COMPUTATIONALLY EFFICIENT MULTISCALE ESTIMATION OF LARGE-SCALE DYNAMIC SYSTEMS

Terrence T. Ho[†], *Paul W. Fieguth*[‡], *Alan S. Willsky*[†]

† Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

‡ Dept. of Systems Design Engineering
University of Waterloo

ABSTRACT

Statistical estimation of large-scale dynamic systems governed by stochastic partial differential equations is important in a wide range of scientific applications. However, the realization of computationally efficient algorithms for statistical estimation of such dynamic systems is very difficult. Conventional linear least squares methods are impractical for both computational and storage reasons.

A recently-developed multiscale estimation methodology has been successfully applied to a number of large-scale static *estimation problems. In this paper we apply the multiscale approach to the more challenging* dynamic *estimation problems, introducing a recursive procedure that efficiently propagates multiscale* models *for the estimation errors in a manner analogous to, but more efficient than, the Kalman filter's propagation of the error covariances. We will illustrate our research in the context of 1-D and 2-D diffusive processes.*

1. INTRODUCTION

Statistical estimation of dynamic systems governed by stochastic partial differential equations finds numerous scientific applications ranging from computing optical flow [1] to tracer tracking in the ocean [7]. Dynamic estimation of large-scale systems governed by even simple dynamics such as advection-diffusion is challenging and would represent a major step forward in numerous applications, such as imaging ocean surface height [3], where methods for assimilating data with dynamics have hitherto been limited. Conventional linear least squares estimation methods such as Kalman filtering are impractical given the sheer size of the error covariance matrices, as the algorithm is of $\mathcal{O}(N^3)$ computational complexity and requires $\mathcal{O}(N^2)$ storage. Computationally more efficient methods such as multigrid and FFT are undesirable because they either do not supply error statistics or require spatially stationary prior models and regular measurement patterns.

In this paper we consider the use of a multiscale stochastic modeling and estimation methodology developed in [2], which has been successfully applied to a number of large-scale *static* estimation problems [5, 3]. This class of models admits an $\mathcal{O}(N)$ algorithm for the calculation of estimates and error variances. For *dynamic* systems, in the same way that the Kalman filter propagates the error covariance matrices through the dynamic prediction and measurement update steps, similarly in our multiscale context we need to propagate the statistics for the multiscale estimation error. Given the sheer size of our estimation problem, the key challenge is that we need to propagate *models* for the estimation error, not covariances, and to do so efficiently. Since each measurement update step has a straightforward interpretation as solving a *static* estimation problem, propagating the estimation error model through the measurement update step is easy and is accomplished as a by-product of the $\mathcal{O}(N)$ estimation algorithm [6]. It is the propagation of the multiscale model through the prediction step that remains.

The challenge here is to untangle the mixing effect of the dynamics in updating the parameters of the model. Doing this requires a careful examination of what variables should be captured at each scale in the multiscale representation of the estimation errors and how the temporal dynamics mix the variables. In this paper we perform that examination for 1-D and 2-D diffusion processes. For 1-D diffusion we can construct an $\mathcal{O}(N)$ algorithm for the dynamic propagation of our multiscale models for the estimation errors resulting in an overall algorithm with near-optimal performance. In 2-D the computational complexity rises to $\mathcal{O}(N^{3/2})$ for exact models, but even this makes the multiscale approach an attractive alternative for large-scale estimation problems when compared to $\mathcal{O}(N^3)$ of the full Kalman filter. Reduced-order models for 2-D diffusion are currently under investigation.

This work was supported by ONR under Grant N00014-91-J-1004, and by AFOSR under Grant F49620-95-1-0083.

0-8186-8821-1/98 $10.00 © 1998 IEEE

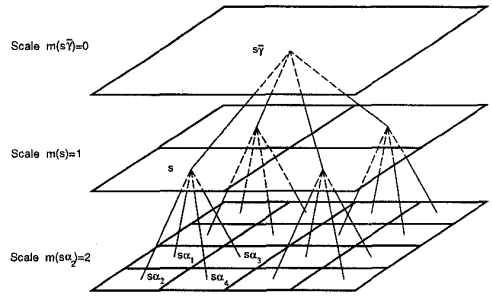

Fig. 1. An example multiscale tree for 2-D fields.

2. MULTISCALE MODELING

In the multiscale estimation framework, random fields are modeled on multiscale tree structures, e.g., in Fig. 1 for a quad tree where each node has $q = 4$ children:

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

where s indexes the nodes of the tree, $s\bar{\gamma}$ is the parent of s, and $\boldsymbol{w}(s)$ is a white noise process uncorrelated with $\boldsymbol{x}(0)$. Conditioned on the state $\boldsymbol{x}(s)$ the $q+1$ subtrees connected to node s are conditionally decorrelated by the whiteness of $\boldsymbol{w}(s)$. This Markovianity property makes possible an efficient scale-recursive estimation algorithm on the tree [2].

Two-dimensional random fields of interest are often represented at the finest scale of a $q = 4$ quad tree (Fig. 1); similarly 1-D processes can be placed at the finest scale of $q = 2$ dyadic trees. The challenge, in both the static and dynamic estimation cases, is the determination of appropriate multiscale states $\boldsymbol{x}(s)$ and model parameters $\boldsymbol{A}(s)$ and $\boldsymbol{B}(s)$ on all scales so that the desired statistical behavior is produced at the finest scale. One specific class of multiscale models defines the multiscale state at s as a linear function $\boldsymbol{L}(s)$ of the finest-scale process $\boldsymbol{\chi}$:

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

The model parameters $\boldsymbol{A}(s)$ and $\boldsymbol{B}(s)$ are uniquely determined given the joint statistics between $\boldsymbol{x}(s)$ and $\boldsymbol{x}(s\bar{\gamma})$, i.e., $\boldsymbol{P}(s)$, $\boldsymbol{P}(s\bar{\gamma})$, and $\boldsymbol{P}(s, s\bar{\gamma})$, which are easily computed once (2) is defined and given the second-order statistics for $\boldsymbol{\chi}$, $\boldsymbol{P}_{\boldsymbol{\chi}}$. It is known that the class of first-order Markov random fields (MRF) can be exactly modeled by letting $\boldsymbol{L}(s)$ be the boundary points of subregions of $\boldsymbol{\chi}$ in 2-D, or the end points of subintervals of $\boldsymbol{\chi}$ in 1-D, as shown in Fig. 2(a) [5].

The goal of our research is the development of multiscale models for large dynamic systems and of computationally tractable methods for propagating such models without the explicit availability of second-order statistics, $\boldsymbol{P}_{\boldsymbol{\chi}}$.

3. MULTISCALE RECURSIVE ESTIMATION

Consider a discrete-time system whose temporal dynamics are governed by

$$\boldsymbol{z}(t+1) = \boldsymbol{A}_d\,\boldsymbol{z}(t) + \boldsymbol{w}_d(t), \tag{3}$$

where $\boldsymbol{w}_d(t) \sim N(0, \boldsymbol{Q}_d)$ is Gaussian with zero mean and diagonal covariance $\boldsymbol{Q}_d$. The measurements are

$$\boldsymbol{y}_d(t) = \boldsymbol{C}_d(t)\,\boldsymbol{z}(t) + \boldsymbol{v}_d(t), \tag{4}$$

where $\boldsymbol{v}_d(t) \sim N(0, \boldsymbol{R}_d)$ with diagonal $\boldsymbol{R}_d$. Let $\hat{\boldsymbol{z}}(t|\tau)$ denote the estimate of $\boldsymbol{z}(t)$ based on measurements through time τ, and let $\boldsymbol{\chi}(t|\tau)$ denote the corresponding estimation error

$$\boldsymbol{\chi}(t|\tau) = \boldsymbol{z}(t) - \hat{\boldsymbol{z}}(t|\tau). \tag{5}$$

The form of a class of recursive estimators (which includes the optimal Kalman filter) consists of a prediction stage

$$\hat{\boldsymbol{z}}(t+1|t) = \boldsymbol{A}_d\,\hat{\boldsymbol{z}}(t|t), \tag{6}$$

and a measurement update stage

$$\hat{\boldsymbol{z}}(t|t) = \hat{\boldsymbol{z}}(t|t-1) + \hat{\boldsymbol{\chi}}(t|t-1), \tag{7}$$

where $\hat{\boldsymbol{\chi}}(t|t-1)$ is the estimate of the one-step prediction error $\boldsymbol{\chi}(t|t-1)$ based on the measurement

$$\begin{aligned} \boldsymbol{\nu}(t) &= \boldsymbol{y}_d(t) - \boldsymbol{C}_d\,\hat{\boldsymbol{z}}(t|t-1) && (8)\\ &= \boldsymbol{C}_d\,\boldsymbol{\chi}(t|t-1) + \boldsymbol{v}_d(t). && (9) \end{aligned}$$

Each measurement update step is essentially a static estimation problem of computing $\hat{\boldsymbol{\chi}}(t|t-1)$. In standard Kalman filtering the estimate $\hat{\boldsymbol{\chi}}(t|t-1)$ is calculated explicitly as

$$\hat{\boldsymbol{\chi}}(t|t-1) = \boldsymbol{K}(t)\boldsymbol{\nu}(t), \tag{10}$$

for which the prediction error covariance $\boldsymbol{P}_{\boldsymbol{\chi}}(t|t-1)$ is required in computing the Kalman filter gain $\boldsymbol{K}(t)$. $\boldsymbol{P}_{\boldsymbol{\chi}}(t|t-1)$ is propagated through time to $\boldsymbol{P}_{\boldsymbol{\chi}}(t+1|t)$ at the next time step via a Riccati equation, which is of $\mathcal{O}(N^3)$ computational complexity. The updated estimation error covariance $\boldsymbol{P}_{\boldsymbol{\chi}}(t|t)$ is exactly the *a posteriori* estimation error covariance from estimating $\boldsymbol{\chi}(t|t-1)$.

The alternative approach that we consider in this paper involves the *implicit* calculation and propagation of the statistics of the estimation error as a sequence of multiscale models. The multiscale update step, no different from static estimation, is known [6]. The multiscale prediction step—deriving a model $\boldsymbol{A}(s; t+1|t)$ and $\boldsymbol{B}(s; t+1|t)$ for the one-step ahead predicted errors

from a model $\boldsymbol{A}(s;t|t)$ and $\boldsymbol{B}(s;t|t)$ for the updated errors—is substantially more difficult. We write the i-th state variable at node s on the model for $\boldsymbol{\chi}(t+1|t)$:

$$\begin{aligned} & x_i(s;t+1|t) \\ = \; & \boldsymbol{l}_i^T(s)\boldsymbol{\chi}(t+1|t) && (11) \\ = \; & \boldsymbol{l}_i^T(s)\boldsymbol{A}_d\boldsymbol{\chi}(t|t) + \boldsymbol{l}_i^T(s)\boldsymbol{w}_d(t) && (12) \\ = \; & \sum_{(\sigma,j)} h_{\sigma,j}^{s,i}\boldsymbol{l}_j^T(\sigma)x_j(\sigma;t|t) + \boldsymbol{l}_i^T(s)\boldsymbol{w}_d(t) && (13) \\ = \; & \boldsymbol{h}_i^T(s)\boldsymbol{\xi}(t|t) + \boldsymbol{l}_i^T(s)\boldsymbol{w}_d(t), && (14) \end{aligned}$$

where (11) follows from (2), and (12) follows from (3) and (5). In (13) we attempt to relate $x_i(s;t+1|t)$ to states on the model for $\boldsymbol{\chi}(t|t)$, rather than to $\boldsymbol{\chi}(t|t)$ itself. Denoting by $\boldsymbol{\xi}(t|t)$ the collection of all state variables on the updated error model, we may write (13) in the vector form of (14). To compute the joint statistics between $\boldsymbol{x}(s;t+1|t)$ and $\boldsymbol{x}(s\bar{\gamma};t+1|t)$, which will give us $\boldsymbol{A}(s;t+1|t)$ and $\boldsymbol{B}(s;t+1|t)$, more than just the joint statistics between $\boldsymbol{x}(s;t|t)$ and $\boldsymbol{x}(s\bar{\gamma};t|t)$ are needed due to the temporal dynamics, since the sum in (13) usually depends on more than just $x_i(s;t|t)$. Specifically which additional joint statistics will be required will depend on the choice of linear functionals and on the temporal dynamics.

The dynamics generally operate locally in space, e.g., $\boldsymbol{A}_d$ is tridiagonal. Fig. 2(a) shows a realization from [5] for 1-D MRFs, where $\boldsymbol{L}(s)$ consists of the boundary points of the subintervals under s. Consider computing the joint statistics between nodes s and $s\bar{\gamma}$: some of the additional functionals needed for prediction are found only at nodes far away from s or $s\bar{\gamma}$. The overall computational complexity for the prediction step is $\mathcal{O}(N \log N)$.

A new type of *non-redundant* model that we introduce in this paper further reduces the complexity for the prediction step. The physical process is no longer conveniently mapped to the finest scale of the tree, but distributed among the nodes, as shown in Fig. 2(b). Each element of the physical process appears only once on the tree, i.e., there are no redundant functionals. The computational benefit is that the additional functionals needed for the prediction step always appear at either a parent or child node. The computational complexity is only $\mathcal{O}(N)$, the same as the update step. The resulting multiscale recursive algorithm is then of $\mathcal{O}(N)$.

In general one might expect $\boldsymbol{L}(s)$ to change with time if the statistics of the prediction errors are time varying. However, we will assume a fixed set of $\boldsymbol{L}(s)$ for the multiscale realization, an assumption similar in spirit to the assertion of a Markov structure in [1]. The choice of $\boldsymbol{L}(s)$ is not arbitrary as the functionals must

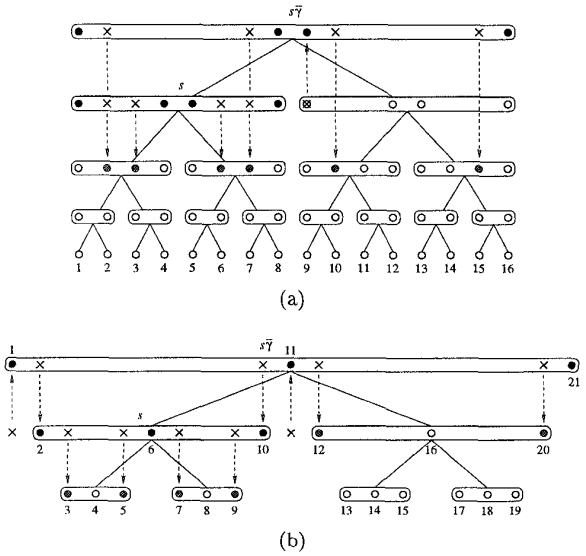

Fig. 2. (a) The linear functionals of the type in [5]; (b) non-redundant linear functionals for modeling the updated estimation errors are labeled with circles (and with filled circles at s and $s\bar{\gamma}$). The × marks indicate the locations of additional functionals needed for the prediction step due to a tridiagonal $\boldsymbol{A}_d$. The arrows point to shaded circles where the additional linear functionals are found. Joint statistics among all nodes with either filled or shaded circles are must be computed to determine $\boldsymbol{A}(s;t+1|t)$ and $\boldsymbol{B}(s;t+1|t)$.

also satisfy the Markovianity property on the tree in order to give an accurate statistical realization. The imposition of a fixed set of linear functionals means that the prediction errors are only approximately realized. We will see that for 1-D diffusion, the linear functionals in Fig. 2 do indeed adequately capture the error statistics under a variety of conditions.

4. 1-D DIFFUSION

Consider a diffusion process governed by the normalized stochastic PDE

$$\frac{\partial z}{\partial t} = \nabla^2 z - \beta \cdot z + \gamma \cdot \omega, \qquad (15)$$

where z is the temperature distribution on a thin rod in 1-D, or on a sheet in 2-D; ω is white Gaussian noise, and β and γ are constants. We discretize this PDE in both space and time to arrive at a system of difference equations in the form of (3).

For small-sized systems, we can solve for the steady-state process $\boldsymbol{P}_z$ and predicted estimation error $\boldsymbol{P}_p$ covariances by exactly solving the Lyapunov and Riccati equations, respectively. We find that $\boldsymbol{P}_z^{-1}$ and $\boldsymbol{P}_p^{-1}$ are

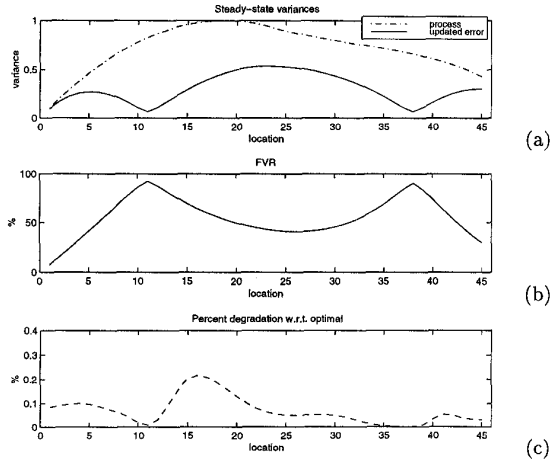

Fig. 3. Pinned cooling fin with two measurements at locations 11 and 38. (SNR = 0 dB.) Heat loss is nonuniform: $\beta = 0$ at locations $1 - 23$ and $\beta = 10$ at $24 - 45$. (a) Steady-state process and steady-state updated estimation error variances. (b) Percent FVR of the optimal and multiscale estimators. The values for the optimal and multiscale estimators are indistinguishable. (c) Percent degradation in FVR of suboptimal multiscale estimator with respect to the optimal.

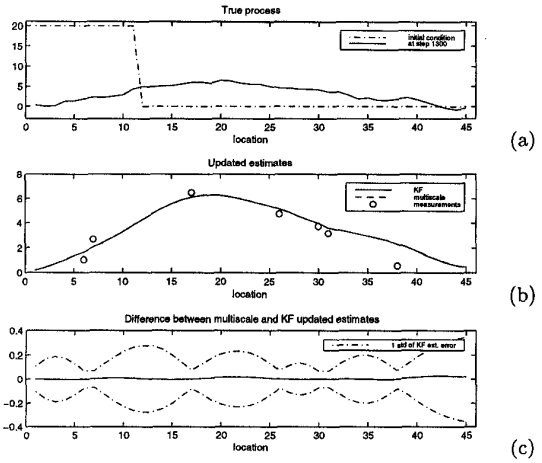

Fig. 4. A warm liquid flows to the right down a cooling pipe. (SNR = 0 dB, $\rho = -10$, $\beta = 0$ at $1 - 32$, $\beta = 10$ at $33 - 64$.) A different measurement pattern is used for every 100 steps (about 1/13 of process time constant). The number of measurements is a Poisson distribution with mean 4. The locations are uniformly distributed. (a) True process at step 1300 and initial values of the process. (b) Updated estimates at step 1300. The optimal and the multiscale suboptimal estimates are indistinguishable at this resolution. The locations and values of the two measurements are labeled with circles. (c) The difference between the multiscale estimates and the optimal estimates. The dotted curves show the magnitude of one standard deviation of the optimal estimation errors.

approximately banded. Since we know that MRFs have banded inverse covariance matrices and that MRFs can be exactly modeled with the linear functionals shown in Fig. 2, we have good reason to believe that these functionals result in good approximate multiscale models for the estimation errors. This is indeed the case as the two examples in Figs. 3 and 4 demonstrate.

The multiscale recursive algorithm can be used as an iterative procedure for finding the steady-state estimator, in the same way that the Kalman filter represents an iterative solution to the Riccati equation. In Fig. 3 we show a 45-element pinned cooling fin example modeled with non-redundant functionals. The multiscale suboptimal estimator is found after 1000 iterations or about one time constant of the optimal steady-state estimator. We use *fractional variance reduction* (FVR) of the updated estimation errors to measure the performance of the estimators:

$$\text{FVR} = \frac{\text{Var(s.s. process)} - \text{Var(s.s. updated error)}}{\text{Var(s.s. process)}}. \tag{16}$$

In Fig. 3(a), we can see the reduction in variance due to the two measurements at 11 and 38. The multiscale steady-state estimators in general perform no more than a few percent below optimal.

The non-redundant linear functionals work well for modeling the estimation errors under a variety of more general conditions. Fig. 4 shows an example in which an advection component $\rho \nabla z$ is added to (15), and the number and locations of measurements are time-varying. Since the problem is time-varying and does not attain steady-state, we compare the multiscale estimates with the optimal ones.

5. 2-D DIFFUSION

Analogous to the 1-D model in Fig. 2(b), we can build 2-D multiscale models which keep the boundary points of s as the state variables in $\boldsymbol{x}(s)$. The 2-D multiscale steady-state estimator performs well compared to the optimal estimator, as the example in Fig. 5 shows. However, these 2-D models have state dimensions that grow with the size of the random field. For a 2-D field of interest with N elements, the complexity of the multiscale recursive estimator becomes $\mathcal{O}(N^{3/2})$. This is still substantially better than $\mathcal{O}(N^3)$ complexity of the Kalman filter. Reduced-order models that subsample the boundary points lead to further reduction in com-

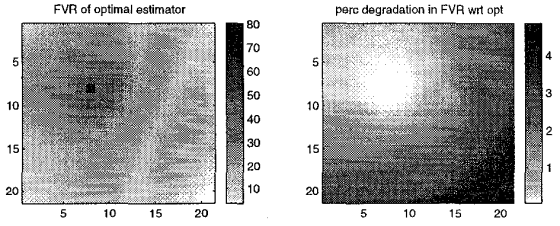

Fig. 5. Performance of multiscale estimator for 2-D diffusion, with boundary points are kept as linear functionals in the model. ($\beta = 1$, one measurement at location $(8,8)$). (a) Steady-state FVR of multiscale suboptimal estimator. (b) Percent degradation in FVR of suboptimal multiscale estimator with respect to the optimal.

putation, although the asymptotic complexity remains the same. Similar models have been successfully used in ocean hydrography applications where conventional estimation methods are infeasible due to the size of the problems.

It is a major challenge to design multiscale models for 2-D random fields that achieve better asymptotic computational complexity. Wavelet coefficients of the boundaries were used for modeling textures in [5]. Alternative approaches, such as canonical correlation realization (CCR) [4]—a singular value decomposition method—produce linear functionals shown in Fig. 6, that resemble the Fourier basis (which is reasonable, since a 2-D boundary resembles a diffusion process, which is diagonalized by the Fourier transform). Such approaches may lead to effective, reduced-order, multiscale state definitions for 2-D systems, and are currently under investigation.

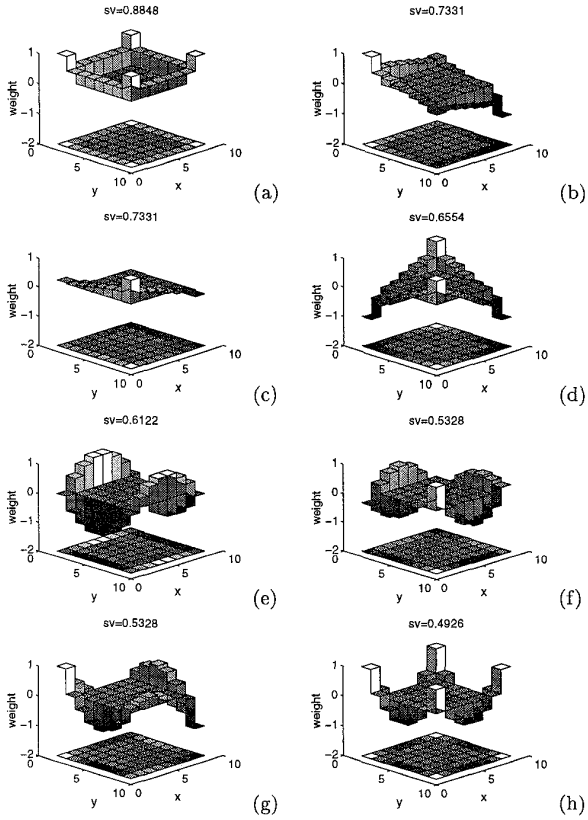

Fig. 6. The eight most significant linear functionals computed by CCR that decorrelate pixels at the "northeast" quadrant, $(1-8, 1-8)$, from the remaining pixels of a 16×16 2-D diffusion process in steady state. The significance (the singular value) associated with each linear functional is displayed above the plots.

6. REFERENCES

[1] T. M. Chin, W. C. Karl, and A. S. Willsky. "Probabilistic and Sequential Computation of Optical Flow Using Temporal Coherence". *IEEE Transactions on Image Processing*, 3(6):773–788, Nov. 1994.

[2] 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.

[3] 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.

[4] W. W. Irving. "A Canonical Correlations Approach to Multiscale Stochastic Realization". *IEEE Transactions on Automatic Control*, 1998. In press.

[5] 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.

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

[7] C. Wunsch. "Transient Tracers as a Problem in Control Theory". *Journal of Geophysical Research*, 93(C7):8099–8110, Jul. 1988.