

# A multiscale, Bayesian and error-in-variables approach for linear dynamic data rectification

Sridhar Ungarala <sup>a,\*</sup>, Bhavik R. Bakshi <sup>b</sup>

<sup>a</sup> Department of Chemical Engineering, Cleveland State University, Cleveland, OH 44115, USA

<sup>b</sup> Department of Chemical Engineering, The Ohio State University, Columbus, OH 43210, USA

## Abstract

A multiscale approach to data rectification is proposed for data containing features with different time and frequency localization. Noisy data are decomposed into contributions at multiple scales and a Bayesian optimization problem is solved to rectify the wavelet coefficients at each scale. A linear dynamic model is used to constrain the optimization problem, which facilitates an error-in-variables (EIV) formulation and reconciles all measured variables. Time-scale recursive algorithms are obtained by propagating the prior with temporal and scale models. The multiscale Kalman filter is a special case of the proposed Bayesian EIV approach. © 2000 Elsevier Science Ltd. All rights reserved.

*Keywords:* Rectification; Bayesian; Wavelets; Error-in-variables; Kalman filter

## 1. Introduction

Most physical systems exhibit inherent multiscale features, where the physical phenomena and the associated measurements may contain contributions at multiple scales in time and frequency. Traditional data rectification methods either reduce the natural multiscale characteristics of the problem to a single scale or fail to utilize the rich variety of information available across all the scales. Adaptive estimation techniques based on multiscale models have been shown to perform better than their single scale counterparts, for example, multiscale principal component analysis (Bakshi, 1998) and multiscale model predictive control (Stephanopoulos, Karsligil & Dyer, 1998). Recently, multiscale methods have been developed for removing errors from measured data based on the ability of wavelets to extract deterministic features and to approximately decorrelate stochastic processes. Among the most popular multiscale filtering methods is wavelet thresholding (Donoho, Johnstone, Kerkycharian & Picard, 1995). A multiscale approach for removing errors from fractal signals has been developed based on the ability of wavelets to decorrelate fractal processes (Wornell & Oppenheim, 1992; Miller & Willsky, 1995).

A similar approach has been used for the development of multiscale Kalman filter banks (Chou, Willsky & Benveniste, 1994; Hirchoren & D'Attellis, 1999; Stephanopoulos et al., 1998).

Similar to their single scale counterparts, the multiscale filtering methods such as the multiscale Kalman filter are fundamentally Bayesian in nature. The recursive prior update is implemented with two-scale models, which are derived from the wavelet transform (Chou et al., 1994) or a combination of the wavelet transform and the single scale temporal models (Stephanopoulos et al., 1998). Existing multiscale filtering approaches do not utilize the multivariate relationship given by the models to accomplish data reconciliation. Bakshi, Nounou, Goel and Shen (1999) developed a multiscale Bayesian approach with empirical prior for data reconciliation with steady-state models. Multiscale methods have so far focused on linear systems and have implicitly assumed that the error statistics remain the same at all scales (Chou et al., 1994). Such approaches are best suited for stochastic underlying signals. Scale dependent probability distributions are required for many common processes such as autocorrelated noise (Bakshi et al., 1999).

This paper presents a multiscale Bayesian approach to data rectification and exploits the multivariate relationships given by a linear dynamic model. A frame-

\* Corresponding author.

work for multiscale moving horizon algorithm is developed which incorporates error-in-variables (EIV) data rectification. The method represents the measurements for each variable in the wavelet domain, and rectifies it based on the probability distribution of the coefficients and errors at each scale while satisfying a process model. The resulting solution provides data reconciliation as opposed to filtering by existing methods (Chou et al., 1994; Stephanopoulos et al., 1998). The primary aim is to rectify the wavelet coefficients, which capture all the stochastic information in the signal. The constrained optimization problem at each scale is solved sequentially from fine to coarse scales. The proposed method supports adaptive, scale varying prior probability distributions by using multiscale linear models similar to those derived by Stephanopoulos et al. (1998). Besides accounting for the multiscale features, such models allow highly parallelizable and computationally efficient algorithms. Existing methods assume scale independent noise processes (Chou et al., 1994; Stephanopoulos et al., 1998). This assumption is relaxed in this work so that the approach can be used for both single scale data as well as multiscale or multirate data. The benefits of the proposed approach are illustrated with a simulation example.

## 2. Multiscale Bayesian data rectification

A family of dyadically discretized wavelet basis functions is represented as,

$$\psi_{mk}(t) = 2^{-m/2}\psi(2^{-m}t - k) \quad (1)$$

where  $\psi(t)$  is the mother wavelet and  $m$  and  $k$  are dilation and translation parameters, respectively. Any signal can be decomposed into its contributions at multiple scales as,

$$y(t) = \sum_{m=1}^M \sum_{k=1}^N d_y(m,k)\psi_{mk}(t) + \sum_{k=1}^N a_y(L,k)\phi_{mk}(t) \quad (2)$$

where  $d_y$  are the wavelet coefficients or detail signal at scale  $m$  and  $a_y$  are the scaled signal coefficients at the coarsest scale  $M$ . The matrix of coefficients at a selected scale may be represented as a product of the corresponding filter matrix with the data matrix as  $d_y(m) = G_m Y$  and  $a_y(L) = H_L Y$ .

Consider the following measurement equation, where  $y$  is the measurement vector of process variables  $x$ , with an additive term  $v$ , representing the errors in measurements,

$$y = Cx + v \quad (3)$$

It is common to assume that the measurement errors are Gaussian,  $v \sim N(0,R)$ . Wavelet decomposition of data corrupted with Gaussian errors preserves the Gaussian distribution for errors at all scales. If the

underlying signal is stochastic, the noise-free wavelet coefficients at each scale tend towards a Gaussian distribution for many common stochastic processes that are non-Gaussian in the time domain. The wavelet coefficients are approximately white with the variance at each scale changing according to the power spectrum of the time domain signal in the corresponding range of frequencies (Beylkin, Coifman & Rokhlin, 1991).

The Bayesian approach to data rectification maximizes the posterior probability distribution of the process variables  $p(x|y)$ . The objective is to propagate the a priori distribution conditioned on the measurements to obtain the a posteriori distribution. Assuming that the errors as well as the noise-free wavelet coefficients behave as Gaussian random variables, multiscale Bayesian data rectification may be written as the following optimization problem subject to a model and applicable constraints at each scale  $m$ ,

$$\min_{\hat{d}_x} \{ [d_y(m) - d_x(m)]^T R^{-1}(m) [d_y(m) - d_x(m)] + [d_x(m) - \mu_{d_x}(m)]^T P_{d_x}^{-1}(m) [d_x(m) - \mu_{d_x}(m)] \} \quad (4)$$

where  $d_y(m)$  and  $d_x(m)$  are the wavelet coefficients of the measurements and noise-free variables, respectively,  $R(m)$  is the covariance of the errors and  $\mu_{d_x}(m)$  and  $P_{d_x}(m)$  are the parameters of the Gaussian prior. The same objective function is valid for the scaled signal coefficient  $a_y(L)$ , if the distributions are Gaussian. Without any inequality constraints, let us assume that a linear process model of the form,

$$F(m)d_x(m) = b(m) \quad (5)$$

is available for both the wavelet and scaled coefficients, which is imposed as algebraic constraints in the optimization. The rectified coefficients can be obtained by an efficient closed form solution to the optimization problem (Ungarala & Bakshi, 2000a).

$$\hat{d}_x(m) = K[C^T R^{-1}(m)d_y(m) + P_{d_x}^{-1}(m)\mu_{d_x}(m)] + Lb \quad (6)$$

where

$$K = [I - JF^T(FJF^T)^{-1}F]J \quad (7)$$

$$J = [C^T R^{-1}(m)C + P_{d_x}^{-1}(m)]^{-1} \quad (8)$$

$$L = JF^T(FJF^T)^{-1} \quad (9)$$

The error covariance matrix associated with the rectified variables is computed by,

$$P_{\hat{d}_x}(m) = \text{cov}[d_x(m) - \hat{d}_x(m)] = K(m) \quad (10)$$

If a model is available to represent the dynamics of the process in scale, it can be used to propagate the prior parameters  $\mu_{d_x}$  and  $P_{d_x}$  for a scale recursive rectification algorithm. Furthermore, dynamic models can be used as equality constraints and an EIV formulation can be implemented to rectify both input and output variables simultaneously (Ungarala & Bakshi,

2000a). The multiscale Bayesian data rectification scheme is depicted in Fig. 1, where single scale data is decomposed over multiple scales and the constrained optimization problem is solved at each scale on both the scaled signal and wavelet coefficients and the prior is propagated from scale to scale. If the distribution of the errors or the prior at any scale is represented as a non-Gaussian distribution, the solution requires nonlinear or quadratic optimization.

### 3. Multiscale models

Let the time evolution of a linear system be represented at the finest scale ( $m = 0$ ) by

$$x(0, k + 1) = Ax(0, k) + Bu(0, k) + w(0, k) \quad (11)$$

The Haar wavelet decomposition of a data sequence  $\{x_k\}$  for  $k = 0, \dots, N - 1$  at scale  $m = 0$  generates a set of  $N/2$  scaled coefficients and  $N/2$  wavelet coefficients at a coarser scale  $m = 1$ . The coefficients can be indexed on a dyadic tree to represent the time-scale evolution of the process. Chou et al. (1994) used the wavelet transform to derive linear dynamic models in scale for

fractal processes and Stephanopoulos et al. (1998) combined the Haar wavelet transform with the finest scale temporal model (Eq. (11)) to derive scale models on binary trees.

Without loss of generality, assume that the process is noise-free ( $w = 0$ ) and the input  $u$  is stochastic with known statistical properties,  $u \sim N(0, Q)$ . The time dynamics of the scaled signal at scale  $m$  can be written as,

$$\begin{aligned} x\left(m, \frac{k}{2^m} + 1\right) \\ = A^{2^m} x\left(m, \frac{k}{2^m}\right) + \sum_{j=0}^{2^m-1} A^j B u\left(m, \frac{k}{2^m} + \frac{2^m-1-j}{2^m}\right) \end{aligned} \quad (12)$$

Note that the inputs to this linear system are not just the nodes of a binary tree generated by downsampling. The system dynamics dictate that the decimated input nodes in between two output nodes separated by  $2^m \Delta t$  also contribute to the evolution in time. Stephanopoulos et al. (1998) used a modified hat function for the decomposition of the inputs and merged the second term on the right in Eq. (12) into a single node.

It is convenient to introduce a notation of shift operators on dyadic trees (Chou et al., 1994). Consider a parent node and its two children nodes on a binary tree (Fig. 2). The parent node  $\tau$  communicates with the left child node  $\tau\alpha$  and the right child node  $\tau\beta$  with the downward shift operators  $\alpha$  and  $\beta$ , respectively. In our case, a new lateral shift operator  $\delta$  is defined to relate the adjacent child nodes and passes through the nodes decimated by downsampling. The scale dynamics of the process can be represented as the relationship between the children nodes at a finer scale and the parent node at a coarser scale (Ungarala & Bakshi, 2000b). The wavelet coefficients of nodes at scale  $m + 1$  can be expressed as a function of the scaled signal coefficients of nodes at scale  $m$ . The wavelet coefficients can also be obtained from the scaled signal coefficients at the same scale and input from a finer scale.

#### 3.1. EIV formulation

Consider the decomposition of the data sequences  $\{x_{k+1}\}$ , its lagged sequence  $\{x_k\}$  and the input sequence  $\{u_k\}$  for  $k = 0, \dots, N - 1$ . The overlapping binary trees are shown in Fig. 3. Define an augmented left node vector  $\mathbf{x}(\tau\alpha) = [x(\tau\alpha\delta), x(\tau\alpha), u(\tau\alpha)]^T$  and an augmented input vector  $\mathbf{u}(\tau\alpha) = [u(\tau\alpha\delta^0), \dots, u(\tau\alpha\delta^{2^m-1})]^T$ . From left child node at scale  $m$  to parent node at scale  $m + 1$ ,

$$\mathbf{x}(\tau) = \mathbf{A}_x^\alpha(m)\mathbf{x}(\tau\alpha) + \mathbf{B}_x^\alpha(m)\mathbf{u}(\tau\alpha) \quad (13)$$

where,

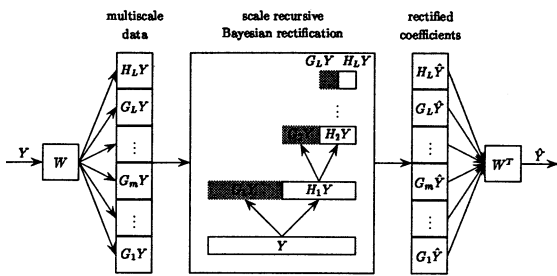


Fig. 1. Multiscale Bayesian data rectification.

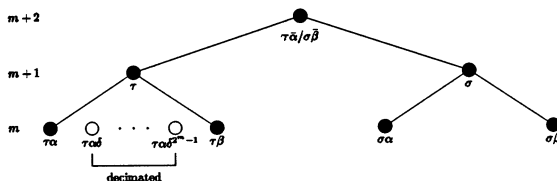


Fig. 2. Shift operator on trees.

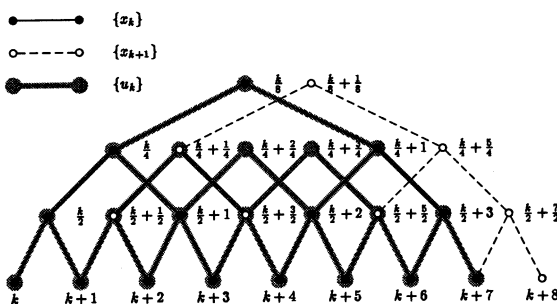


Fig. 3. Binary trees with and without downsampling.

$$\mathbf{A}_x^z(m) = h \begin{bmatrix} (A^{2^m} + I) & 0 & 0 \\ 0 & (A^{2^m} + I) & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{B}_x^z(m) = h \begin{bmatrix} 0 & A^{2^m-1}B & \cdots & AB & B \\ A^{2^m-1}B & A^{2^m-2}B & \cdots & B & 0 \\ I & 0 & \cdots & 0 & I \end{bmatrix}$$

where  $h = 1/\sqrt{2}$ . From right node to parent node,

$$\mathbf{x}(\tau) = \mathbf{A}_x^\beta(m)\mathbf{x}(\tau\beta) + \mathbf{B}_x^z(m)\mathbf{u}(\tau\alpha) \quad (14)$$

where,

$$\mathbf{A}_x^\beta(m) = \begin{bmatrix} (A^{2^m})^{-1} & 0 & 0 \\ 0 & (A^{2^m})^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{A}_x^z(m)$$

$$\mathbf{B}_x^\beta(m) = \begin{bmatrix} -(A^{2^m})^{-1} & 0 & 0 \\ 0 & -(A^{2^m})^{-1} & 0 \\ 0 & 0 & I \end{bmatrix} \mathbf{B}_x^z(m)$$

Wavelet coefficients from scaled signal coefficients,

$$\mathbf{d}(\tau) = \mathbf{A}_d(m)\mathbf{x}(\tau) + \mathbf{B}_d(m)\mathbf{u}(\tau\alpha) \quad (15)$$

where,

$$\mathbf{A}_d(m) = \begin{bmatrix} \frac{(A^{2^m} - I)}{(A^{2^m} + I)} & 0 & 0 \\ 0 & \frac{(A^{2^m} - I)}{(A^{2^m} + I)} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{B}_d(m) = 2h \begin{bmatrix} (A^{2^m} + I)^{-1} & 0 & 0 \\ 0 & (A^{2^m} + I)^{-1} & 0 \\ 0 & 0 & I \end{bmatrix}$$

$$\times \begin{bmatrix} 0 & A^{2^m-1}B & \cdots & AB & B \\ A^{2^m-1}B & A^{2^m-2}B & \cdots & B & 0 \\ -I & 0 & \cdots & 0 & I \end{bmatrix}$$

The augmented vector  $\mathbf{x}$ , satisfies the following algebraic relationship at all nodes, which is based on the temporal model at  $m = 0$ ,

$$\mathbf{F}\mathbf{x} = 0 \quad (16)$$

where  $\mathbf{F} = [I, -A, -B]$ . The wavelet coefficient  $\mathbf{d}$  also satisfies the same equality relationship at all scales.

The input to the two-scale models  $\mathbf{u}(\cdot)$ , is serially correlated in time at all scales except the finest, which is the result of not downsampling. The covariance matrix  $\mathbf{Q}_u$  can be computed using a Haar filter matrix without

downsampling (Ungarala & Bakshi, 2000b). Stephanopoulos et al. (1998) used a modified hat function for a similar task. It is generally assumed that the process evolves at all scales driven by scale independent inputs, that is,  $\mathbf{u}(m+1)$  is not correlated with  $\mathbf{u}(m)$  (Chou et al., 1994; Stephanopoulos et al., 1998). This assumption, natural to multiscale process noise sequences, is not valid in the present case because the input data at coarser scales are generated by decomposing the inputs at the finest scale. Furthermore,  $\mathbf{u}$  cannot be expressed as a linear system driven by white noise in scale. In the EIV formulation, we are interested in rectifying the noisy input measurements at all scales. Since the input vector doubles in length at every scale, state augmentation leads to very large system matrices. Hence,  $\mathbf{x}(\tau\alpha)$ , and  $\mathbf{u}(\tau\alpha)$  remain correlated at all scales except the finest.

$$\mathbf{Q}_{\mathbf{x}(\tau\alpha)\mathbf{u}(\tau\alpha)}(m) = \mathbf{E}\{\mathbf{x}(\tau\alpha)\mathbf{u}^T(\tau\alpha)\} \quad (17)$$

which can be calculated with a modified Haar filter without downsampling (Ungarala & Bakshi, 2000b). Similarly, in the scale model for wavelet equation, the correlation between  $\mathbf{x}(\tau)$  and  $\mathbf{u}(\tau\alpha)$  is obtained from,

$$\mathbf{Q}_{\mathbf{x}(\tau\alpha)\mathbf{u}(\tau\alpha)}(m) = \mathbf{A}_x^z(m)\mathbf{Q}_{\mathbf{x}(\tau\alpha)\mathbf{u}(\tau\alpha)}(m) + \mathbf{B}_x^z(m)\mathbf{Q}_u(m) \quad (18)$$

#### 4. Multiscale measurements

Measured data at the finest scale  $y(0,k)$  are decomposed into multiscale data  $y(m, k/2^m)$  with the Haar wavelet transform. The measurements of the input signal are also assumed to be corrupted with Gaussian noise  $v \sim N(0,R)$ ,

$$z_k = Du_k + v_k \quad (19)$$

which are decomposed into noisy inputs at coarser scales. For the EIV formulation, the augmented measurement vector is defined as  $\mathbf{y}(\tau\alpha) = [y(\tau\alpha\delta), y(\tau\alpha), z(\tau\alpha)]^T$ . Thus the measurement equations for scaled signal  $\mathbf{x}$  and wavelet coefficients  $\mathbf{d}$  are

$$\mathbf{y}_x(\tau\alpha) = C\mathbf{x}(\tau\alpha) + \mathbf{v}_x(m) \quad (20)$$

$$\mathbf{y}_d(\tau\alpha) = C\mathbf{d}(\tau\alpha) + \mathbf{v}_d(m) \quad (21)$$

In the case of truly multiscale processes, where independent data are available at all scales, the measurement noise  $\mathbf{v}_x(m)$  and  $\mathbf{v}_d(m)$  can be assumed to be white Gaussian processes in time as well as in scale (Chou et al., 1994; Stephanopoulos et al., 1998). When multiscale data are generated by wavelet decomposition of single scale data, it results in measurement errors that are correlated in scale. In the EIV formulation, it also gives rise to cross correlation at all scales coarser than the finest scale due to the presence of lagged quantities in  $\mathbf{y}$ .

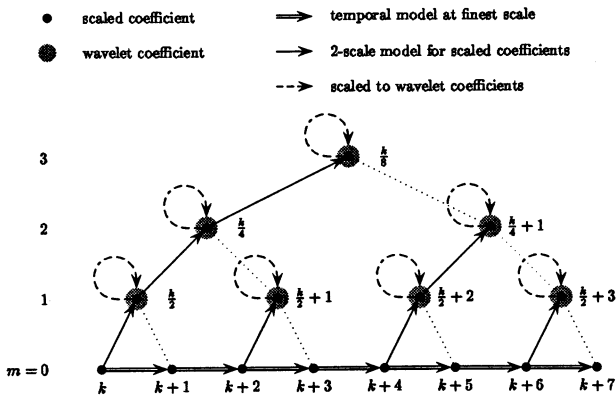


Fig. 4. Recursive rectification on multiscale trees.

$$\mathbf{R}(m) = \mathbf{E}\{\mathbf{v}\mathbf{v}^T\} = \begin{bmatrix} R & \frac{2^m - 1}{2^m} R & 0 \\ \frac{2^m - 1}{2^m} R & R & 0 \\ 0 & 0 & R \end{bmatrix} \quad (22)$$

The errors  $\mathbf{v}$  can be represented as the output of a linear dynamic system in scale driven by white noise,

$$\mathbf{v}(\tau) = h\mathbf{v}(\tau\alpha) + h\mathbf{v}(\tau\beta) \quad (23)$$

Suppose there are  $N$  measurements at the finest scale, in order to propagate the prior parameters from  $N/2$  left nodes at scale to their  $N/2$  parent nodes at scale  $m+1$ , it is necessary to decorrelate the errors between these two sets of measurements. A measurement differencing scheme can be implemented to generate pseudo measurements corrupted with additive white noise in scale as well as time (Jazwinski, 1970). Define new measurements as linear combinations of data at two adjacent scales connected by an upward left shift operator  $\bar{\alpha}$  (see Fig. 2):

$$\xi(\tau) = \mathbf{y}_x(\tau\bar{\alpha}) - h\mathbf{y}_x(\tau) = h\mathbf{y}_x(\sigma) = \mathbf{M}(m+1)\mathbf{x}(\tau) + \epsilon(\tau) \quad (24)$$

The covariance matrix of the new measurement error  $\epsilon(\tau)$  is computed with,

$$\mathbf{S}(m+1) = \mathbf{C}\mathbf{B}_x^z(m+1)\mathbf{Q}_u(m+1)\mathbf{B}_x^{zT}(m+1)\mathbf{C}^T + h^2\mathbf{R}(m+1) \quad (25)$$

Considering the newly developed correlation between the measurement errors  $\epsilon$  and the process input  $\mathbf{u}$ , the covariance matrix is given by,

$$\mathbf{Q}_{ue}(m+1) = \mathbf{Q}_u(m+1)\mathbf{B}_x^{zT}(m+1)\mathbf{C}^T \quad (26)$$

Note that the derived measurements  $\xi(\tau)$  contain information from the neighboring node  $\sigma$ . The use of future information is necessary to whiten serially correlated measurement noise. As a consequence, measurement differencing schemes always result in smoothing (Jazwinski, 1970). From a Bayesian perspective, it is

also necessary to use information from  $\sigma$  to rectify the scaled signal measurements at  $\tau$  because all the measurement information available until  $\tau$  has already been used for processing the scaled signal at  $\tau\alpha$  and  $\tau\beta$ . Since measurement differencing cannot be implemented for wavelet coefficients, the prior parameters are propagated from the scaled signal to the wavelet coefficients at the same scale.

Wavelet decomposition approximately decorrelates the wavelet coefficients and the scaled signal coefficients at the same scale, the extent of decorrelation being dependent on the particular family of wavelets selected. Hence, although the wavelet coefficients at scale  $m+1$  are correlated with the scaled signal coefficients at scale  $m$  from which they are generated, they are approximately decorrelated from the scaled signal coefficients at scale  $m+1$ . This is a common assumption in many wavelet-based methods. Based on this assumption wavelet coefficients may be treated as new information at that scale, even though all the measured information has already been used for the rectification of the scaled signal. This assumption breaks down at coarser scales because the decorrelation between scaled signal and wavelet coefficients is progressively weakened.

## 5. Time-scale recursion

The multiscale Bayesian rectification problem is posed as an optimization problem, which is solved at multiple scales inside a moving window to rectify the scaled and wavelet coefficients. Rectification is initialized at the finest scale in the window with the Bayesian data rectification solution at the finest scale (Ungarala & Bakshi, 2000a). Further processing follows the direction of fine to coarse decomposition on the binary tree based on the Haar wavelet transform. The recursive scheme for multiscale Bayesian data rectification is shown in Fig. 4. Recursion in time is followed at the finest scale and recursion in scale follows the direction of left node to parent node connections for the scaled signal. The wavelet coefficients are rectified with recursion from the scaled signal coefficient at the same node. The fine to coarse scale recursion steps proceed independently for each node and allow for parallel processing. Finally the rectified signal at the finest scale is synthesized by the inverse wavelet transform using the rectified values of the coarsest scaled signal and the wavelet coefficients at all scales. Details are described in Ungarala and Bakshi (2000b).

## 6. Simulation example

Consider the following linear dynamic model for a level control process, where  $h$ ,  $F_3$ ,  $x$  and  $e$  are deviation

variables for the level, feed flowrate, valve stem position and controller output, respectively (Narasimhan & Mah, 1988),

$$\begin{bmatrix} h_{k+1} \\ x_{k+1} \end{bmatrix} = \begin{bmatrix} 0.995 & -0.1373 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} h_k \\ x_k \end{bmatrix} + \begin{bmatrix} 0.00012 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} F_{3k} \\ e_k \end{bmatrix} \quad (27)$$

The input variables  $F_3$  and  $e$  are stochastic and vary as iid Gaussian random processes with variances 625 and 0.0025, respectively. The simulated data consists of 256 measurements of inputs and outputs, all corrupted by white noise of S.D. 0.5. The optimization problem is solved at each scale to rectify  $x_k = [h_{k+1}, x_{k+1}, h_k, x_k, F_{3k}, e_k]^T$  at each node. A normalized mean squared error of rectification is defined as,

$$\text{M.S.E.} = \frac{1}{nN} \sum_{k=1}^N (\mathbf{x}_k - \hat{\mathbf{x}}_k)^T \mathbf{R}^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_k) \quad (28)$$

The results of rectification are shown for decomposition to three scales in Fig. 5. For comparison, filtered data by a multiscale Kalman filter based on two-scale models from Stephanopoulos et al. (1998) are also shown in the figure. The improved performance of the Bayesian method can be attributed to the use of the temporal model as a constraint.

The average normalized M.S.E. of approximation of  $\hat{\mathbf{x}}$  for 500 realizations is shown in Fig. 6. The average error is reported for wavelet decomposition to increasing levels of coarseness. At scale  $m = 0$ , the single scale data rectification is the result of the time recursive algorithm (Ungarala & Bakshi, 2000a). The results indicate an optimum scale for data rectification, which is natural since the decorrelation ability of wavelets deteriorates at coarser scales. It violates the fundamental assumption of Bayesian methods that the prior and measurements contain independent information. Furthermore, the assumption that all the dynamics are captured only by the last scaled signal becomes invalid for coarser decompositions as the coarser wavelet coefficients tend to retain the deterministic characteristics of the signal. The figure also compares the results of a multiscale Kalman filter without an EIV approach (Ungarala & Bakshi, 2000b). The multiscale Bayesian data rectification performs consistently better since the constrained optimization results in reconciliation with the process model.

## 7. Conclusions

A general multiscale Bayesian method for data rectification of linear systems with an EIV approach is proposed. Since process data are usually multiscale in nature, the multiscale Bayesian approach provides better rectification than traditional single scale methods. The improved performance of the multiscale approach is due to the ability of orthonormal wavelets to approximately decorrelate most stochastic processes, and extract deterministic features. Our approach extends previous multiscale methods (Chou et al., 1994; Stephanopoulos et al., 1998) by reducing the errors in all the variables and constraining the solution to satisfy the dynamic model. The time domain dynamic model is converted into two scale linear dynamic models for the evolution of the process in scale. The constrained optimization at each scale results in data reconciliation with the temporal model as opposed to multiscale filtering. The scale models are used to adapt the prior across scales. This paper has focused primarily on multiscale recursive data rectification of linear systems with Gaussian distributions. Current research is focused on the

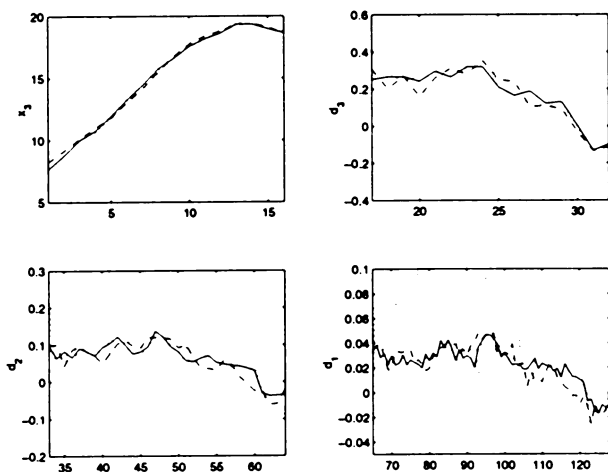


Fig. 5. Rectified scaled signal and wavelet coefficients of  $\{h_k\}$  for  $L=3$ ; (...) multiscale Kalman filter, (- - -) multiscale Bayesian EIV, (—) noise-free data.

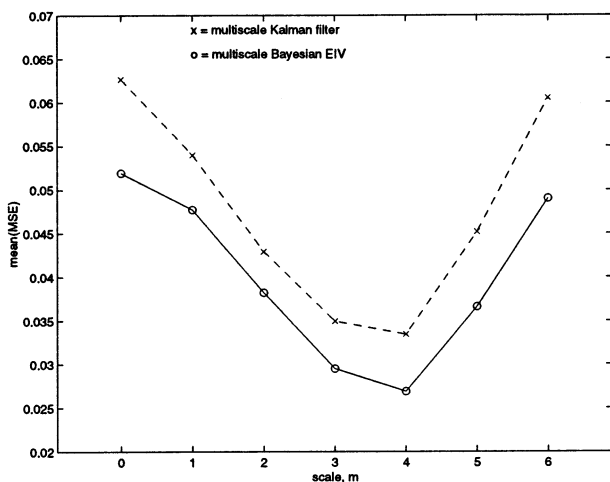


Fig. 6. Average estimation error of  $\hat{\mathbf{x}}$  for 500 realizations.

development of multiscale rectification for nonlinear systems. Extension to using a library of basis functions is expected to result in even better performance.

### Acknowledgements

Financial support from the National Science Foundation through grant CTS-9733627 and the DuPont Education Foundation are acknowledged.

### References

- Bakshi, B. R. (1998). Multiscale PCA with application to multivariate statistical process monitoring. *American Institute of Chemical Engineering Journal*, 44, 1596.
- Bakshi, B. R., Nounou, M. N., Goel, P. K., Shen, X. (1999). Multiscale Bayesian data rectification with steady-state models. Technical Report, The Ohio State University, <http://www.che.eng.ohio-state.edu/~bakshi/msbrec.pdf>.
- Beylkin, G., Coifman, R., & Rokhlin, V. (1991). Wavelet transforms and numerical algorithms I. *Comm. Pure Applied Mathematics*, XLIV, 141.
- Chou, K. C., Willsky, A. S., & Benveniste, A. (1994). Multiscale recursive estimation, data fusion, and regularization. *IEEE Transactions on Automatic Control*, 39, 464.
- Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., & Picard, D. (1995). Wavelet shrinkage: asymptopia. *Journal of the Royal Statistical Society Series B*, 57, 41.
- Hirchoren, G. A., & D'Attellis, C. E. (1999). Estimation of fractional Brownian motion with multiresolution Kalman filter banks. *IEEE Transactions on Signal Processing*, 47, 1431.
- Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic Press.
- Miller, E., & Willsky, A. S. (1995). A multiscale approach to sensor fusion and the solution of linear inverse problems. *Applied and Comparative Harmonic Analysis*, 2, 127.
- Narasimhan, S., & Mah, R. S. H. (1988). Generalized likelihood ratios for gross error identification in dynamic processes. *American Institute of Chemical Engineering Journal*, 34, 1321.
- Stephanopoulos, G., Karsligil, O., & Dyer, M. S. (1998). Multiscale aspects in model predictive control. In R. Berber, & C. Kravaris, *Nonlinear Model Based Control NATO ASI* (pp. 667–734). The Netherlands: Kluwer Academic Publishers.
- Ungarala, S., & Bakshi, B. R. (2000a). Bayesian error-in-variables rectification of data from linear dynamic systems. Technical Report, The Ohio State University, <http://www.che.eng.ohio-state.edu/~bakshi/bayrec.pdf>.
- Ungarala, S., & Bakshi, B. R. (2000b). A multiscale, Bayesian and error-in-variables approach for linear dynamic data rectification. Technical Report, The Ohio State University, <http://www.che.eng.ohio-state.edu/~bakshi/mbdrec.pdf>.
- Wornell, G. W., & Oppenheim, A. V. (1992). Wavelet-based representations for a class of self-similar signals with applications to fractal modulation. *IEEE Transactions on Information Theory*, 38, 785.