Multiscale Realization and Estimation for Space and Space-Time Problems

Terrence T. Ho, Austin B. Frakt, Alan S. Willsky

Laboratory for Information and Decision Systems, MIT, Cambridge, MA[1]

1 Introduction

Since the introduction of the multiscale framework by Willsky, Benveniste, Chou, et al. [3], a significant amount of work has been done on its theoretical development and on its application to real problems. The continued interest in this framework stems in large part from the existence of an efficient multiscale smoothing algorithm capable of handling non-stationary processes and measurements, as well as generating error statistics—information demanded in many applications—with no additional computation beyond what is needed to compute the estimates themselves [4].

In recent years, each advancement in modeling techniques has resulted in new areas of application. Simple multiscale models for $1/f$-like random fields developed early on have proven effective in application to surface reconstruction [7, 8], optical flow [14], and MRI image segmentation [13]. Multiscale models for exact and approximate Markov random fields (MRF) fall under a class of so-called internal models and have been applied to texture classification [15] and ocean hydrography [18]. Other internal models have been used for modeling fractional Brownian motion [6] and for solving inversion problems in ground-water hydrology [5]. Techniques for modeling random fields with any second-order statistical structure are available, but suffer from high computational complexity [11, 12].

Although there have been significant advances in techniques for the realization of multiscale models, there remain a number of important open problems two of which will be addressed in this paper. After reviewing the basics of the multiscale framework we will outline in Section 3 a new static realization method that is two orders of magnitude more efficient than methods previously available [11, 12] but no less general. The second area of current effort (Section 4) focuses on the extension of multiscale methodology to dynamic systems [10], motivated by applications including ocean circulation models and numerical methods for distributed parameter systems.

[1]Supported by ONR Grant N00014-91-J-1004 and by AFOSR Grant F49620-96-1-0028 through Boston University.

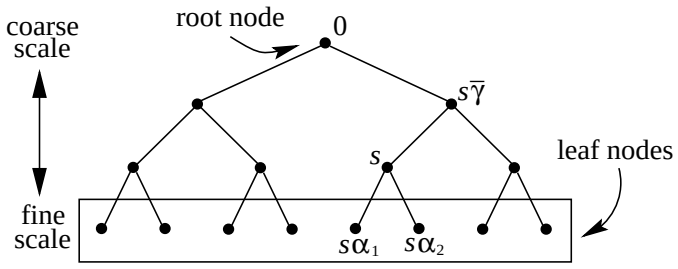


Figure 1: A dyadic tree. We denote the finest level as scale M; the coarsest scale is scale 0. Nodes $s\alpha_1$ and $s\alpha_2$ are the left and right children of node s, respectively while $s\bar{\gamma}$ is the parent of node s.

2 Multiscale Framework

In the multiscale framework, zero-mean random fields are modeled on multiscale tree structures whose scale-to-scale relationship is expressed as:

$$\boldsymbol{x}(s) = \boldsymbol{A}(s)\boldsymbol{x}(s\bar{\gamma}) + \boldsymbol{B}(s)\boldsymbol{w}(s), \tag{1}$$

where s is any node of the tree, $s\bar{\gamma}$ is the parent node of s, and $\boldsymbol{w}(s)$ is a white noise process uncorrelated with the root-node state of the tree $\boldsymbol{x}(0)$, which has mean $\mathbf{0}$ and covariance $\boldsymbol{P}_{\boldsymbol{x}(0)}$. As a result, the multiscale process $\boldsymbol{x}(\cdot)$ has the following Markov property: conditioned on $\boldsymbol{x}(s)$ the subprocesses of $\boldsymbol{x}(\cdot)$ indexed on the disjoint subtrees extending from node s are mutually conditionally uncorrelated. Because of this Markov property and in analogy with standard state-space models, $\boldsymbol{x}(s)$ is referred to as the state at node s. This Markov property leads to an efficient scale-recursive smoothing algorithm which is a generalization of the Rauch-Tung-Striebel smoother [4].

While the multiscale framework and algorithms apply to any tree whatsoever, one-dimensional random processes are typically represented at the finest level of a dyadic tree, as illustrated in Figure 1. Two-dimensional random fields are typically represented at the finest level of a quadtree. The challenge is to determine appropriate multiscale state definitions at coarser scales and the model parameters $\boldsymbol{A}(s)$, $\boldsymbol{B}(s)$, and $\boldsymbol{P}_{\boldsymbol{x}(0)}$ so that the process $\boldsymbol{\chi}$ modeled at the finest scale of the tree has the desired statistics. This is the stochastic realization problem. A convenient parameterization for the realization problem is given by the class of internal models, in which state variables at s, $\boldsymbol{x}(s)$, are linear functions of $\boldsymbol{\chi}$:

$$\boldsymbol{x}(s) = \boldsymbol{L}^T(s)\boldsymbol{\chi}\,. \tag{2}$$

We refer to the individual rows of $\boldsymbol{L}^T(s)$ as linear functionals and the collection $\boldsymbol{L}^T(s)$ as an internal matrix. Once the internal matrices are found, the model parameters can be determined from second-order statistics for $\boldsymbol{\chi}$ as follows. Viewing (1) as the linear least squares estimation (LLSE) of $\boldsymbol{x}(s)$ based on $\boldsymbol{x}(s\bar{\gamma})$ plus the estimation error, and using standard LLSE formulae, we find that

$$\boldsymbol{A}(s) = \boldsymbol{L}^T(s)\boldsymbol{P}_{\boldsymbol{\chi}}\boldsymbol{L}(s\bar{\gamma})\left(\boldsymbol{L}^T(s\bar{\gamma})\boldsymbol{P}_{\boldsymbol{\chi}}\boldsymbol{L}(s\bar{\gamma})\right)^{-1}, \tag{3}$$

$$\boldsymbol{B}(s)\boldsymbol{B}^T(s) = \boldsymbol{L}^T(s)\boldsymbol{P}_{\boldsymbol{\chi}}\boldsymbol{L}(s) - \boldsymbol{A}(s)\boldsymbol{L}^T(s\bar{\gamma})\boldsymbol{P}_{\boldsymbol{\chi}}\boldsymbol{L}(s)\,. \tag{4}$$

The state covariance at the root node $\boldsymbol{P}_{\boldsymbol{x}(0)}$ is simply $\boldsymbol{L}^T(0)\boldsymbol{P}_{\boldsymbol{\chi}}\boldsymbol{L}(0)$.

The internal matrices for MRFs can be specified without computation. The linear functionals retained at each node s consist of edge points of $\boldsymbol{\chi}(s)$ in 2-D or end points in 1-D [15], where $\boldsymbol{\chi}(s)$ is the portion of $\boldsymbol{\chi}$ that descends from s. One algorithm for computing internal matrices for general random fields is based on canonical correlations analysis [1, 11]. An advantage of this canonical correlations approach is that it provides a prioritized ordering of the information retained at each state. Thus, if a reduced-order approximate realization of $\boldsymbol{\chi}$ is desired, it is clear what state variables are least important and can be discarded. Reduced-order realizations are often desired because the multiscale estimator has computational complexity that is cubic in state dimension. A serious disadvantage of the canonical correlations approach of [11, 12], however, is its computational complexity of $\mathcal{O}(N^4)$, where N is the size of $\boldsymbol{\chi}$. The new approach outlined in the next section is only quadratic in problem size.

3 Efficient Multiscale Realization

3.1 Decorrelating Random Vectors

The heart of the realization problem is finding for each s the internal matrix $\boldsymbol{L}^T(s)$ so that (i) the state $\boldsymbol{x}(s) = \boldsymbol{L}^T(s)\boldsymbol{\chi}$ contains the appropriate decorrelating information as dictated by the Markov property associated with (1), (ii) the number of rows of $\boldsymbol{L}^T(s)$ is minimal, and (iii) the information is prioritized. We will address these three challenges by considering a pair-wise decorrelation problem which is then easily generalized for the problem of decorrelating three (or more) random vectors. While our approach is general, for simplicity and clarity we shall restrict attention to the realization

of a one-dimensional processes modeled on dyadic trees. Details omitted here may be found in [9].

Let $\boldsymbol{z} = [\boldsymbol{z}_1^T \; \boldsymbol{z}_2^T]^T$ have zero-mean where $\boldsymbol{z}_i$ has length n_i and $n_1 \ll n_2$. Also,

$$\boldsymbol{P}_i \stackrel{\triangle}{=} E[\boldsymbol{z}_i \boldsymbol{z}_i^T] \;\text{ and }\; \boldsymbol{P}_{12} \stackrel{\triangle}{=} E[\boldsymbol{z}_1 \boldsymbol{z}_2^T]\,. \tag{5}$$

Let $\mathcal{M}_r$ be the set of $r \times n_1$ matrices. Canonical correlations analysis provides the matrix $\boldsymbol{V} \in \mathcal{M}_r$ such that conditioned on $\boldsymbol{V}\boldsymbol{z}_1$ the vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are maximally conditionally decorrelated. Since $\boldsymbol{V}$ is restricted to have r rows, we cannot expect that $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are exactly conditionally decorrelated by $\boldsymbol{V}\boldsymbol{z}_1$. What canonical correlations analysis provides is the best possible decorrelating information subject to this row constraint. The drawback to this approach, which forms the engine of the method of [12], is that it is computationally intensive, requiring $\mathcal{O}(n_2^3)$ operations.

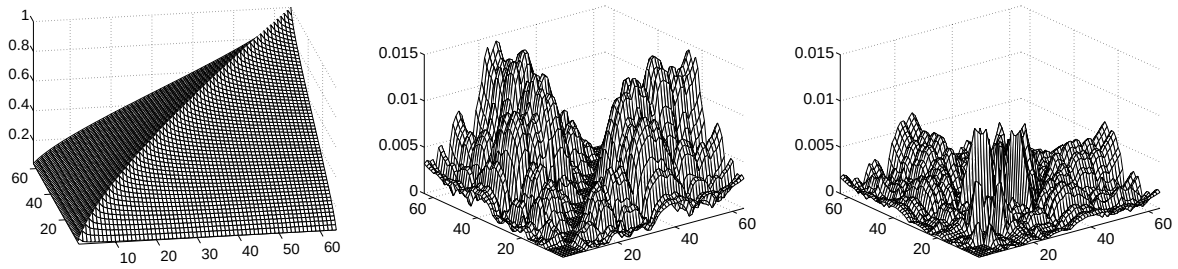
Our approach[2] instead considers the mean square estimation error (MSEE) in estimating $\boldsymbol{z}_2$ based on a r linear functions of $\boldsymbol{z}_1$. By considering the optimal estimation gain matrix $\boldsymbol{P}_{12}^T \boldsymbol{P}_1^{-1}$, it can be shown that the element of $\mathcal{M}_r$ which minimizes the MSEE is given by the first r rows of the matrix $\boldsymbol{U}^T \boldsymbol{P}_1^{-1/2}$ where $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ is the eigen-decomposition of $\boldsymbol{P}_1^{-1/2}\boldsymbol{P}_{12}\boldsymbol{P}_{12}^T\boldsymbol{P}_1^{-T/2}$ with the eigenvalues forming the diagonal of $\boldsymbol{\Lambda}$ in decreasing order from upper-left to lower-right. The computational complexity of computing $\boldsymbol{U}^T \boldsymbol{P}_1^{-1/2}$ is $\mathcal{O}(n_2)$.

To mutually conditionally decorrelate the three sub-vectors of $\boldsymbol{z} = [\boldsymbol{z}_1^T \; \boldsymbol{z}_2^T \; \boldsymbol{z}_3^T]^T$, as in the case for realization on dyadic trees, we simply consider the two pair-wise problems of finding the best $\boldsymbol{V}_1\boldsymbol{z}_1$ to conditionally decorrelate $\boldsymbol{z}_1$ and $[\boldsymbol{z}_2^T \; \boldsymbol{z}_3^T]^T$ and the best $\boldsymbol{V}_2\boldsymbol{z}_2$ to conditionally decorrelate $\boldsymbol{z}_2$ and $[\boldsymbol{z}_1^T \; \boldsymbol{z}_3^T]^T$. What "best" means here depends on the approached used—canonical correlations minimizes correlation coefficients while our approach minimizes MSEE. The stacked vector $\left[(\boldsymbol{V}_1\boldsymbol{z}_1)^T \; (\boldsymbol{V}_2\boldsymbol{z}_2)^T\right]^T$ will exactly conditionally decorrelate $\boldsymbol{z}_1$, $\boldsymbol{z}_2$, and $\boldsymbol{z}_3$ if no restrictions on the number of rows of $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are applied. Approximate conditional decorrelation is obtained under a row constraint.

3.2 Realization Algorithm

For internal multiscale processes the Markov property mentioned before is equivalent to the following simpler condition: $\boldsymbol{x}(s)$ at scale $n \neq M$ conditionally decorrelates $\boldsymbol{x}(s\alpha_1)$, $\boldsymbol{x}(s\alpha_2)$, and $\{\boldsymbol{x}(t) \mid t \text{ is at scale } n+1, t \neq s\alpha_1, s\alpha_2\}$. Therefore, if this simpler condition is satisfied at each

[2] A similar approach applied to the classical stationary AR stochastic realization problem is discussed in [2].



(a) (b) (c)

Figure 2: (a) is $\boldsymbol{P}_{\boldsymbol{\chi}}$, the covariance for 64 samples of fBm(0.3) on $(0, 1]$. (b) and (c) are the element-wise absolute value of the difference between the $\boldsymbol{P}_{\boldsymbol{\chi}}$ and the finest scale covariance realized by: (b) the method of [12]; (c) our approach. In both cases the maximum state dimension is four.

node s then the Markov property is automatically satisfied. The algorithm proceeds as follows. For each scale $n = M-1, M-2, \ldots, 0$ and for each node s at scale n do the following:

1. For $i = 1, 2$, find the vector $\boldsymbol{V}_{s\alpha_i}\boldsymbol{x}(s\alpha_i)$ which minimizes the MSEE in estimating the other states at scale $n+1$ from $\boldsymbol{V}_{s\alpha_i}\boldsymbol{x}(s\alpha_i)$. This is a pair-wise problem as described previously.

2. Form the stacked vector $\left[(\boldsymbol{V}_{s\alpha_1}\boldsymbol{x}(s\alpha_1))^T \ (\boldsymbol{V}_{s\alpha_2}\boldsymbol{x}(s\alpha_2))^T\right]^T$. This solves the three-way problem involving $\boldsymbol{x}(s\alpha_1)$, $\boldsymbol{x}(s\alpha_2)$ and all the other states at scale $n+1$ as described previously.

3. If the resulting state dimension is too large, delete the appropriate number of rows of the $\boldsymbol{V}_{s\alpha_i}$ matrices. Because the rows of $\boldsymbol{V}_{s\alpha_i}$ are in priority order, it is a simple matter to decide which information is least important.

Together the $\{\boldsymbol{V}_{s\alpha_i}\}$ define the internal matrices. Notice that in the above algorithm each node is visited once and the complexity of the calculation at each node is $\mathcal{O}(N)$. Therefore, the overall complexity is $\mathcal{O}(N^2)$—constant computational complexity per element of $\boldsymbol{P}_{\boldsymbol{\chi}}$ the matrix that we are trying to realize.

3.3 An Example

As an example we compare our approach to that described in [12] for the case where $\boldsymbol{\chi}$ consists of samples of fractional Brownian motion (as defined in [17]) with Hurst parameter $H = 0.3$ (denoted by fBm(0.3)). Figure 2(a) illustrates $\boldsymbol{P}_{\boldsymbol{\chi}}$, the covariance matrix for 64 equally spaced samples

of fBm(0.3) on the interval $(0, 1]$. Figure 2(b) illustrates the element-wise absolute value of the difference between $\boldsymbol{P}_{\boldsymbol{\chi}}$ and the finest-scale covariance realized by the method described in [12] based on canonical correlations. Figure 2(c) illustrates the element-wise absolute value of the difference between $\boldsymbol{P}_{\boldsymbol{\chi}}$ and the finest-scale covariance realized by our MSEE-based method. For both cases the maximum state dimension is four. The main message of Figure 2 is that the two methods yield approximations to $\boldsymbol{P}_{\boldsymbol{\chi}}$ of comparable quality. However, our approach is $\mathcal{O}(N^2)$ while that of [12] is $\mathcal{O}(N^4)$.

4 Multiscale estimation of Dynamic Systems

The fast realization algorithm just discussed assumes that the covariance of the random process to be realized is available. This, however, is not always true. Suppose that we wish to construct the steady-state estimator of a large dynamic system. Solving for the exact steady-state estimation error covariance is an $\mathcal{O}(N^3)$ operation with a storage requirement of $\mathcal{O}(N^2)$.

Our proposed multiscale dynamic estimation algorithm is best explained by drawing a parallel with the discrete-time Kalman filter. Similar to the Kalman filter's initialization with prior error covariance, we initialize with a multiscale model for the errors. The strategy of our algorithm is to then propagate the multiscale *models* for the updated and predicted estimation errors through time much like the Kalman filter propagates the error covariances, but in a computationally efficient manner and without ever computing or storing the full error covariance matrix. We assume a fixed set of linear functionals for modeling the estimation errors, but allow the model parameters to be modified over time.

We will illustrate the development of multiscale dynamic estimation in the context of 1-D diffusion processes governed by

$$\frac{\partial z}{\partial t} = \frac{\partial^2 z}{\partial x^2} - \beta \cdot z + \gamma \cdot \xi, \tag{6}$$

where z is the temperature distribution on a ring or rod depending on boundary conditions, ξ is white Gaussian noise, and β and γ are constants. We discretize this PDE in space and time to arrive at a system of difference equations

$$\boldsymbol{z}(t+1) = \boldsymbol{F}\boldsymbol{z}(t) + \boldsymbol{\xi}(t). \tag{7}$$

Measurements are assumed to be stationary in time

$$\boldsymbol{y}(t) = \boldsymbol{H}\boldsymbol{z}(t) + \boldsymbol{v}(t), \tag{8}$$

where $\boldsymbol{v}(t)$ is white.

The Kalman filter update step incorporating new measurements at time t can be interpreted as static estimation of the predicted error $\hat{\boldsymbol{z}}(t|t-1)$ given innovation $\boldsymbol{\nu}(t) = \boldsymbol{y}(t) - \boldsymbol{H}\hat{\boldsymbol{z}}(t|t-1)$. Given a multiscale model, $\boldsymbol{A}(s,t|t-1)$, $\boldsymbol{B}(s,t|t-1)$ for the predicted estimation error and given $\boldsymbol{\nu}(t)$, we already know how to use the multiscale smoothing algorithm to efficiently compute the updated estimates of $\hat{\boldsymbol{z}}(t|t-1)$, as well as to produce a multiscale error model, $\boldsymbol{A}(s,t|t)$, $\boldsymbol{B}(s,t|t)$, for the updated estimation errors [16]. The missing piece now is a multiscale counterpart to the Kalman filter prediction step, i.e., a method for constructing a multiscale model for the predicted error $\tilde{\boldsymbol{z}}(t+1|t)$ given a model for $\tilde{\boldsymbol{z}}(t|t)$.

Constructing a multiscale model for the predicted error requires knowledge of cross-correlations among certain nodes on the tree. For internal realization models of (2), such statistics can be obtained from the error dynamics equation $\tilde{\boldsymbol{z}}(t+1|t) = \boldsymbol{F}\tilde{\boldsymbol{z}}(t|t) + \boldsymbol{\xi}(t)$, since for the ith state at node s

$$x_i(s,t+1|t) = \boldsymbol{l}_i^T(s)\tilde{\boldsymbol{z}}(t+1|t) = \boldsymbol{l}_i^T(s)\boldsymbol{F}\tilde{\boldsymbol{z}}(t|t) + \boldsymbol{l}_i^T(s)\boldsymbol{\xi}(t). \tag{9}$$

and $\tilde{\boldsymbol{P}}(t|t) = E\left(\tilde{\boldsymbol{z}}(t|t)\tilde{\boldsymbol{z}}(t|t)^T\right)$ can in principle be computed, but is to be avoided due to its large size. Suppose that we can write

$$\boldsymbol{l}_i^T(s)\boldsymbol{F} = \sum_{(\sigma,j)\in S} h_{\sigma,j}\boldsymbol{l}_j^T(\sigma), \tag{10}$$

where (σ, j) indexes the j-th state at node σ and S is the set of indices of all states on the tree. We can now write $x_i(s,t+1|t)$ in terms of selected states on the updated error model,

$$x_i(s,t+1|t) = \sum_{(\sigma,j)\in S} h_{\sigma,j}x_j(\sigma,t|t) + \boldsymbol{l}_i^T(s)\boldsymbol{\xi}(t), \tag{11}$$

and depending on $\boldsymbol{L}^T(s)$ and $\boldsymbol{F}$ a smaller set of statistics than the full $\tilde{\boldsymbol{P}}(t|t)$ needs to be computed.

By examining the exact steady-state estimation error covariances of 1-D diffusion, we found the steady-state errors to be close to Markov, i.e., the inverse of the error covariance matrix is close to banded. We may, therefore, use the end-point linear functionals for MRFs to approximately

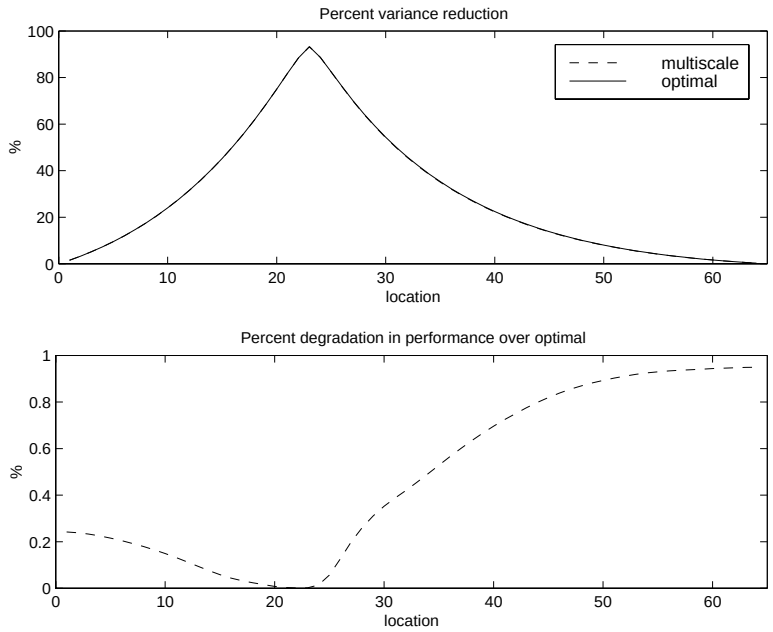


Figure 3: Performance comparison of the steady-state multiscale estimator and the optimal estimator (with $\beta = 10$ and SNR $= 0$dB).

model the estimation errors. With this choice of linear functionals and a banded dynamics matrix $\boldsymbol{F}$, (10) amounts to needing $\mathcal{O}(N)$ number of entries of the $N \times N$ matrix $\tilde{\boldsymbol{P}}(t|t)$ for the multiscale prediction step. These entries can be filled by computing covariances among certain nodes on the tree model, but filling all such entries exactly is an $\mathcal{O}(N^2)$ operation, as extraneous entries are computed in the process. However, if we allow exact computation of only covariances among nodes along direct descendent lines and among nodes across descendent lines over $\log \log N$ scales, and approximate the few remaining entries, the overall complexity for the prediction step is reduced to $\mathcal{O}(N \log N)$.

We show one example here for a 64-element diffusion process with Dirichlet boundary condition and one measurement at location 23. We iterate the update and prediction steps until the multiscale estimator converges. We then use that as our suboptimal steady-state estimator and compare with the optimal steady-state estimator. We use error variance reduction, defined as

$$\text{variance reduction} = \frac{\text{Var(s.s. process)} - \text{Var(s.s. updated error)}}{\text{Var(s.s. process)}} \tag{12}$$

to measure performance. The top panel of Figure 3 shows the variance reduction comparison be-

tween the steady-state optimal estimator and the steady-state approximate multiscale estimator, and the bottom panel shows the percent degradation in variance reduction of the multiscale estimator in relation to the optimal. The suboptimal multiscale estimator performed well for the case shown in the figure as well as for other combinations of process and measurement parameters. This level of performance is achieved with computational complexity of $\mathcal{O}(N)$ and $\mathcal{O}(N \log N)$ for the update and predictions steps, respectively.

5 Conclusion

In this paper we have outlined some recent developments in multiscale realization and estimation. We have summarized a novel algorithm which can produce an exact or approximate realization for any fine-scale covariance structure. One advantage of our approach to this problem is that it is quadratic in problem size while the only other known general purpose realization algorithm is quartic in problem size. Our future work in this area will focus on even more efficient algorithms for the static realization problem. In particular, we seek algorithms capable of handling the extremely large problems which arise in image processing contexts.

We have also introduced an approach to realization and estimation for dynamic processes by propagating over time a multiscale model for the estimation errors. For 1-D diffusion, we were able to achieve $\mathcal{O}(N \log N)$ computational complexity and $\mathcal{O}(N)$ storage requirements versus the Kalman filter's $\mathcal{O}(N^3)$ and $\mathcal{O}(N^2)$. This algorithm has been shown to work equally well for other types of dynamic systems, e.g. advection-diffusion. The challenge lying ahead rests in the expansion to 2-D systems where controlling state dimensions becomes far more critical to computational efficiency.

References

[1] H. Akaike. Markovian representation of stochastic processes by canonical variables. *SIAM Journal of Control*, 13(1):162–173, January 1975.

[2] K. S. Arun and S. Y. Kung. Balanced approximation of stochastic systems. *SIAM Journal of Matrix Analysis and Applications*, 11(1):42–68, January 1990.

[3] M. Basseville, A. Benveniste, K. Chou, S. Golden, R. Nikoukhah, and A. Willsky. "Modeling and Estimation of Multiresolution Stochastic Processes". *IEEE Transactions on Information Theory*, 38(2):766–784, Mar. 1992.

[4] K. Chou, A. Willsky, and A. Benveniste. "Multiscale Recursive Estimation, Data Fusion, and Regularization". *IEEE Transactions on Automatic Control*, 39(3):464–478, Mar. 1994.

[5] M. Daniel and A. Willsky. "A Multiresolution Methodology for Signal-Level Fusion and Data Assimilation with Applications to Remote Sensing". *Proceedings of the IEEE*, 85(1):164–183, Jan. 1997.

[6] M. Daniel and A. Willsky. "Modeling and Estimation of Fractional Brownian Motion Using Multiresolution Stochastic Processes". In J. L. Vehel, E. Lutton, and C. Tricot, editors, *Fractals in Engineering*, pages 124–137. Springer, 1997.

[7] P. Fieguth, W. Karl, A. Willsky, and C. Wunsch. "Multiresolution Optimal Interpolation and Statistical Analysis of TOPEX/POSEIDON Satellite Altimetry". *IEEE Transactions on Geoscience and Remote Sensing*, 33(2):280–292, Mar. 1995.

[8] P. Fieguth, D. Menemenlis, T. Ho, A. Willsky, and C. Wunsch. "Mapping Mediterranean Altimeter Data with a Multiresolution Optimal Interpolation Algorithm". *Journal of Atmospheric and Oceanic Technology*, 15:535–546, April 1998.

[9] A. B. Frakt and A. S. Willsky. Efficient multiscale stochastic realization. In *Proceedings of the International Conference on Acoustics, Speech, and Singal Proccessing*, Seattle, Washington, May 1998. Submitted.

[10] T. Ho, P. Fieguth, and A. Willsky. "Multiresolution Stochastic Models for the Efficient Solution of Large-Scale Space-Time Estimation Problems". In *Proceedings of ICASSP*, volume 6, pages 3098–4001, Atlanta, Georgia, May 1996.

[11] W. W. Irving. *Theory for Multiscale Stochastic Realization and Identification.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, Sep. 1995.

[12] W. W. Irving, W. C. Karl, and A. S. Willsky. A theory for multiscale stochastic realization. In *Proceedings of the 33rd IEEE Conference on Decision and Control*, volume 1, pages 655–62, Lake Buena Vista, Florida, USA, December 1994.

[13] J. Kaufhold, M. Schneider, W. Karl, and A. Willsky. "A Statistical Approach to Segmentation of MRI Imagery". *Journal of Pattern Recognition and Artificial Intelligence.* To appear.

[14] M. Luettgen, W. Karl, and A. Willsky. "Efficient Multiscale Regularization with Applications to the Computation of Optical Flow". *IEEE Transactions on Image Processing*, 3(1):41–64, Jan. 1994.

[15] M. Luettgen, W. Karl, A. Willsky, and R. Tenney. "Multiscale Representations of Markov Random Fields". *IEEE Transactions on Signal Processing*, 41(12):3377–3395, Dec. 1993.

[16] M. Luettgen and A. Willsky. "Multiscale Smoothing Error Models". *IEEE Transactions on Automatic Control*, 40(1):173–175, Jan. 1995.

[17] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437, 1968.

[18] D. Menemenlis, P. Fieguth, C. Wunsch, and A. Willsky. "Adaptation of a Fast Optimal Interpolation Algorithm to the Mapping of Oceanographic Data". *Journal of Geophysical Research*, 102(C5):10573–10584, 1997.