables in computing the latent variables or principal components. It determines the projection directions that capture the variations in the input variables which are the closest to the output by maximizing the following objective function [19]:

$$\{\hat{\alpha}\}_{\text{PLS}} = \arg \max_{\hat{\alpha}} \{\text{corr}^2(Y, X\hat{\alpha})\text{var}(X\hat{\alpha})\}. \quad (8)$$

A similar formulation of PLS has also been used to extend PLS to deal with nonlinear problems, where the projection directions are estimated by minimizing the sum of input and output errors as follows [20]:

$$\{\hat{\alpha}\}_{\text{PLS}} = \arg \min_{\hat{\alpha}} \left\{ \sum_{i=1}^n (x_i - \hat{x}_i)^T (x_i - \hat{x}_i) + \sum_{i=1}^n (y_i - \hat{y}_i)^T (y_i - \hat{y}_i) \right\}, \quad (9)$$

subject to the constraints shown in (6) and (7).

It can be seen from the formulation of both PCR and PLS that they partially reduce the effect of measurement noise as they reduce the redundancy among the input variables. However, improvement can be made if the noise content within each variable is also reduced, which can be achieved using multiscale representation of data.

### 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 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

$$\phi_{jk}(t) = \sqrt{2^{-j}}\phi(2^{-j}t - k) \quad (10)$$

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. 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

$$\psi_{jk}(t) = \sqrt{2^{-j}}\psi(2^{-j}t - k) \quad (11)$$

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

$$x(t) = \sum_{k=1}^{n2^{-j}} a_{jk}\phi_{jk}(t) + \sum_{j=1}^J \sum_{k=1}^{n2^{-j}} d_{jk}\psi_{jk}(t), \quad (12)$$

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 [21, 22].

Fast wavelet transform algorithms of  $O(n)$  complexity for a discrete signal of dyadic length have been developed [23]. 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,

$$a_j = H_j a_{j-1}, \quad d_j = G_j a_{j-1}, \quad (13)$$

where

$$H_j = \begin{bmatrix} h_1 & \cdot & h_r & 0 & 0 \\ 0 & h_1 & \cdot & h_r & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & h_1 & \cdot & h_r \end{bmatrix}_{n2^{-j} \times n2^{-j+1}}, \quad (14)$$

$$G_j = \begin{bmatrix} g_1 & \cdot & g_r & 0 & 0 \\ 0 & g_1 & \cdot & g_r & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & g_1 & \cdot & g_r \end{bmatrix}_{n2^{-j} \times n2^{-j+1}}.$$

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.,  $j = 0$ ),

$$Y_o = [y_o(1) \ y_o(2) \ \cdot \ y_o(k) \ \cdot \ y_o(n)]^T, \quad (15)$$

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

$$Y_j = [y_j(1) \ \cdot \ y_j(k) \ \cdot \ y_j(n2^{-j})]^T, \quad (16)$$

can be computed as follows:

$$Y_j = H_j Y_{j-1} = H_j H_{j-1} \dots H_1 Y_o. \quad (17)$$

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 [24].

#### 4. Multiscale Latent Variable Regression (MSLVR)

In this section, the feature extraction abilities of multiscale data representation are utilized to develop multiscale latent variable regression models, which 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 latent variable regression model at each scale using the scaled signal approximations of the data. Then, among all scales, select the optimum latent variable regression model that provides the best noise-feature separation and least prediction error.

*4.1. Multiscale latent variable Regression Model Formulation.* Denoting the model input and output matrices in the time domain as  $X_o$  and  $Y_o$ , the scaled signal approximations of these matrices at the first scale can be computed using the low-pass filter matrices as shown in (17) as follows:

$$\begin{aligned} X_1 &= H_1 X_o, \\ Y_1 &= H_1 Y_o. \end{aligned} \quad (18)$$

Note that  $X_1$  and  $Y_1$  are of sizes  $(n/2) \times m$  and  $(n/2) \times 1$ , respectively because of the downsampling used in wavelet decomposition. Having computed the matrices,  $X_1$  and  $Y_1$ , the latent variable regression model at the first scale, which can be estimated using either PCR or PLS, can be represented as follows:

$$Y_1 = Z_1 \beta_1, \quad (19)$$

where

$$Z_1 = X_1 \alpha_1^T. \quad (20)$$

Repeating this process at coarser scales, the latent variable regression model at scale ( $j$ ) can be expressed as follows:

$$Y_j = Z_j \beta_j, \quad (21)$$

where,

$$Z_j = X_j \alpha_j^T, \quad (22)$$

$$X_j = H_j H_{j-1} \cdots H_1 X_o, \quad (23)$$

$$Y_j = H_j H_{j-1} \cdots H_1 Y_o. \quad (24)$$

*4.2. Multiscale Latent Variable Regression (MSLVR) Modeling Algorithm.* Based on the above discussion, the following algorithm is proposed for multiscale latent variable regression.

- (1) Given the input-output data, construct the matrices  $X$  and  $Y$  and estimate a latent variable regression model either using PCR or PLS as described in Section 2. This will require estimating the rank of the latent variable regression model, which can be done using cross-validation [25].
- (2) Decompose the input and output data at coarser scales as shown in (23) and (24).
- (3) At each scale ( $j$ ) and using the scaled signal approximations of the data,
  - (a) estimate either a PCR or PLS model as described in step (1), and use it to predict the output at that scale;
  - (b) reconstruct the predicted output of either the PCR or PLS model back to the time domain;
  - (c) use the reconstructed output prediction from each scale to compute the following cross-validation mean squares errors (CVMSE) at each scale [25]

$$\text{CVMSE}(j) = \frac{1}{n} \sum_{k=1}^n \left[ (y_k^* - \hat{y}_k)^2 \right], \quad (25)$$

where

$$y_k^* = \frac{1}{2} [y_{k-1} + y_{k+1}]; \quad (26)$$

- (4) select among all scales the optimum multiscale latent variable regression model that minimizes the cross-validation mean squares error criterion shown in (25).

### 5. Illustrative Example

In this section, the performance of the developed MSLVR modeling algorithm is illustrated and compared to those of the time-domain latent variable regression techniques and to modeling prefiltered data using an Exponentially Weighted Moving Average (EWMA) filter. The MSLVR models are used as inferential models that estimate the distillation column composition using temperature measurements. The data are simulated using a 30-tray distillation column that is used to separate methanol, ethanol, 1-propanol, and 1-butanol under temperature control. The objective of this distillation process is to maintain high-purity separation of the light and heavy components. A more detailed description of the process and its operating conditions is provided in [26].

For control purposes, measuring compositions on-line is very expensive, and thus it is desired to build accurate inferential models that estimate the product composition in the distillate and bottom streams. In this example, an inferential model is constructed to estimate the composition of ethanol in the bottom stream from temperature measurements at different trays. As shown in [26], the distillate composition can be estimated using temperature measurements from nine trays, which are  $T_3, T_6, T_9, T_{12}, T_{15}, T_{19}, T_{22}, T_{25}$ , and  $T_{28}$ .

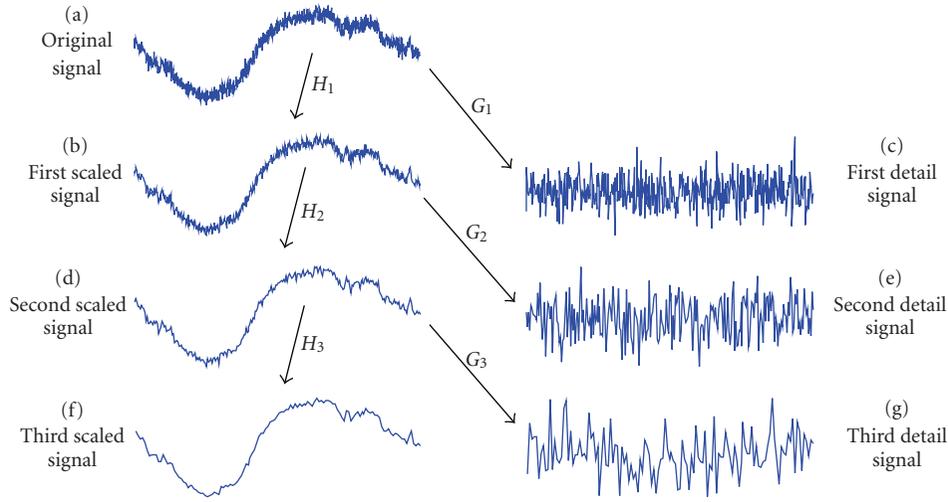


FIGURE 1: A schematic diagram of data representation at multiple scales.

The simulated data, which consist of 1024 samples, are assumed to be noise-free. Then, all variables, inputs, and outputs are contaminated with zero mean Gaussian noise. Different levels of noise, that is, signal-to-noise ratios (SNR) of 10 and 50, are used to test the robustness of the developed MSLVR modeling algorithm.

To show the effect of prefiltering of the model prediction accuracy, the data are filtered using an EWMA filter, and then the filtered data are used to construct the LVR models. The EWMA filter has the following structure:

$$\hat{w}(k) = \alpha w(k) + (1 - \alpha)\hat{w}(k - 1), \quad (27)$$

where  $w$  is a raw data sample,  $\hat{w}$  is a filtered data sample, and  $\alpha$  is a filter parameter that is optimized using cross-validation. Of course, for different level of noise, the optimum value of  $\alpha$  changes. For example, for a SNR of 10,  $\alpha$  is found to be 0.41, while for a SNR of 50, it is found to be 0.52.

The performances of the multiscale and time-domain latent variable regression methods are compared by comparing the prediction mean square errors with respect to the noise-free output, that is,

$$\text{MSE} = \frac{1}{n} \sum_{k=1}^n (\hat{y}_k - \tilde{y}_k)^2, \quad (28)$$

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.

To make statistically valid conclusions about the performances of the various modeling methods, a Monte Carlo simulation of 100 realizations is performed, and the results are presented in Tables 1 and 2.

Table 1 shows that PLS outperformed PCR in the time domain. It also shows that for both PCR and PLS, prefiltering

TABLE 1: Comparison between the output prediction mean square errors for the various modeling methods.

Modeling method	MSE ( $\times 10^7$ )	MSE ( $\times 10^8$ )
	SNR = 10	SNR = 50
PCR	0.907	0.533
EWMA + PCR	0.787	0.482
MS PCR	0.484	0.403
PLS	0.822	0.487
EWMA + PLS	0.711	0.469
MS PLS	0.423	0.365

the data using an EWMA filter helps the prediction of the estimated models. However, the multiscale variable regression modeling techniques (MS PCR and MS PLS) provide improved prediction results than their time domain counterparts and modeling prefiltered data, and the level of improvement increases for larger noise contents (smaller SNR). This improvement is illustrated in Figures 2 and 3, which show the advantages of constructing latent variable regression models at multiple scales.

Figures 2 and 3 also show that the accuracy of the estimated MSLVR models improves at coarser scales but up to a certain scale beyond which the quality of estimated models deteriorates. This can also be noted from Table 2, which reports the MSE of the estimated PCR and PLS models at different scales and shows that there is an intermediate scale at which MSLVR is the best. This is because at very coarse scales, important features get eliminated which affects 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 cross-validation mean squares error criterion (shown in (25)) and shows that the optimum scale increases or gets coarser (higher  $j$ ) for higher noise levels or smaller SNR. This

TABLE 2: Comparison between the output prediction mean square errors obtained by MSLVR at multiple scales (the numbers in parenthesis indicate the percentages that each scale is selected as optimum using the CVMSE criterion).

Scale	PCR MSE( $\times 10^7$ ) SNR = 10	PLS MSE( $\times 10^7$ ) SNR = 10	PCR MSE( $\times 10^8$ ) SNR = 50	PLS MSE( $\times 10^8$ ) SNR = 50
$j = 0$	0.907 (0)	0.822 (0)	0.533 (0)	0.487 (0)
$j = 1$	0.705 (0)	0.643 (0)	0.484 (0)	0.436 (0)
$j = 2$	0.595 (0)	0.537 (0)	0.449 (0)	0.404 (0)
$j = 3$	0.523 (3)	0.463 (1)	0.416 (11)	0.373 (13)
$j = 4$	0.479 (97)	0.422 (99)	0.401 (89)	0.364 (87)
$j = 5$	0.636 (0)	0.584 (0)	0.572 (0)	0.541 (0)

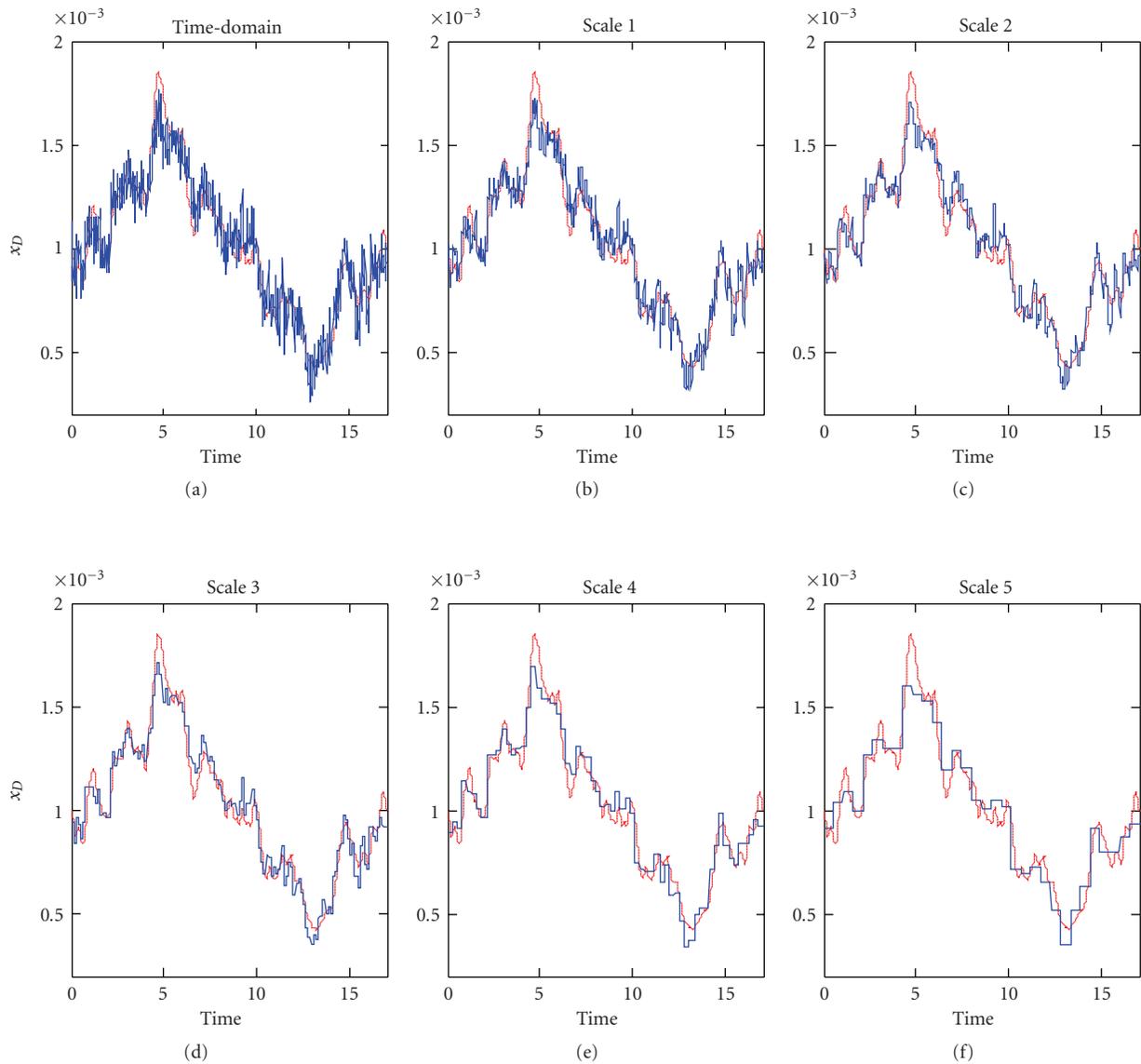


FIGURE 2: The performance of MS PCR modeling at multiple scales for the case where the data SNR = 10 [dashed line: noise-free output; solid line: predicted output].

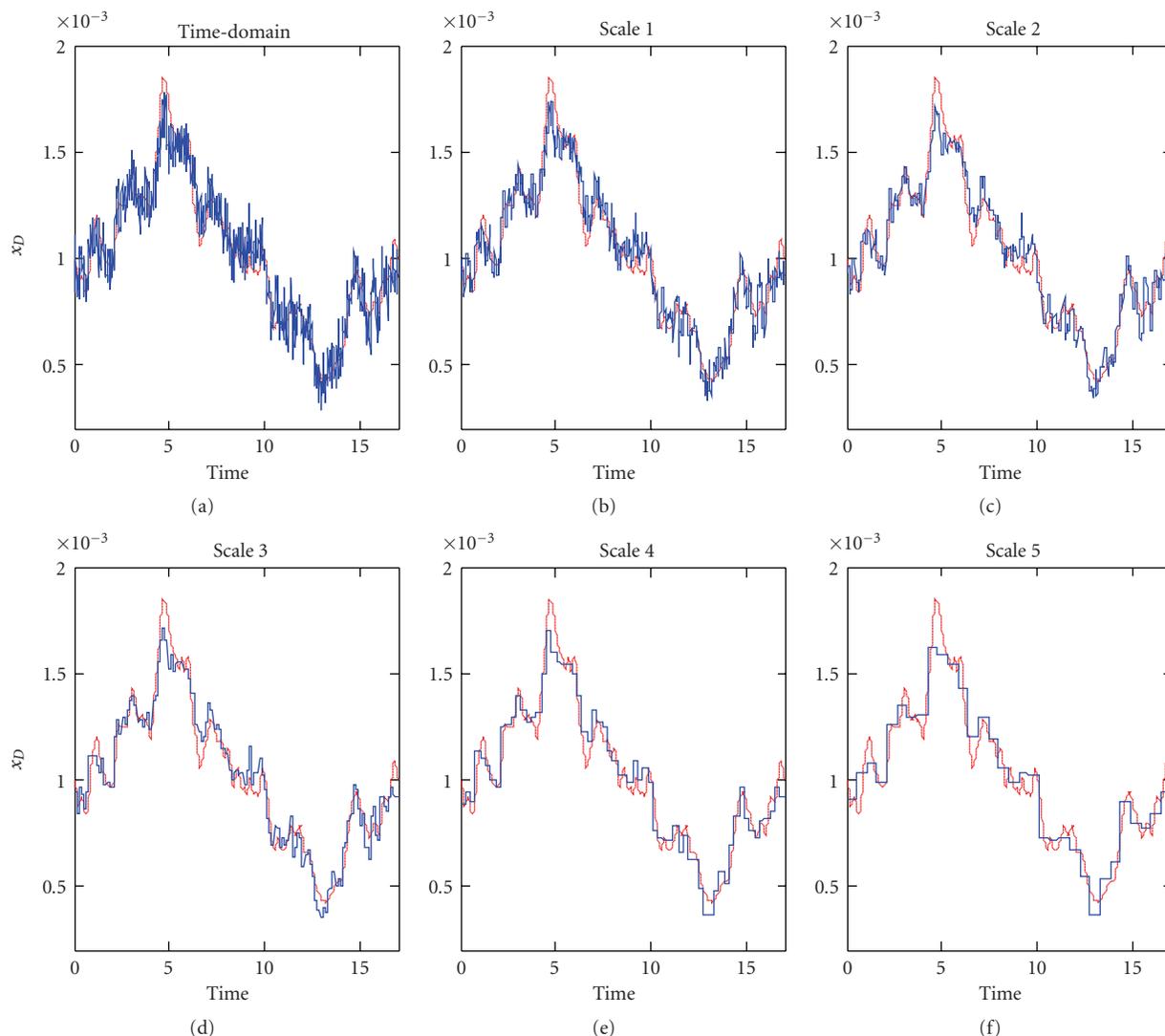


FIGURE 3: The performance of MS PLS modeling at multiple scales for the case where the data SNR = 10 (dashed line: noise-free output and solid line: predicted output).

makes sense because, for higher noise levels, more filtering is needed for good noise-feature separation.

## 6. Conclusions

In this paper, the noise-feature separation capabilities of multiscale representation of data are exploited to improve the prediction accuracy of the latent variable regression models, by presenting a multiscale latent variable regression (MSLVR) modeling algorithm that enhances the robustness of the latent variable regression models to measurement noise in the data. The MSLVR algorithm integrates modeling and filtering by decomposing the input-output data and using the scaled signal approximations to construct different latent variable regression models at different scales. Then, among all scales, the model that minimizes a cross-validation mean squares error criterion is selected as the optimum model. The estimated models using the developed MSLVR

modeling algorithm are shown to outperform their time domain counterparts and modeling prefiltered data through a simulated example, which clearly shows the advantages of integrating filtering and model estimation on the prediction of the estimated models.

## Acknowledgment

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

## References

- [1] S. Qin and T. Badgwell, "An overview of industrial model predictive control technology," in *Proceedings of the 5th International Conference on Chemical Process Control (CPC '97)*, C. Garcia and J. Kantor, Eds., vol. 93 of *AICHE Symposium Series 316*, pp. 232–256, 1997.

- [2] R. Braatz and G. Mijares, "Control relevant identification and estimation," in *AIChE Annual Meeting*, p. 183a, Miami Beach, Fla, USA, 1995.
- [3] B. M. Wise and N. L. Ricker, "Identification of finite impulse response models by principal components regression: frequency-response properties," *Process Control and Quality*, vol. 4, no. 1, pp. 77–86, 1992.
- [4] N. L. Ricker, "The use of biased least-squares estimators for parameters in discrete-time pulse-response models," *Industrial and Engineering Chemistry Research*, vol. 27, no. 2, pp. 343–350, 1988.
- [5] I. Frank and J. Friedman, "A statistical view of some chemometric regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–148, 1993.
- [6] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 42, no. 1, pp. 80–86, 2000.
- [7] J. McGregor, T. Kourti, and J. Kresta, "Multivariate identification: a study of several methods," in *IFAC ADCHEM Conference Proceedings*, pp. 369–375, Toulouse, France, 1991.
- [8] B. R. Bakshi, "Multiscale analysis and modeling using wavelets," *Journal of Chemometrics*, vol. 13, no. 3–4, pp. 415–434, 1999.
- [9] 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.
- [10] 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.
- [11] B. R. Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," *AIChE Journal*, vol. 44, no. 7, pp. 1596–1610, 1998.
- [12] 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.
- [13] M. Nikolaou and P. Vuthandam, "FIR model identification: parsimony through Kernel compression with wavelets," *AIChE Journal*, vol. 44, no. 1, pp. 141–150, 1998.
- [14] M. N. Nounou, "Dealing with collinearity in FIR models using multiscale estimation," in *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference (CDC-ECC '05)*, pp. 8162–8167, Seville, Spain, December 2005.
- [15] M. N. Nounou, "Multiscale ARX process modeling," in *Proceedings of the 45th IEEE Conference on Decision and Control (CDC '06)*, pp. 823–828, San Diego, Calif, USA, December 2006.
- [16] M. N. Nounou and H. N. Nounou, "Enhanced prediction accuracy of fuzzy models using multiscale estimation," in *Proceedings of the 43rd IEEE Conference on Decision and Control (CDC '04)*, vol. 5, pp. 5170–5175, Atlantis, Bahamas, December 2004.
- [17] J. F. Carrier and G. Stephanopoulos, "Wavelet-based modulation in control-relevant process identification," *AIChE Journal*, vol. 44, no. 2, pp. 341–360, 1998.
- [18] W. Massy, "Principal components regression in exploratory statistical research," *Journal of the American Statistical Association*, vol. 60, pp. 234–246, 1965.
- [19] S. Wold, "Soft modeling. The basic design and some extensions," in *Systems under Indirect Observations*, K. Joreskog and H. Wold, Eds., Elsevier, Amsterdam, The Netherlands, 1982.
- [20] E. C. Malthouse, A. C. Tamhane, and R. S. H. Mah, "Nonlinear partial least squares," *Computers and Chemical Engineering*, vol. 21, no. 8, pp. 875–890, 1997.
- [21] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.
- [22] G. Strang, "Wavelets and dilation equations: a brief introduction," *SIAM Review*, vol. 31, no. 4, pp. 614–627, 1989.
- [23] S. G. Mallat, "A theory of multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 764–693, 1989.
- [24] 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.
- [25] G. Nason, "Wavelet shrinkage using cross-validation," *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 2, pp. 463–479, 1996.
- [26] M. Kano, K. Miyazaki, S. Hasebe, and I. Hashimoto, "Inferential control system of distillation compositions using dynamic partial least squares regression," *Journal of Process Control*, vol. 10, no. 2, pp. 157–166, 2000.



**Hindawi**

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

