A Canonical Correlations Approach to Multiscale Stochastic Realization

William W. Irving and Alan S. Willsky, *Fellow, IEEE*

***Abstract*—We develop a realization theory for a class of multiscale stochastic processes having white-noise driven, scale-recursive dynamics that are indexed by the nodes of a tree. Given the correlation structure of a 1-D or 2-D random process, our methods provide a systematic way to realize the given correlation as the finest scale of a multiscale process. Motivated by Akaike's use of canonical correlation analysis to develop both exact and reduced-order model for time-series, we too harness this tool to develop multiscale models. We apply our realization scheme to build reduced-order multiscale models for two applications, namely linear least-squares estimation and generation of random-field sample paths. For the numerical examples considered, least-squares estimates are obtained having nearly optimal mean-square errors, even with multiscale models of low order. Although both field estimates and field sample paths exhibit a visually distracting blockiness, this blockiness is not an important issue in many applications. For such applications, our approach to multiscale stochastic realization holds promise as a valuable, general tool.**

***Index Terms*—Least squares methods, realization theory, simulation, singular value decomposition, stochastic processes.**

I. Introduction

In this paper, we exploit well-established control concepts to develop a realization theory for the class of multiscale stochastic processes introduced in [4], [5]. These processes have white-noise driven, scale-recursive dynamics, directly analogous to the time-recursive dynamics of Gauss-Markov time-series models. Experimental and theoretical results have demonstrated that this class of processes is quite rich statistically; the self-similarity of fractional Brownian motion can be represented [5], as can any given 1-D wide-sense (WS) reciprocal process or 2-D Markov random field (WSMRF) [18].[1] Complementing this statistical richness are the efficient image processing algorithms to which multiscale models lead. For example, a scale-recursive algorithm has been developed that directly generalizes the Kalman filter and related Rauch–Tung–Striebel (RTS) smoother [5]. This algorithm incorporates noisy measurements of a given multiscale process to calculate both smoothed estimates *and* associated error covariances. Another algorithm has been developed for likelihood calculation [17]. In contrast to traditional 2-D optimal estimation formulations based on Markov random fields, which lead to iterative algorithms having a per-pixel computational complexity that typically grows with image size, these multiscale algorithms are noniterative and have a per-pixel complexity independent of image size [18], [19]. Substantial computational savings can thus result, as evidenced by exploitation of the multiscale framework in many applications, including computer vision (e.g., calculation of optical flow [19]), remote sensing (e.g., optimal interpolation of sea level variations in the North Pacific Ocean from satellite measurements, treating the ocean as either static [11], or more recently as dynamic [12]) and geophysics (e.g., characterizing spatial variations in hydraulic conductivity from both point and nonlocal measurements [6]).

In exactly the same way that Kalman filtering requires the prior specification of a state-space model before least-square estimation or likelihood calculation can be carried out, so does multiscale statistical processing require such prior modeling. The techniques developed in this paper provide the needed modeling tools: given the correlation structure of a 1-D or 2-D random process,[2] these tools can be used to realize the correlation as the finest of a multiscale process. Because there is typically a conflict between model complexity and accuracy, we mainly focus on the case where a constraint is imposed on the allowed model state dimension; the objective then is to build a model whose finest-scale correlation structure best matches the desired correlation, subject to the dimension constraint.

Our focus on realizing *finest*-scale statistics is motivated by the not insignificant class of applications in which the finest scale is really the only one of interest. For instance, in many image-processing applications (such as de-noising problems [15], or terrain segmentation [17]), the finest-scale process is a pixel-by-pixel representation of the image, the measurements are noisy observations of some subset of the pixels, and the objective is either to estimate the value of each image pixel [15] or to calculate the likelihood of the observations [17]. Similarly, in distributed parameter estimation problems (such as those arising in remote sensing [12] or geophysics [6]), the finest-scale process is a sampled version of the parameter of interest (e.g., hydraulic conductivity [6]), the measurements are noisy observations related to the parameter, and the objective

Manuscript received December 2, 1996; revised April 29, 1997, February 20, 1998, and November 27, 2000. Recommended by Associate Editor J. Spall. This work was supported by the Air Force Office of Scientific Research, under Grant AFOSR-F496-20-93-1-0604 and the Army Research Office under Grant ARO DAAL03-92-G-0015.

W. W. Irving is with Fidelity Investments, Merrimack, NH 03054 USA (e-mail: William.Irving@FMR.COM).

A. S. Willsky is with the Department of Electrical Engineering and Computer Science, the Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: willsky@mit.edu).

Publisher Item Identifier S 0018-9286(01)09591-5.

[1]The definition and properties of wide-sense reciprocal processes and WSMRFs are nicely summarized in [8].

[2]The terminology 1-D, 2-D, or multidimensional random process is used here to indicate that the dimension of the independent variable of the process is 1-D, 2-D, or multidimensional.

is to estimate the parameter. For all of these problems, the multiscale framework provides an efficient statistical approach for obtaining estimates and error covariances or calculating likelihoods, even though every other aspect of these problems involves only the finest scale.

Our approach to multiscale stochastic realization takes advantage of the close relationship the problem shares with its more traditional, time-series counterpart. To clarify this relationship, we first observe that each node in a q-th order tree has q children and a single parent, and hence partitions the remaining nodes into $(q+1)$ subtrees, one associated with each of these child and parent nodes. (A second-order tree, which is often used to index multiscale representations of time series, is illustrated in Fig. 1.) Related to this partitioning property is the Markov property that multiscale processes possess: if $x(s)$ is the value of the state at node s, then conditioned on $x(s)$ the states in the corresponding $(q+1)$ subtrees of nodes extending away from s are uncorrelated [18]. The connection to the time-series realization problem is that in both contexts, the role of state information is to provide an information interface among subsets of the process. This interface must store just enough process information to make the corresponding process subsets conditionally uncorrelated. The key challenge is that while in the time-series case this interface is between two subsets of the process (i.e., the past and the future), in the multiscale case, the interface is among *multiple* (i.e., $(q+1)$) subsets of the process.

We exploit the connection between the time-series and multiscale realization problems by adapting to the multiscale context work done in [1] and [2] on reduced-order time-series modeling. The work in [1] and [2] addresses two issues. First, for exact realizations, a method is devised for finding the minimal dimension and corresponding information content of the state. Second, for reduced-order, approximate realizations, a method is devised for measuring the relative importance of the components of the information interface provided by the state, so that a decision can be made about which components to discard in a reduced-order realization. The latter is accomplished using a classical tool from multivariate statistics, namely *canonical correlation analysis* [13]. We decompose our multiscale problem of decorrelating jointly $(q+1)$ process subsets into a collection of q problems of decorrelating pairs of process subsets. We then demonstrate that with respect to a particular decorrelation metric, canonical correlation analysis can in principle be used to solve optimally each of the pairwise decorrelation problems. Furthermore, these pairwise solutions can be concatenated to yield a sub-optimal solution to the multi-way decorrelation problem. The solution to this decorrelation problem leads readily to values for all the multiscale model parameters.

We apply our realization scheme to build reduced-order multiscale models for two applications, namely linear least-squares estimation and generation of random-field sample paths. For the numerical examples considered, we obtain least-squares estimates having mean-square errors that are nearly optimal, even with multiscale models of very low order. Although both field estimates and field sample paths exhibit a visually distracting blockiness, this blockiness is not really an issue in many applications. For such applications, our approach to multiscale realization holds promise as a valuable, general tool.

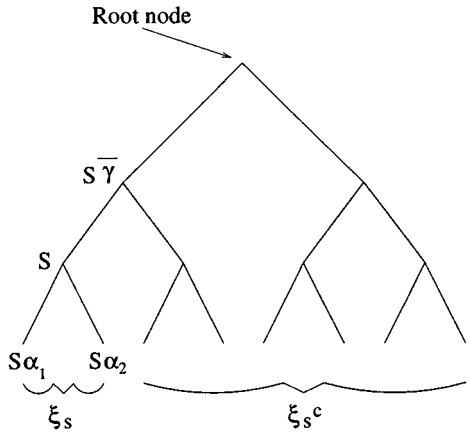


Fig. 1. The first four levels of a second-order tree are shown. The parent of node s is denoted by $s\bar{\gamma}$ and the two offspring are denoted by $s\alpha_1$ and $s\alpha_2$. The random vectors ξ_s and ξ_{s^c} contain, respectively, the finest-scale state information that does and does not descend from the node s.

The remainder of this paper is organized in the following way. In Section II, the multiscale framework is more formally introduced, and a measure of decorrelation is defined. In Section III, the modeling problem is precisely formulated, and a solution is presented for the case that full-order, exact models are sought. Then, in Section IV, the solution to the modeling problem is presented for the more challenging case that reduced-order, approximate models are sought. In Section V, two numerical examples are presented, and finally in Section VI, a summary is provided. Details of the proofs are relegated to Appendices A–C at the end of the paper.

II. Preliminaries

A. State-Space Models on qth-Order Trees

The models introduced in [4], [5], [19] describe multiscale stochastic processes indexed by nodes on a *tree*. A qth-order tree is a pyramidal structure of nodes connected such that each node (other than the leaf nodes) has q offspring nodes. We associate with each node s a vector-valued state $x(s)$, where in general, the q^m state vectors at the mth level of the tree (for $m = 0, 1, \ldots$) can be interpreted as information about the m-th scale of the process. In keeping with the conventions established in [4], [5], we define an upward (fine-to-coarse) shift operator $\bar{\gamma}$ such that $s\bar{\gamma}$ is the *parent* of node s, and a corresponding set of downward (coarse-to-fine) shift operators $\alpha_i,\ i = 1, 2, \ldots, q$, such that the q offspring of node s are given by $s\alpha_1, s\alpha_2, \ldots, s\alpha_q$. Fig. 1 depicts an example of the relative locations of s, $s\bar{\gamma}$, and $s\alpha_1$, $s\alpha_2$ in a second-order tree.

The dynamics implicitly providing a statistical characterization of $x(s)$ have the form of an autoregression in scale

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

This regression is initialized at the root node, $s = 0$, with a state variable $x(0)$ having mean zero and covariance $P(0)$. The term $w(s)$ represents white driving noise, uncorrelated across scale and space, and also uncorrelated with the initial condition $x(0)$; this noise is assumed to have mean zero and covariance $Q(s)$. Since $x(0)$ and $w(s)$ are zero-mean, it follows that $x(s)$

is a zero-mean random process.[3] Furthermore, since the driving noise $w(s)$ is white, the correlation structure of the process $x(s)$ is characterized completely by $P(0)$ and the auto-regression parameters $A(s)$ and $Q(s)$ for all nodes $s \neq 0$.

The statistical structure of multiscale processes can be exploited to yield an extremely efficient algorithm for estimating $x(\cdot)$, based upon noisy observations $y(\cdot)$. These observations take the form

$$y(s) = C(s)x(s) + v(s)$$

where the noise $v(s)$ is white, has covariance $R(s)$, and is uncorrelated with $x(\cdot)$ at all nodes on the tree. Just like the Kalman filter and the RTS smoother, this estimation algorithm has a recursive structure, and yields both state estimates and associated error covariances. For a multiscale process having states of dimension λ and indexed on a tree with N nodes, the number of required computations is $\mathcal{O}(N\lambda^3)$. Thus, the algorithm is quite efficient, particularly when the dimension of the state vectors is low.

B. Markov Property of Multiscale Processes

Multiscale processes possess an important Markov property, stemming directly from the whiteness of $w(s)$. To describe the form of this property of interest to us, we associate with each tree node s a set $\mathcal{F}_s$, where $\mathcal{F}_s$ contains all of the finest-scale nodes that descend from s. We also associate with each node s the random vectors ξ_s and ξ_{s^c}. The random vector ξ_s contains the elements of the set $\{x(\sigma); \sigma \in \mathcal{F}_s\}$, stacked into a vector, while ξ_{s^c} contains the elements of the complementary set $\{x(\sigma); \sigma \in \mathcal{F}_0\} - \{x(\sigma); \sigma \in \mathcal{F}_s\}$, stacked into a vector. These conventions are illustrated in Fig. 1. It will sometimes prove convenient to denote ξ_{s^c} by $\xi_{s\alpha_{q+1}}$; we freely use both forms. To relate ξ_σ and $x(s)$, we introduce the matrix $H_{\sigma|s}$, where $H_{\sigma|s}x(s)$ is the linear least-squares estimate of ξ_σ, given $x(s)$, and σ is a node or represents a set of nodes.

The Markov property, as it relates explicitly to the finest scale, can now be stated as follows [7]:

$$\xi_0 = \begin{pmatrix} \xi_{s\alpha_1} \\ \xi_{s\alpha_2} \\ \vdots \\ \xi_{s\alpha_{q+1}} \end{pmatrix} = \begin{pmatrix} H_{s\alpha_1|s} \\ H_{s\alpha_2|s} \\ \vdots \\ H_{s\alpha_{q+1}|s} \end{pmatrix} x(s) + \begin{pmatrix} \tilde{\xi}_{s\alpha_i|s} \\ \tilde{\xi}_{s\alpha_2|s} \\ \vdots \\ \tilde{\xi}_{s\alpha_{q+1}|s} \end{pmatrix} \tag{2}$$

with

$$x(s), \tilde{\xi}_{s\alpha_1|s}, \tilde{\xi}_{s\alpha_2|s}, \ldots, \tilde{\xi}_{s\alpha_{q+1}|s} \text{ uncorrelated.} \tag{3}$$

We use this property to relate the dimension of $x(s)$ to the correlation among the vectors $\xi_{s\alpha_1}, \xi_{s\alpha_2}, \ldots, \xi_{s\alpha_{q+1}}$. Toward this end, (2) and (3) together imply that

$$P_{\xi_{s\alpha_i}\xi_{s\alpha_j}} = H_{s\alpha_i|s} P_{x(s)} H^T_{s\alpha_j|s} \quad (i \neq j). \tag{4}$$

(Here and elsewhere, we adhere to the notational convention that P_x is the covariance of random vector x and P_{xy} is the cross-covariance of random vectors x and y). By elementary linear algebra [22], the rank of the cross-covariance in (4) is upper-bounded by the rank of $P_{x(s)}$, which in turn is upper-bounded by the dimension of $x(s)$. The following proposition thus follows.

Proposition 1:

$$\text{dimension}(x(s)) \geq \max_{i \neq j} \text{rank} \left(P_{\xi_{s\alpha_i}\xi_{s\alpha_j}} \right).$$

If the finest-scale covariance P_{ξ_0} must match exactly some prespecified covariance, then this proposition provides a lower bound on the required multiscale state dimension. In the rather likely case that the involved cross-covariance matrices have full rank, this dimension constraint becomes quite stringent. Thus, to keep the multiscale estimation algorithm computationally efficient, we find considerable motivation to turn to reduced-order (approximate) realizations.

C. The Generalized Correlation Coefficient

For the purposes of developing reduced-order models, it will prove convenient to have a scalar measure of the correlation among a collection of random vectors. We thus introduce a so-called *generalized correlation coefficient*. In keeping with standard conventions, we define as follows the correlation coefficient $\rho(\eta_1, \eta_2)$ between two scalar valued random variables η_1 and η_2:

$$\rho(\eta_1, \eta_2) \equiv \begin{cases} \frac{P_{\eta_1\eta_2}}{\sqrt{P_{\eta_1} P_{\eta_2}}} & \text{if } P_{\eta_i} > 0, \text{ for } i = 1, 2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Then, for a pair of vector-valued random variables η_1 and η_2, we define their generalized correlation coefficient $\bar{\rho}(\eta_1, \eta_2)$ by

$$\bar{\rho}(\eta_1, \eta_2) \equiv \max_{f_1, f_2} \left\{ \rho \left(f_1^T \eta_1, f_2^T \eta_2 \right) \right\}$$

where the dummy argument f_i (for $i = 1, 2$) is a column vector having the same dimension as η_i. Finally, we extend the definition of $\bar{\rho}(\cdot, \cdot)$ to a collection of random vectors $\eta_1, \eta_2, \ldots, \eta_k$, in the following way:

$$\bar{\rho}(\eta_1, \eta_2, \ldots, \eta_k) \equiv \max_{i \neq j} \bar{\rho}(\eta_i, \eta_j).$$

Each of these correlation coefficients has a conditioned version. To define them, we first let random vector z contain the conditioning information. Also, we let $\tilde{\eta}_i \equiv \eta_i - \hat{E}(\eta_i|z)$, where (here and elsewhere) we adhere to the convention that $\hat{E}(x|y)$ is the linear least-squares estimate of x given y. Finally, we define

$$\bar{\rho}(\eta_1, \eta_2, \ldots, \eta_k|z) \equiv \bar{\rho}(\tilde{\eta}_1, \tilde{\eta}_2, \ldots, \tilde{\eta}_k). \tag{6}$$

III. Formulation and Initial Investigation of Realization Problem

The realization problem of interest in this paper is to build a multiscale model, indexed on a given tree structure, to realize some prespecified, finest-scale covariance. We begin with a random vector χ_0, having the prespecified covariance P_{χ_0} and having an established correspondence with the finest scale of the given tree. For example, χ_0 might be a random field (written for

[3]The mean of $x(\cdot)$ can be set to any arbitrary value, by suitably adjusting the mean of $x(0)$ and $w(\cdot)$.

simplicity as a vector), with the finest scale of the tree (e.g., a quadtree) being a pixel-by-pixel representation of the field. Our objective is to specify values for the model parameters $P(0)$, $A(\cdot)$ and $Q(\cdot)$, to achieve the best match possible between the actual, realized covariance P_{ξ_0} and the desired covariance P_{χ_0}. In making the specification, we impose on each node s a dimension constraint λ_s

$$\text{dimension}\,(x(s)) \leq \lambda_s. \tag{7}$$

With the dimension constraint imposed, Proposition 1 shows that it may not be possible to achieve strict equality between P_{ξ_0} and P_{χ_0}, thus justifying our distinct notation for each. For convenience, we also introduce the random vectors χ_s and χ_{s^c}. These are defined to have the same relation to χ_0 as ξ_s and ξ_{s^c} have to ξ_0. To be more precise, suppose that the ith component of ξ_s (ξ_{s^c}) maps to the $n_s(i)$th ($n_{s^c}(i)$th) component of ξ_0; then, the ith component of χ_s (χ_{s^c}) maps to the $n_s(i)$th ($n_{s^c}(i)$th) component of χ_0, where χ_s (χ_{s^c}) has the same dimension as ξ_s (ξ_{s^c}). It will sometimes prove convenient to denote χ_{s^c} by $\chi_{s\alpha_{q+1}}$; we freely use both forms.

A. Full-Order Realizations

When the dimension constraint is discarded, the realization problem becomes conceptually simpler and exact realizations (i.e., realizations for which $P_{\xi_0} = P_{\chi_0}$) become possible. We begin by analyzing this case.

A notable class of multiscale processes in the context of exact realizations is those in which each state variable $x(s)$ is a linear function of the finest-scale process

$$x(s) = W_s \xi_s. \tag{8}$$

A state vector $x(s)$ obeying this relationship can clearly be seen to represent an aggregate (coarse) description of the finest-scale process descending from s. We refer to the matrix W_s associated with node s as the node's *internal* matrix, and we refer to multiscale processes for which (8) holds $\forall\, s$ as *internal* realizations. The notion of internal stochastic realizations is standard in time-series analysis [16], [20], with our use of the concept representing a natural adaptation.

Our interest in internal realizations stems from the convenient fact that the model parameters $P(0)$, $A(\cdot)$, and $Q(\cdot)$ can be specified completely in terms of the internal matrices and the finest-scale covariance. In other words, once values values for the internal matrices have been determined, values for the model parameters $P(0)$, $A(\cdot)$ and $Q(\cdot)$ follow easily. To see this fact, we begin by substituting (8) evaluated at $s = 0$ into $P(0) = E\left[x(0)x^T(0)\right]$ and using the fact that $P_{\xi_0} = P_{\chi_0}$ to yield

$$P(0) = W_0 P_{\chi_0} W_0^T. \tag{9}$$

The parameters $A(s)$ and $Q(s)$ can then be computed by noting that (1) represents the linear least-squares prediction of $x(s)$ based upon $x(s\bar{\gamma})$, plus the associated prediction error

$$x(s) = \hat{E}\left[x(s)|x(s\bar{\gamma})\right] + w(s). \tag{10}$$

Comparing (1) and (10), and using standard results from linear least-squares estimation, the model parameters can be seen to satisfy the following relations:

$$A(s) = P_{x(s)x(s\bar{\gamma})} P^{-1}_{x(s\bar{\gamma})} \tag{11a}$$

$$Q(s) = P_{x(s)} - P_{x(s)x(s\bar{\gamma})} P^{-1}_{x(s\bar{\gamma})} P^T_{x(s)x(s\bar{\gamma})}. \tag{11b}$$

Finally, again using (8) and the facts that $P_{\xi_s} = P_{\chi_s}$ and $P_{\xi_s\xi_{s\bar{\gamma}}} = P_{\chi_s\chi_{s\bar{\gamma}}}$, the covariances appearing in (11a) and (11b) can be expressed as simple functions of the internal matrices and the finest-scale covariance

$$P_{x(s)x(s\bar{\gamma})} = W_s P_{\chi_s\chi_{s\bar{\gamma}}} W^T_{s\bar{\gamma}} \tag{12a}$$

$$P_{x(s)} = W_s P_{\chi_s} W_s^T. \tag{12b}$$

A natural question at this point is how to devise the internal matrices for an exact, internal realization of a given finest-scale covariance. At the finest-scale nodes, the internal matrices are easy to determine; they are implicitly defined by the association between finest-scale nodes and the components of χ_0. For example, if each scalar component of χ_0 is assigned to a distinct finest-scale node, then clearly $W_s = 1$ for each finest-scale node. At any coarse-scale node s (i.e., any nodes above the finest scale), a necessary condition for (2) and (3) to hold in an exact, internal model is that W_s fulfill the following decorrelating role:

$$\bar{\rho}\left(\chi_{s\alpha_1}, \chi_{s\alpha_2}, \ldots, \chi_{s\alpha_{q+1}} | W_s\chi_s\right) = 0. \tag{13}$$

The necessity of (13) can be justified in the following manner:

$$\begin{aligned}\bar{\rho}\left(\chi_{s\alpha_1}, \ldots, \chi_{s\alpha_{q+1}} | W_s\chi_s\right) &= \bar{\rho}\left(\xi_{s\alpha_1}, \ldots, \xi_{s\alpha_{q+1}} | x(s)\right) \\ &= \bar{\rho}\left(\tilde{\xi}_{s\alpha_1|s}, \ldots, \tilde{\xi}_{s\alpha_{q+1}}\right) = 0.\end{aligned}$$

Here, the first equality follows from the facts that the realization is exact and internal, the second equality follows from the definitions of $\tilde{\xi}_{s\alpha_i}$ and the conditional generalized correlation coefficient, and the final equality follows from (2) and (3).

As we now show, (13) is not only *necessary* for an exact, internal realization, but is also *sufficient* for building an exact realization.

Proposition 2: Suppose that for all nodes s, the W_s matrices satisfy (13), and that the multiscale system matrices $P(0)$, $A(s)$, and $Q(s)$ are defined in terms of the W_s matrices via (9) and

$$A(s) = W_s P_{\chi_s\chi_{s\bar{\gamma}}} W^T_{s\bar{\gamma}} \left(W_{s\bar{\gamma}} P_{\chi_{s\bar{\gamma}}} W^T_{s\bar{\gamma}}\right)^{-1} \tag{14a}$$

$$Q(s) = W_s P_{\chi_s} W_s^T - A(s) W_{s\bar{\gamma}} P_{\chi_{s\bar{\gamma}}\chi_s} W_s^T. \tag{14b}$$

Then, for any nodes s and t at the same scale (possibly with $s = t$)

$$P_{x(s)x(t)} = W_s P_{\chi_s\chi_t} W_t^T. \tag{15}$$

A proof of Proposition 2 is contained in Appendix A. The following special case of the proposition is of particular interest.

Corollary 1: If the conditions in Proposition 2 hold, then $P_{\chi_0} = P_{\xi_0}$.

Proof: For any s and t at the finest scale, the internal matrices W_s and W_t are identity matrices, and hence (15) implies that $P_{x(s)x(t)} = P_{\chi_s\chi_t}$. But also, by definition, $P_{x(s)x(t)} =$

$P_{\xi_s\xi_t}$. Thus, $P_{\chi_s\chi_t} = P_{\xi_s\xi_t}$, and the corollary immediately follows. **QED**

To summarize, there is a three-stage procedure for realizing exactly any desired finest-scale covariance: (i) establish a correspondence between finest-scale nodes and components of the vector χ_0, thereby implicitly specifying W_s for each finest-scale node, (ii) find a matrix W_s satisfying (13) for every coarse-scale node,[4] and finally (iii) substitute the resulting W_s values into (9), (14a), and (14b) to calculate values for $P(0)$, $A(\cdot)$, and $Q(\cdot)$. An attractive feature of this procedure is that it decomposes the realization problem into a collection of independent subproblems, each myopically focused on determining the information content of the state vector at a single node to fulfill the decorrelating role (13). In the case that χ_0 is either a wide-sense, bilateral Markov random process or aWSMRFdefined on a discrete lattice, this realization approach is particularly attractive, since the W_s matrices can be determined by inspection; this fact is proved by construction in [18], which provides a nice, concrete illustration of the ideas presented in this section. We hasten to add, however, that even in the WSMRF case, the state vectors for an exact realization will typically have an impractically high dimension, and thus the construction is mainly of interest for motivating our approach to reduced-order modeling.

B. Reduced-Order Realizations

For the rest of the paper, we reinstate the constraint (7) on multiscale model state dimension. With this constraint in effect, we no longer look for W_s matrices that fulfill the decorrelation condition (13) exactly; instead, we look for matrices that do the best decorrelation job possible, subject to the dimension constraint. Specifically, we seek W_s matrices satisfying

$$W_s = \arg \min_{W \in \mathcal{M}_{\lambda_s}} \bar{\rho}\left(\chi_{s\alpha_1}, \chi_{s\alpha_2}, \ldots, \chi_{s\alpha_{q+1}} | W\chi_s\right) \quad (16)$$

where $\mathcal{M}_{\lambda_s}$ is the set of matrices having λ_s or fewer rows (and a number of columns implicitly defined by context). We refer to (16) as the *decorrelation problem*. Once values for the W_s matrices have been found, we mimic our approach to the full-order realization problem: values for the multiscale parameters $P(0)$, $A(\cdot)$ and $Q(\cdot)$ are set using (9), (14a), and (14b). Thus, our reduced-order modeling procedure is identical to our three-stage, full-order modeling procedure (see Section III-A), except that now condition (16) is used in lieu of condition (13).

There are several comments to make about this modeling approach. First, the approach shares with its full-order counterpart the computational benefit that we can find all the model parameters in a single sweep from coarse to fine scales, determining W_s for each node as we go along, and thereby implicitly specifying $P(0)$, $A(\cdot)$ and $Q(\cdot)$. We emphasize, though, that the condition (16) is a heuristic one. Certainly, this condition is reasonable, from a myopic, node-by-node view of the realization problem; however, the condition *does not* provide tight control over the overall match between P_{ξ_0} and P_{χ_0}. Indeed, an open research challenge is to find a way to build a reduced-order model, in which the parameters $P(0)$, $A(\cdot)$ and $Q(\cdot)$ are chosen explicitly to minimize some global measure of the discrepancy between P_{χ_0} and P_{ξ_0}. This problem appears to be very challenging. We will focus only on the more myopic problem of solving (16).

As an additional comment, models constructed with our approach will not in general be internal realizations (i.e., (8) will not hold in general). Consequently, the W_s matrices should be viewed in general as merely auxiliary constructs that aid in setting values for the parameters $P(0)$, $A(\cdot)$ and $Q(\cdot)$.

Finally, the definition of the generalized correlation coefficient makes it clear that for any given matrix W_s

$$\bar{\rho}\left(\chi_{s\alpha_1}, \chi_{s\alpha_2}, \ldots, \chi_{s\alpha_{q+1}} | W_s\chi_s\right) = \bar{\rho}(\chi_{s\alpha_1}, \chi_{s\alpha_2}, \ldots, \chi_{s\alpha_{q+1}} | W_s'\chi_s)$$

where W_s' is a matrix whose rows form an orthonormal basis for the row space of W_s. Hence, defining the set $\mathcal{N}_\lambda$ to be the subset of $\mathcal{M}_\lambda$ having orthonormal rows, we see that

$$\min_{W \in \mathcal{M}_{\lambda_s}} \bar{\rho}\left(\chi_{s\alpha_1}, \chi_{s\alpha_2}, \ldots, \chi_{s\alpha_{q+1}} | W\chi_s\right) = \min_{W \in \mathcal{N}_{\lambda_s}} \bar{\rho}(\chi_{s\alpha_1}, \chi_{s\alpha_2}, \ldots, \chi_{s\alpha_{q+1}} | W\chi_s).$$

Thus, without loss of optimality, we can replace the constraint set $\mathcal{M}_{\lambda_s}$ in (16) with the set $\mathcal{N}_{\lambda_s}$. When convenient, we will freely make this replacement.

IV. Decorrelating Sets of Random Vectors

A. Decorrelating a Pair of Random Vectors

We here analyze a special case of the decorrelation problems in which there are only two vectors to decorrelate. Denoting these vectors by η_1 and η_2 and stacking them as $\eta = \begin{pmatrix} \eta_1^T & \eta_2^T \end{pmatrix}^{\mathrm{T}}$, our objective is to find the optimal matrix solution to the following optimization problem:

$$W = \arg \min_{W \in \mathcal{M}_\lambda} \bar{\rho}(\eta_1, \eta_2 | W\eta_1). \quad (17)$$

Playing a central role in the solution is a standard result from canonical correlation theory. For the purposes of stating this result precisely, we denote the rank of the $n_i \times n_i$ covariance matrix P_{η_i} by m_i (for $i = 1, 2$), and the rank of $P_{\eta_1\eta_2}$ by m_{12}. Also, we let I_n be an identity matrix having n rows and columns.

Theorem 1: There exist matrices T_1 and T_2, of dimension $m_1 \times n_1$ and $m_2 \times n_2$, respectively, such that

$$\underbrace{\begin{pmatrix} T_1 & \mathbf{0} \\ \mathbf{0} & T_2 \end{pmatrix}}_{T} \begin{pmatrix} P_{\eta_1} & P_{\eta_1\eta_2} \\ P_{\eta_1\eta_2}^T & P_{\eta_2} \end{pmatrix} \begin{pmatrix} T_1 & \mathbf{0} \\ \mathbf{0} & T_2 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} I_{m_1} & D \\ D^T & I_{m_2} \end{pmatrix}$$

and

$$\begin{pmatrix} T_1^+ & \mathbf{0} \\ \mathbf{0} & T_2^+ \end{pmatrix} \begin{pmatrix} I_{m_1} & D \\ D^T & I_{m_2} \end{pmatrix} \begin{pmatrix} T_1^+ & \mathbf{0} \\ \mathbf{0} & T_2^+ \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} P_{\eta_1} & P_{\eta_1\eta_2} \\ P_{\eta_1\eta_2}^T & P_{\eta_2} \end{pmatrix}.$$

The matrix D has dimension $m_1 \times m_2$ and is given by $D = \mathrm{diag}\left(\hat{D}, \mathbf{0}\right)$, where $\hat{D} = \mathrm{diag}(d_1, d_2, \ldots, d_{m_{12}})$, $1 \geq d_1 \geq d_2 \geq \cdots \geq d_{m_{12}} > 0$; for a given $(P_{\eta_1}, P_{\eta_2}, P_{\eta_1\eta_2})$, the matrix

[4]The choice $W_s = I$, so that $x(s) = \xi_s$, is universally valid, though of very little practical use, owing to the high dimension for $x(s)$ to which it leads.

$\hat{D}$ is unique. Finally, T_i^+ is the Moore–Penrose pseudoinverse of T_i, and is given by $T_i^+ = P_{\eta_i} T_i^T$, $(i = 1, 2)$.

We refer to the triple of matrices (T_1, T_2, D) as the *canonical correlation matrices* associated with (η_1, η_2). For convenience, we introduce truncated versions of T_i (for $i = 1, 2$), denoted by $T_{i,k}$ and defined to contain the first k rows of T_i; as a special case, we define $\hat{T}_i$ to contain the first m_{12} rows of T_i. Results very similar to Theorem 1 can be found in several places, including [3] and [9]. A proof of the theorem, as exactly stated here, can be found in [14]. As these proofs reveal, the calculation of the canonical correlation matrices can be carried out in a numerically sound fashion using the singular value decomposition; this calculation requires $\mathcal{O}(N^3)$ floating point operations, where $N = \max(n_1, n_2)$.

We use Theorem 1 to perform a change of basis on the vectors η_1 and η_2, to simplify maximally the correlation between them, and thus to simplify analysis of the decorrelation problem. Proceeding, we define the random vectors μ, μ_1 and μ_2 via

$$\mu \equiv \begin{pmatrix} \mu_1^T & \mu_2^T \end{pmatrix}^T, \quad \mu_i = T_i \eta_i, \quad (i = 1, 2)$$

where thanks to Theorem 1, μ_1 and μ_2 have covariance

$$P_\mu = \begin{pmatrix} P_{\mu_1} & P_{\mu_1\mu_2} \\ P_{\mu_1\mu_2}^T & P_{\mu_2} \end{pmatrix} = \begin{pmatrix} I_{m_1} & D \\ D^T & I_{m_2} \end{pmatrix}$$

and the transformation from (η_1, η_2) to (μ_1, μ_2) is invertible in a mean-square sense

$$E\left[\left(\eta_i - T_i^+ \mu_i\right)\left(\eta_i - T_i^+ \mu_i\right)^T\right] = \mathbf{0}, \quad (i = 1, 2).$$

The following lemma now provides the key simplification.

Lemma 1:

$$\bar{\rho}(\eta_1, \eta_2 | WT\eta) = \bar{\rho}(\mu_1, \mu_2 | W\mu) \tag{18a}$$

$$\bar{\rho}(\eta_1, \eta_2 | W\eta) = \bar{\rho}\left(\mu_1, \mu_2 | WT^+\mu\right) \tag{18b}$$

$$\min_{W \in \mathcal{M}_\lambda} \bar{\rho}(\eta_1, \eta_2 | W\eta) = \min_{V \in \mathcal{N}_\lambda} \bar{\rho}(\mu_1, \mu_2 | V\mu). \tag{18c}$$

As a special case of (18a) and (18b), we note that $\bar{\rho}(\eta_1, \eta_2) = \bar{\rho}(\mu_1, \mu_2)$. The lemma is a direct consequence of the definition of the generalized correlation coefficient, together with Theorem 1; we omit the details of the proof.

Equipped with the foregoing theorem and lemma, we can now solve (17).

Proposition 3: For $0 \leq \lambda < m_{12}$ and for $i = 1, 2$

$$\min_{W \in \mathcal{M}_\lambda} \bar{\rho}(\eta_1, \eta_2 | W\eta_1) = \bar{\rho}(\eta_1, \eta_2 | T_{1,\lambda}\eta_1) = d_{\lambda+1}. \tag{19a}$$

For $\lambda \geq m_{12}$

$$\min_{W \in \mathcal{M}_\lambda} \bar{\rho}(\eta_1, \eta_2 | W\eta_1) = \bar{\rho}\left(\eta_1, \eta_2 | \hat{T}_1 \eta_1\right) = 0. \tag{19b}$$

Proof: In Appendix B, we demonstrate that for $\lambda < m_{12}$

$$\min_{W \in \mathcal{N}_\lambda} \bar{\rho}(\mu_1, \mu_2 | W\mu_1) = \bar{\rho}\left(\mu_1, \mu_2 | \begin{pmatrix} I_\lambda & \mathbf{0} \end{pmatrix} \mu_1\right) = d_{\lambda+1} \tag{20a}$$

while for $\lambda \geq m_{12}$

$$\min_{W \in \mathcal{N}_\lambda} \bar{\rho}\left(\mu_1, \mu_2 | W\mu_1\right) = \bar{\rho}(\mu_1, \mu_2 | \begin{pmatrix} I_{m_{12}} & \mathbf{0} \end{pmatrix} \mu_1) = 0. \tag{20b}$$

Once these facts are established, the results (19a) and (19b) then follow. In particular, with regard to (19a), we have the following sequence of identities:

$$\begin{aligned} &\min_{W \in \mathcal{M}_\lambda} \bar{\rho}(\eta_1, \eta_2 | W\eta_1) \\ &\quad = \min_{W \in \mathcal{N}_\lambda} \bar{\rho}(\mu_1, \mu_2 | W\mu_1) = \bar{\rho}(\mu_1, \mu_2 | \begin{pmatrix} I_\lambda & \mathbf{0} \end{pmatrix} \mu_1) \\ &\quad = \bar{\rho}(\eta_1, \eta_2 | \begin{pmatrix} I_\lambda & \mathbf{0} \end{pmatrix} T_1 \eta_1) = d_{\lambda+1}. \end{aligned} \tag{21}$$

The first equality follows from (18c), the second from (20a), the third from (18a) and the fourth from (20a). The result (19b) can be proved from (20b) in a very similar fashion; the details are omitted. **QED**

The important point to note about this proposition is that solving (17) is essentially a problem of calculating the canonical correlation matrices associated with (η_1, η_2). Indeed, (17) can be solved simultaneously for all values of λ by calculating just once these canonical correlation matrices.

B. Decorrelating Multiple Random Vectors

We now turn to the general decorrelation problem (16), for which we develop a suboptimal solution. This solution has an intuitively appealing structure motivated by the solution to the simpler problem (17). We emphasize that to the best of our knowledge, the task of characterizing the *optimal* solution to (16) is an unsolved problem.

Our approach is to decompose the decorrelation problem into a collection of q subproblems. In the ith subproblem, we focus on decorrelating $\chi_{s\alpha_i}$ from $\chi_{s\alpha_j}$ for all $j \neq i$; specifically, we exploit Proposition 3 to solve

$$W_{i,k_i} = \arg \min_{W \in \mathcal{M}_{k_i}} \bar{\rho}\left(\chi_{s\alpha_i}, \chi_{(s\alpha_i)^c} | W\chi_{s\alpha_i}\right) \tag{22}$$

where for now we treat $k_1, \ldots, k_q$ as free parameters. By choosing W_{i,k_i} as in (22), we effectively decorrelate $\chi_{s\alpha_i}$ from $\chi_{s\alpha_j}$ for all $j \neq i$ at once; in particular, it is clear that

$$\bar{\rho}\left(\chi_{s\alpha_i}, \chi_{s\alpha_j} | W_{i,k_i}\chi_{s\alpha_i}\right) \leq \bar{\rho}\left(\chi_{s\alpha_i}, \chi_{(s\alpha_i)^c} | W_{i,k_i}\chi_{s\alpha_i}\right), \quad j \neq i \tag{23}$$

and so, if the right side of (23) is small, then the left side will also be for all $j \neq i$.

To see how we combine $W_{1,k_1}, \ldots, W_{q,k_q}$ to solve (16) approximately, let us consider the quantity

$$\bar{\rho}\left(\chi_{s\alpha_1}, \ldots, \chi_{s\alpha_q} | W_{1,k_1}\chi_{s\alpha_1}, \ldots, W_{q,k_q}\chi_{s\alpha_q}\right) \tag{24}$$

which we can express more succinctly as $\bar{\rho}(\chi_{s\alpha_1}, \ldots, \chi_{s\alpha_q} | W_s(k_1, \ldots, k_q)\chi_s)$ by defining the block-diagonal matrix $W_s(k_1, \ldots, k_q) \equiv \operatorname{diag}(W_{1,k_1}, \ldots, W_{q,k_q})$. Since the ith block component of this matrix has been specially designed to decorrelate $\chi_{s\alpha_i}$ from $\chi_{s\alpha_j}$, $j \neq i$, we intuitively expect that all the block components will work together to make (24) small. Furthermore, if

$$\sum_{i=1}^{q} k_i \leq \lambda_s \tag{25}$$

then $W_s(k_1,\ldots,k_q) \in \mathcal{M}_{\lambda_s}$, implying that $W_s(k_1,\ldots,k_q)$ is in the feasible set of the optimization problem (16), and can indeed be used as an approximate solution to (16).

To characterize precisely the value of (24), we first must establish a result describing the nonincreasing nature of the generalized correlation coefficient as the amount of conditioning information increases.

Proposition 4:

$$\bar{\rho}(\eta_1,\eta_2|W_i\eta_i) \le \bar{\rho}(\eta_1,\eta_2), \quad i=1,2. \tag{26}$$

Proof: In Appendix C, we demonstrate that

$$\bar{\rho}\left(\mu_1,\mu_2|W_i\mu_i\right) \le \bar{\rho}\left(\mu_1,\mu_2\right). \tag{27}$$

Once this fact is established, the result (26) follows. In particular, we have the following sequence of relations:

$$\bar{\rho}\left(\eta_1,\eta_2|W_i\eta_i\right) = \bar{\rho}\left(\mu_1,\mu_2|W_iT_i^{+}\mu_i\right) \le \bar{\rho}\left(\mu_1,\mu_2\right) = \bar{\rho}\left(\eta_1,\eta_2\right).$$

The first relation follows from (18b), the second from (27) and the third from (18a). **QED**

We emphasize that if the conditioning information is not a function of either η_1 or η_2 alone, then the function $\bar{\rho}(\cdot,\cdot|\cdot)$ may become an increasing one. For instance, if

$$\begin{pmatrix} P_{\eta_1} & P_{\eta_1\eta_2} \\ P_{\eta_1\eta_2} & P_{\eta_2} \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

then $\bar{\rho}(\eta_1,\eta_2)=0.5$, but $\bar{\rho}(\eta_1,\eta_2|\eta_1+\eta_2)=1$.

We can, however, slightly strengthen Proposition 4 by relaxing our restriction that *all* of the conditioning information be a linear function of either η_1 or η_2; in lieu of this restriction, we restrict *each individual scalar component* of this conditioning information to be a function of either η_1 or η_2. We state this result as a corollary.

Corollary 2:

$$\bar{\rho}(\eta_1,\eta_2|W_1\eta_{i_1},W_2\eta_{i_2}) \le \bar{\rho}(\eta_1,\eta_2|W_1\eta_{i_1}), \quad (i_1,i_2)\in\{\{1,2\}\times\{1,2\}\}$$

Proof:

$$\begin{aligned}\bar{\rho}&(\eta_1,\eta_2|W_1\eta_{i_1},W_2\eta_{i_2}) \\ &= \bar{\rho}\Big(\eta_1-\hat{E}\left(\eta_1|W_1\eta_{i_1}\right),\eta_2-\hat{E}\left(\eta_1|W_1\eta_{i_1}\right)|W_2 \\ &\qquad \times\left(\eta_{i_2}-\hat{E}\left(\eta_{i_2}|W_1\eta_{i_1}\right)\right)\Big) \\ &\le \bar{\rho}\left(\eta_1-\hat{E}\left(\eta_1|W_1\eta_{i_1}\right),\eta_2-\hat{E}\left(\eta_1|W_1\eta_{i_1}\right)\right) \\ &= \bar{\rho}\left(\eta_1,\eta_2|W_1\eta_{i_1}\right).\end{aligned}$$

The first and third lines here represent direct applications of (6), while the second line represents application of Proposition 4. **QED**

Using this corollary, we now return to consideration of the value of (24).

Proposition 5:

$$\begin{aligned}\bar{\rho}&\left(\chi_{s\alpha_1},\ldots,\chi_{s\alpha_{q+1}}|W_s(k_1,\ldots,k_q)\chi_s\right) \\ &\le \max_{i=1,\ldots,q}\bar{\rho}\left(\chi_{s\alpha_i},\chi_{(s\alpha_i)^c}|W_s(k_1,\ldots,k_q)\chi_s\right) \quad (28a) \\ &\le \max_{i=1,\ldots,q}\bar{\rho}\left(\chi_{s\alpha_i},\chi_{(s\alpha_i)^c}|W_{s,k_i}\chi_{s\alpha_i}\right). \quad (28b)\end{aligned}$$

Proof: The first inequality in (28) is a direct consequence of the definition of the generalized correlation coefficient. The second is then a direct consequence of the corollary to Proposition 4. **QED**

The important point to note about this proposition is that $W_s(k_1,\ldots,k_q)$ leads to a value for the objective function in (16) that is no greater than the maximum of the values obtained in the q subproblems (22). In other words, by concatenating together the solutions to the q subproblems (22) into the block-diagonal matrix $W_s(k_1,\ldots,k_q)$, we obtain an approximate solution to (16) having a value upper-bounded by the maximum value of these q solutions to (22). This observation suggests a way to select values for the parameters $k_1,\ldots,k_q$. In particular, subject to the constraint (25), we should choose these parameters to fulfill the following minimax condition:

$$\begin{aligned}&\left(k_1^*,\ldots,k_q^*\right) \\ &= \arg\min_{k_1,\ldots,k_q}\left\{\max_{i=1,\ldots,q}\bar{\rho}\left(\chi_{s\alpha_i},\chi_{(s\alpha_i)^c}|W_{i,k_i}\chi_{s\alpha_i}\right)\right\}. \quad (29)\end{aligned}$$

By choosing the k_i parameters in this fashion, we minimize the right side of (28b), which upper-bounds the left side of (28a). The matrix $W_s\left(k_1^*,\ldots,k_q^*\right)$ then serves as a suboptimal solution to (16).

To describe the solution to (29), we denote by $(\hat{T}_{s\alpha_i},\hat{T}_{(s\alpha_i)^c},\hat{D}_{s,i})$ the canonical correlation matrices associated with $(\chi_{s\alpha_i},\chi_{(s\alpha_i)^c})$, where the diagonal elements of $\hat{D}_{s,i}$ are denoted by $d_1^{s,i},d_1^{s,i},\ldots$. For simplicity of exposition only, we assume that k_i is strictly less than the rank of the cross-covariance $P_{\chi_{s\alpha_i}\chi_{(s\alpha_i)^c}}$, for $i=1,\ldots q$. Then, thanks to Proposition 3, it follows that:

$$\bar{\rho}\left(\chi_{s\alpha_i},\chi_{(s\alpha_i)^c}|W_{i,k_i}\chi_{s\alpha_i}\right) = d_{k_i+1}^{s,i}.$$

Hence, the minimax definition (29) is equivalent to the following, where we again impose the constraint (25):

$$\left(k_1^*,\ldots,k_q^*\right) = \arg\min_{k_1,\ldots,k_q}\left\{\max_{i=1,\ldots,q} d_{k_i+1}^{s,i}\right\}.$$

This discrete optimization problem can easily be solved, once the canonical correlation quantities $\left(\hat{T}_{s\alpha_i},\hat{D}_{s,i}\right)$ associated with $\left(\chi_{s\alpha_i},\chi_{(s\alpha_i)^c}\right)$ have been calculated, for $i=1,2,\ldots,q$ [14].

C. Calculating the Canonical Correlation Matrices

For problems of practical interest to us, the dimension of χ_s and χ_{s^c} can be on the order of a thousand (or greater), thus prohibiting the exact calculation of the associated canonical correlation matrices $(\hat{T}_s,\hat{D}_s)$. However, if χ_0 is a WSMRF, then as we show in this section, the calculation of $\left(\hat{T}_s,\hat{D}_s\right)$ can be simplified drastically; the structure of this simplification is important, because it leads naturally to a computationally simple technique for calculating an approximation to $(\hat{T}_s,\hat{D}_s)$ in the case that χ_0 lacks the WSMRF structure.

Suppose that χ_0 is a first-order, scalar-valued WSMRF over a discrete lattice having dimensions 256×256. We focus on a particular node s for which χ_s and χ_{s^c} contain the values of the field at the subsets of points displayed in Fig. 2(a). Specifically, χ_s contains the values of the field at the 64 grid points marked with circles, both filled and not filled, in the white region, while

χ_{s^c} contains the values at the all other grid points; subsets of these other grid points are marked with squares, both filled and not filled. Also, let μ_s and μ_{s^c} contain the values of χ_s and χ_{s^c} at their respective boundary points, where these boundary points are marked with filled-in circles and squares, respectively. Finally, let Θ_s and Θ_{s^c} be the selection matrices[5] that select μ_s and μ_{s^c} from χ_s and χ_{s^c}, respectively:

$$\mu_s = \Theta_s \chi_s \text{ and } \mu_{s^c} = \Theta_{s^c} \chi_{s^c}. \tag{30}$$

The key point is that if $\left(\hat{T}_s^{\mu}, \hat{T}_{s^c}^{\mu}, \hat{D}^{\mu}\right)$ are the canonical correlation matrices for (μ_s, μ_{s^c}), then

$$\hat{T}_s = \hat{T}_s^{\mu} \Theta_s, \text{ and } \hat{D}_s = \hat{D}_s^{\mu}. \tag{31}$$

Since the dimension of μ_{s^c} is roughly 5×10^{-4} times the dimension of χ_{s^c}, this approach reduces the computational cost of determining $(\hat{T}_s, \hat{D}_s)$ by roughly a factor of 6×10^9.

For non-Markov random fields, we modify our approach, making the boundary region as thick as possible for each of χ_s and χ_{s^c}, subject to the constraint that the resulting vectors μ_s and μ_{s^c} have dimension no greater than some prescribed limit. Using the same graphical conventions as in Fig. 2(a), this idea is illustrated in Fig. 2(b), where 132 is the limiting dimension of both μ_s and μ_{s^c}. Once μ_s and μ_{s^c} have been defined, we proceed exactly as in the Markov case.

V. Numerical Examples

In this section, we present numerical examples that suggest the promise of our modeling approach. All models are indexed on quadtrees. Also, for the purposes of calculating the canonical correlation matrices (see discussion at end of Section IV-C), we choose the selection matrices Θ_s and Θ_{s^c} to have at most 260 rows.

A. Reduced-Order Representations of Isotropic Random Fields

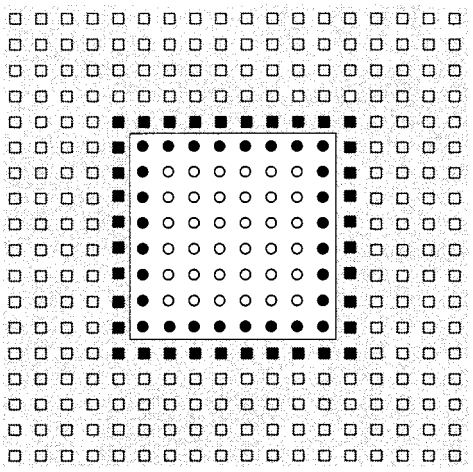
For our first example, we consider a scalar, wide-sense stationary, zero-mean, isotropic random field $y(m,n)$ that is of interest in the geological sciences [21]. We build multiscale models to realize the field (32) on a 128×128 grid. We build four models, each involving a different constraint on the state dimension; the state dimension is constrained to be no greater than the respective values 64, 32, 16, and 8.

The correlation function for this field can be expressed analytically as follows:

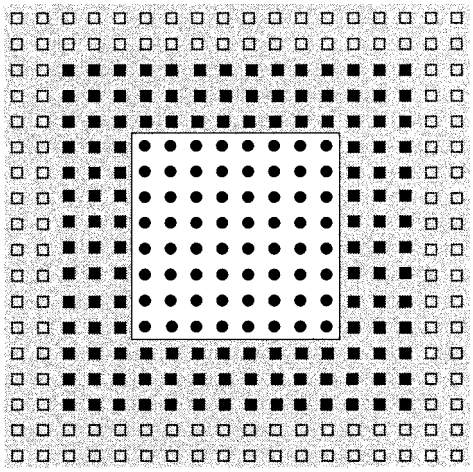
$$\begin{aligned} R_{yy}(i,j) &= E\left[y(m+i,n+j)y(m,n)\right] = R_{yy}(r) \\ &= \begin{cases} 1 - \frac{3}{2}\left(\frac{r}{\ell}\right) + \frac{1}{2}\left(\frac{r}{\ell}\right)^3 & 0 \le r \le \ell, \\ 0 & r > \ell\,, \end{cases} \end{aligned} \tag{32}$$

where $r = \sqrt{i^2 + j^2}$, and ℓ is the characteristic length of the function. A plot of this function for $\ell = 80$ is represented by the solid curve in Fig. 3; we see from this plot that there is significant long-range correlation, at least relative to the 128×128 grid we are using.

[5]A *selection* matrix consists solely of zeros and ones, with the additional restriction that each row have exactly one nonzero component and each column have at most one nonzero column.



(a)

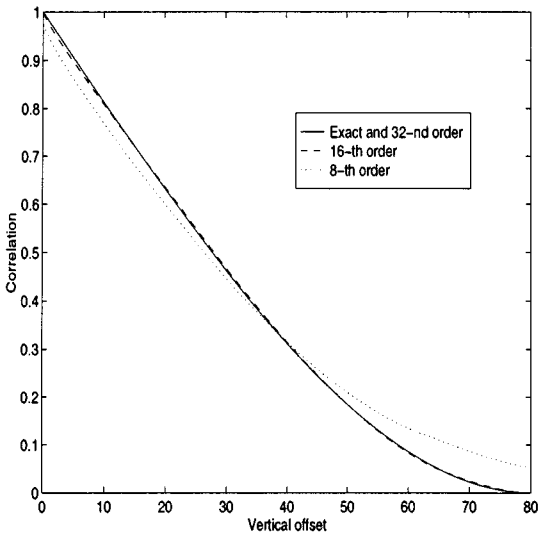


(b)

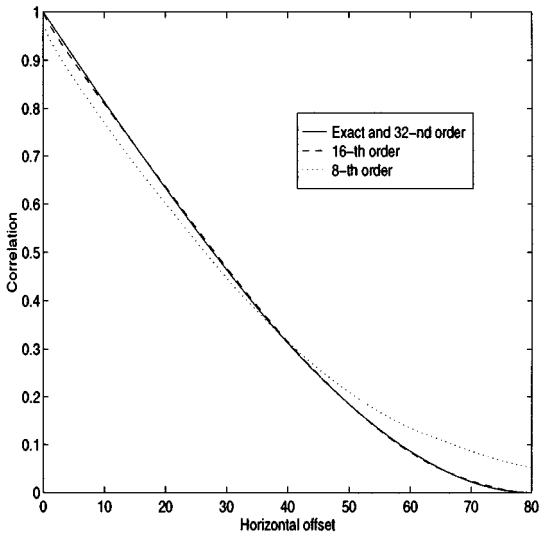
Fig. 2. Illustration of our approach to finding the canonical correlation matrices associated with (χ_s, χ_{s^c}) for (a) a first-order WSMRF, and (b) a non-Markov random field.

In Fig. 4(a), we display as a contour plot the exact correlation function (32). Then, in Fig. 4(b), we display as contour plots the correlation function associated with our multiscale model of order 8. Because our multiscale models have reduced order, they lead to correlation structures that are only approximately stationary, and thus we must define carefully what is being plotted in Fig. 4(b). Toward this end, we let ξ_0^8 denote the random vector comprising the finest-scale of a multiscale process in which the state vectors are constrained to have dimension no greater than 8. We denote the (i,j)th component of ξ_0^8 by $\xi_0^8(i,j)$ (for $i,j = 0,1,\ldots,127$). In terms of these conventions, Fig. 4(b) displays contours of the function $R_8(\cdot,\cdot)$, where

$$\begin{aligned} R_8(m,n) \equiv & \frac{1}{(128-m)(128-n)} \\ & \times \sum_{i=0}^{127-m} \sum_{j=0}^{127-n} E\left[\xi_0^8(i+m,j+n)\,\xi_0^8(i,j)\right]. \end{aligned} \tag{33}$$

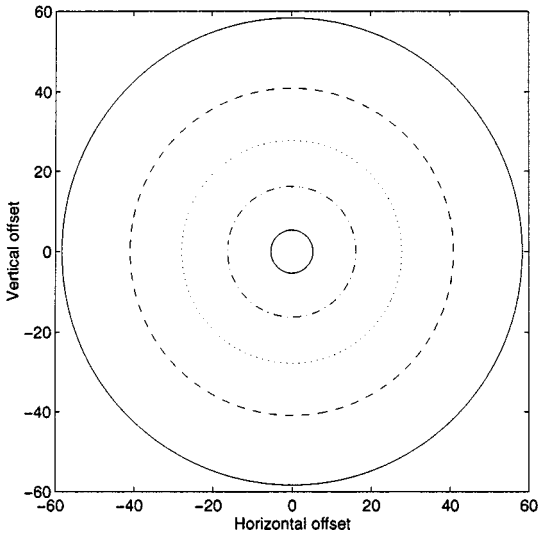


(a)

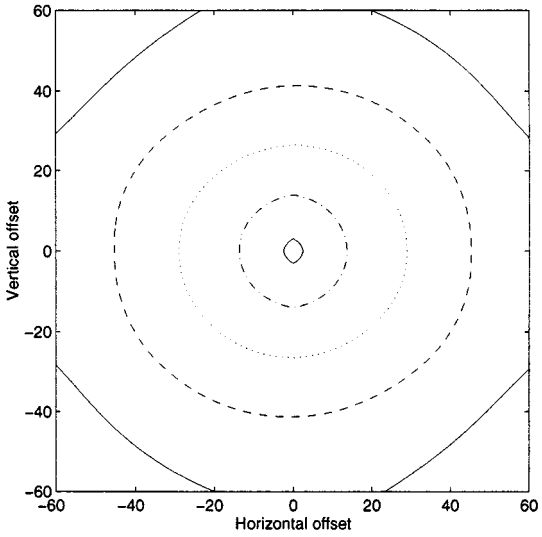


(b)

Fig. 3. Comparison of (a) vertical and (b) horizontal slices of the correlation contour plots in Fig. 4. Again, these plots are based on Monte Carlo simulation, where each point is within 0.005 of its correct value with 95% confidence.



(a)



(b)

Fig. 4. These figures display contour plots associated with $R_{yy}(\cdot,\cdot)$, defined in (32), with the contour levels at 0.9, 0.7, 0.5, 0.3, and 0.1. (a) The exact, desired correlation function. (The correlation function associated with a multiscale model of order 32 is visually indistinguishable from this exact function.) (b) The correlation function associated with a multiscale model of order 8. The latter has been determined by Monte Carlo simulation, using enough trials so that every estimated correlation value is within 0.005 of its correct value.

We remark that for model orders greater than just 16, our multiscale models capture virtually all of the significant correlation structure.

Let us consider the use of these multiscale models to carry out linear least-squares estimation. In Fig. 5(a), we display the original signal that we will be attempting to estimate. This signal consists of 128×128 pixels and has a Gaussian distribution. It is drawn from the *exact* distribution implied by (32) with $\ell = 80$. This field generation is effected by embedding the 128×128 grid into a larger 256×256 toroidal lattice, and extending the definition of $R_{yy}(\cdot,\cdot)$ to have periodic boundary conditions; for $\ell = 80$, this approach leads to a valid (i.e., positive definite) correlation function.

We consider two estimation problems related to the signal in Fig. 5(a). For the first, we corrupt the signal with spatially stationary white noise having covariance one, thus leading to an SNR of 0 dB [ince the signal also has a variance of one, as indicated by (32)]. In Fig. 5(b), we display an estimate based on our multiscale model of order 64. The sample MSE here is 0.0498. While there is no computationally feasible way to determine the mean-square error of an optimal estimator for this problem, we can obtain a fairly tight lower bound for the optimal MSE. In particular, let us consider the problem of estimating the value of the 256×256 signal, from which our 128×128 signal has been extracted. Since this 256×256 signal is stationary and is indexed on a toroidal lattice, exact calculations are possible. In particular, for estimating this signal in 0 dB white noise, the

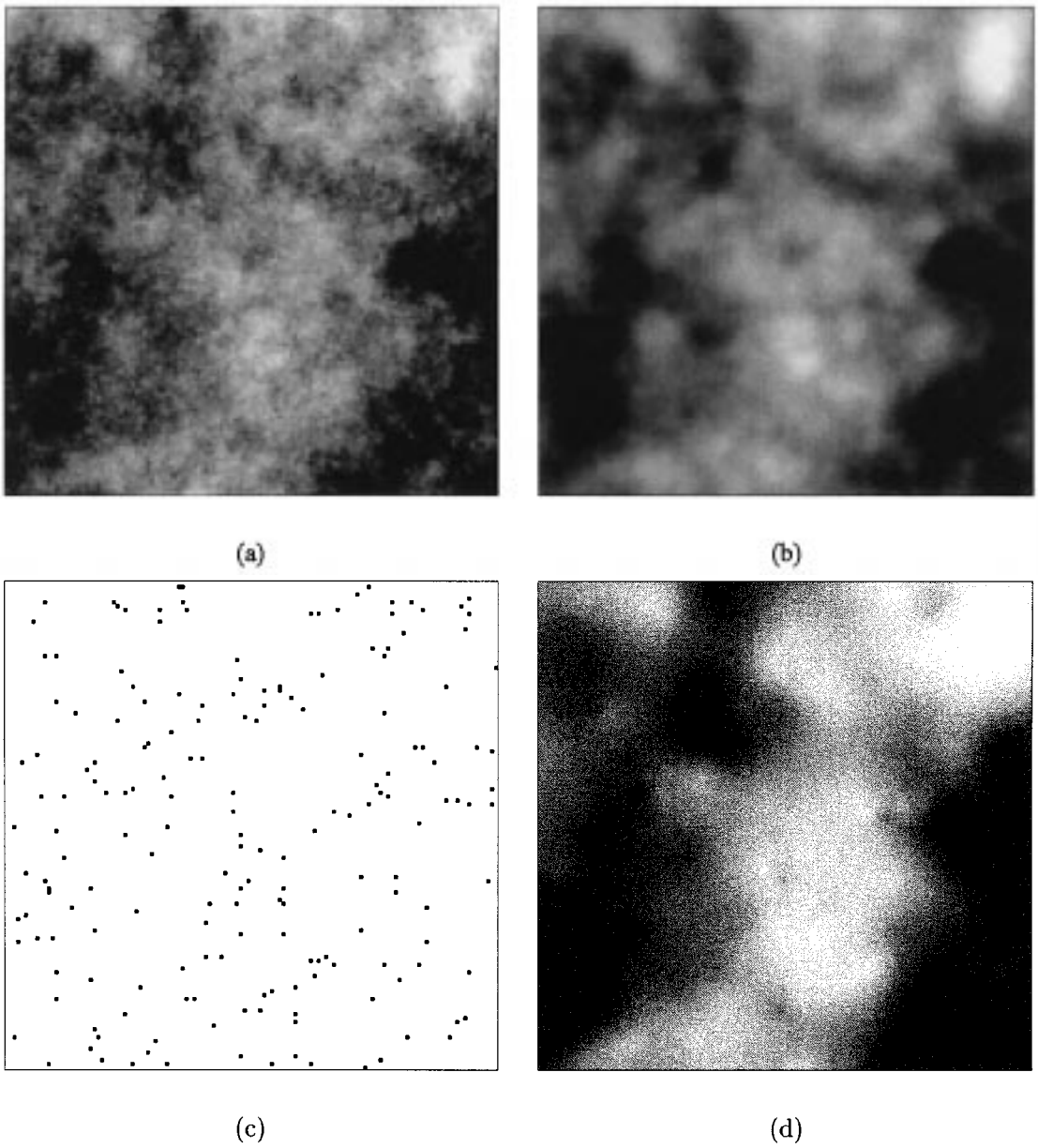


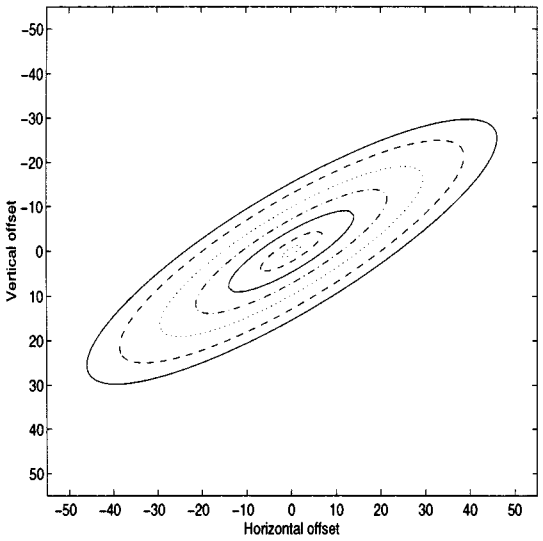
Fig. 5. These four figures relate to linear least-squares estimation of a signal having the isotropic correlation function in (32). (a) The original signal, with Gaussian deviates, drawn from the exact distribution using FFT-based techniques. (b) Estimate of the sample path in (a), based on noisy, densely distributed measurements of the signal, with 0 dB SNR; a 64-th order multiscale model is used to obtain this estimate. (c) Locations of observed pixels, for a second estimation experiment; these observed pixels provide only 1.11% coverage of the image. (d) Estimate of the sample path in (a), based on noiseless observations of the observed pixels [displayed in (c)].

optimal, FFT-based estimator has an MSE of 0.0458, which must lower-bound the MSE of an optimal estimator in our original estimation problem. By comparison, then, our measured MSE of 0.0498 is quite satisfactory. Although not shown in the figure, the same level of performance is also achieved by our lower-order multiscale models; specifically, our models of order 32, 16 and 8 achieve sample MSEs of 0.0501, 0.0533 and 0.0544, respectively, which are all close to the optimal.

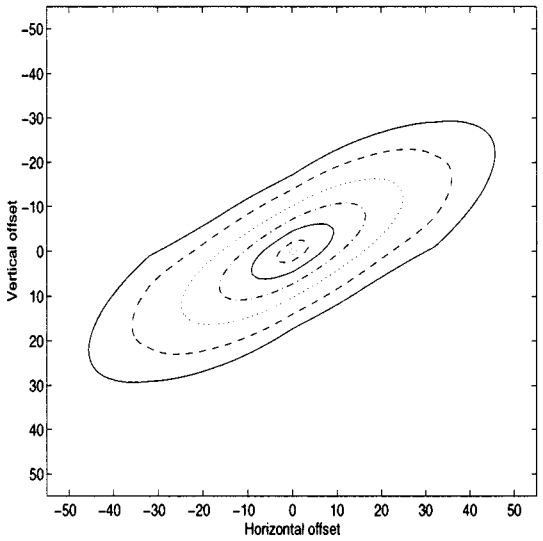
The second estimation problem we consider is one for which the FFT is of little practical use. In particular, we consider the problem of estimating the signal displayed in Fig. 5(a), based on noiseless measurements at the extremely sparse set of points displayed in Fig. 5(c). These points, chosen as the realization of a low-rate, homogeneous Poisson point process, provide only 1.11% coverage of the image region. Their irregular distribution is the key reason that FFT techniques are not useful. On the other hand, in Fig. 5(d), we display the estimate that results from use of our multiscale model of order 64. In light of the sparsity of our measurement coverage, this estimate has impressively captured the coarse qualitative features of the true signal; in fact, the sample MSE of this estimate is only 0.1147.

B. Reduced-Order Representations of Warped-Version of Isotropic Correlation Function

For our second example, we build multiscale representations for a stationary random field having a correlation function that is a warped version of the isotropic correlation function $R_{yy}(k,l)$

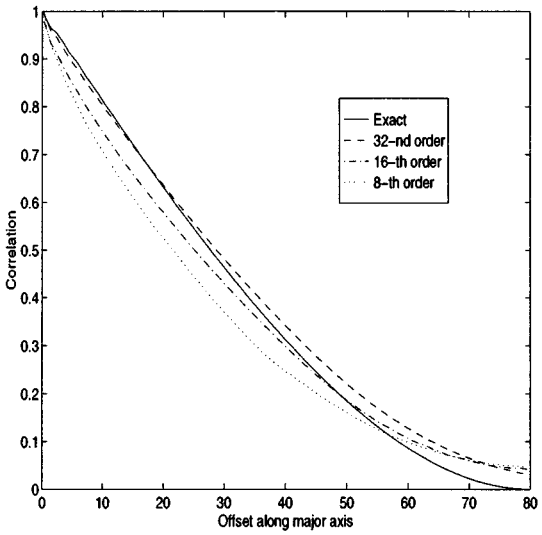


(a)

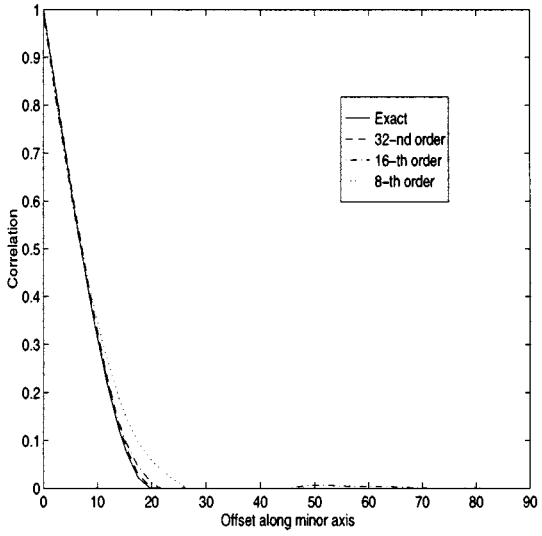


(b)

Fig. 6. These figures display contour plots associated with $R'_{yy}(\cdot,\cdot)$, defined in (34b), with the contour levels at 0.95, 0.85, 0.75, 0.6, 0.45, 0.3, and 0.15. (a) The exact, desired correlation function. (b) The correlation function associated with a multiscale model of order 8. The latter has been determined by Monte Carlo simulation, using enough trials so that every estimated correlation value is within 0.005 of its correct value.



(a)



(b)

Fig. 7. Comparison of slices of correlation contour plots in the previous figure. (a) A slice along the direction of the major axis of the ellipses in part (a) of the previous figure. (b) A slice along the direction of the minor axis of the ellipses in part (a) of the previous figure. Again, these plots are based on Monte Carlo simulation, where each point is within 0.005 of its correct value with 95 percent confidence.

in (32). Our warped version, which we denote by $R'_{yy}(k,l)$ is defined as follows:

$$R'_{yy}(i,j) = R_{yy}(i',j') \tag{34a}$$

$$\begin{pmatrix} i' \\ j' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} i \\ j \end{pmatrix}$$

$$\theta = \frac{\pi}{4} - \frac{\pi}{13}. \tag{34b}$$

The characteristic length ℓ of $R_{yy}(i,j)$ [see (32)] is again set to $\ell = 80$. A contour plot of $R'_{yy}(i,j)$ is displayed in Fig. 6(a), while slices of this correlation function along the directions of strongest and weakest correlation are displayed in Fig. 7(a) and (b), respectively.

We consider the problem of building a multiscale model, indexed on a quadtree, to realize the correlation function (34b) on a 128×128 grid. We constrain the multiscale model dimension to the respective values of 64, 32, 16, and 8.

In Fig. 6(b), we display as contour plots the correlation function associated with a multiscale model of order 8. Just as in our previous example, we must carefully define the precise meaning of these contours. As in the previous example, we let ξ_0^8 denote the random vector comprising the finest-scale of the particular multiscale process in which the state vectors are constrained to have dimension no greater than 8. We denote the (i,j)th component of ξ_0^8 by $\xi_0^8(i,j)$ (for $i,j = 0,1,\ldots,127$). In terms of these conventions, Fig. 6(b) displays contours of the function $R_8(\cdot,\cdot)$,

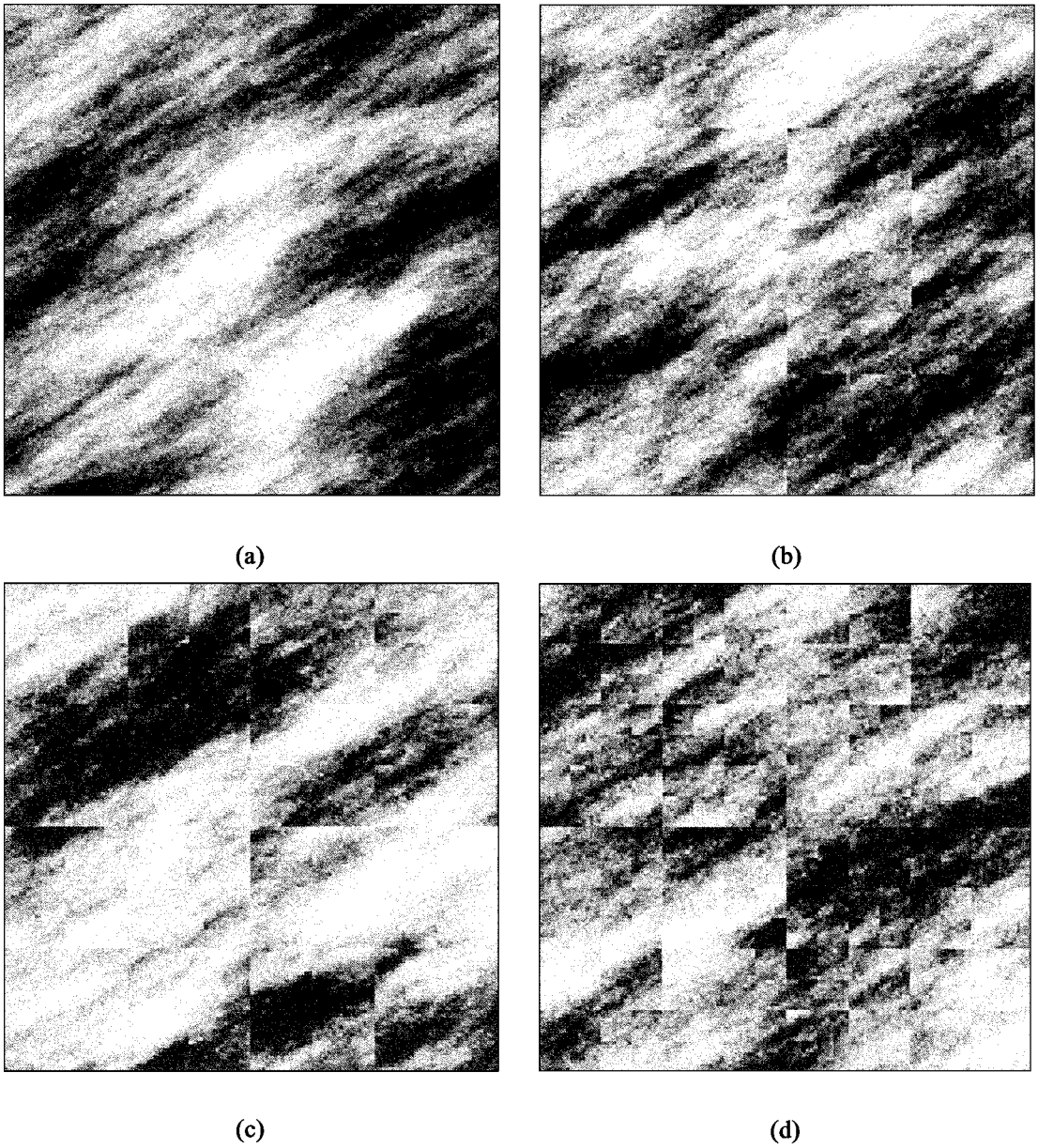


Fig. 8. These four figures display sample paths of a random field having the correlation function given in (34b), for a 128 × 128 pixel region. The sample paths in (a), (b), (c), and (d) correspond to multiscale models of order 64, 32, 16, and 8, respectively, using Gaussian deviates.

where $R_\lambda(\cdot,\cdot)$ is defined in (33). We remark that for model orders greater than or equal to 32, the contour plot becomes indistinguishable from the ideal, desired correlation in (a). To allow for more direct comparison of these contours, we overlay slices of them in Fig. 7(a) and (b) more specifically, Fig. 7(a) represents a slice of the contour plots, along the direction of strongest correlation, while Fig. 7(b) represents a slice of the contour plots along the direction of weakest correlation.

In Fig. 8, we display sample paths of this random field using Gaussian deviates, generated with our models of order 64, 32, 16, and 8. We see that unless a relatively high order model is used, the sample paths exhibit visually distracting blocky artifacts at the quadrantal boundaries. While in many applications, these artifacts are devoid of any statistical significance, they may be important in other contexts. One way to eliminate these artifacts is to employ a relatively high-order model multiscale model; for instance, as shown in Fig. 8(a), the 64-th order model is effective in this regard. An alternative, arguably more elegant approach to eliminating these artifacts is to use so-called *overlapping tree* models, in which distinct tree nodes correspond to overlapping portions of the image domain; this idea is described in detail in [14].

VI. Conclusion

This paper develops elements of a theory for multiscale stochastic realization, focusing on the problem of building multiscale models to realize, either exactly or approximately, prespecified finest-scale statistics. A key challenge has been to generalize the time-series notion of state vectors serving as an interface between the past and the future of a random process. The generalization is made by introducing a generalized correlation

coefficient, which is used to make precise the notion of multiscale state vectors serving to decorrelate multiple subsets of a multiscale process. Once the reduced-order multiscale modeling problem has been formalized, we harness canonical correlation analysis to develop a suboptimal model-building algorithm. We demonstrate the practicality of this algorithm in problems of random-field estimation and generation. In the context of field generation, we demonstrate an ability to build multiscale processes having a finest-scale correlation matching very closely desired correlations. In the context of field estimation, we build multiscale models that are in turn used to efficiently carry out least-squares estimation, with the resulting field estimates having nearly optimal mean-square error.

APPENDIX A
PROOF OF PROPOSITION 2

In this appendix, we prove Proposition 2. The following lemma is instrumental to the proof.

Lemma 2: Let s and t be nodes at a common, coarse scale of a tree, with internal matrices W_s and W_t satisfying (13). Then, $\forall\,(i,j) \in \{1,\ldots,q\}^2$

$$P_{\chi_{s\alpha_i}\chi_{s\alpha_j}} = P_{\chi_{s\alpha_i}\chi_s} W_s^T \left(W_s P_{\chi_s} W_s^T\right)^{-1} \times W_s P_{\chi_s\chi_t} W_t^T \left(W_t P_{\chi_t} W_t^T\right)^{-1} W_t P_{\chi_t\chi_{t\alpha_j}}. \tag{35}$$

Proof: Since W_s and W_t satisfy (13), it follows that:

$$E\left[\left(\chi_{s\alpha_i} - \hat{E}\left(\chi_{s\alpha_i}|W_s\chi_s\right)\right) \times \left(\chi_{t\alpha_j} - \hat{E}\left(\chi_{t\alpha_j}|W_t\chi_t\right)\right)^T\right] = \mathbf{0}$$

and hence

$$P_{\chi_{s\alpha_i}\chi_{t\alpha_j}} = E\left[\hat{E}\left(\chi_{s\alpha_i}|W_s\chi_s\right)\chi_{t\alpha_j}^T\right]. \tag{36}$$

Now, combining (36) with the fact that

$$\hat{E}\left(\chi_{s\alpha_i}|W_s\chi_s\right) = P_{\chi_{s\alpha_i}\chi_s} W_s^T \left(W_s P_{\chi_s} W_s^T\right)^{-1} W_s\chi_s,$$

the result (35) immediately follows. **QED**

Returning to the proof of Proposition 2, we proceed inductively by scale. At the coarsest scale, (15) is automatically satisfied, thanks to (9). Now, suppose that (15) holds for states at scale m, and consider the states $x(s)$ and $x(t)$ at scale $m+1$. There are two possibilities: either $s = t$ or $s \neq t$. If $s = t$, then

$$\begin{aligned} P_{x(s)x(t)} &= A(s)P_{x(s\bar{\gamma})}A^T(s) + Q(s) \\ &= A(s)W_{s\bar{\gamma}}P_{\chi_{s\bar{\gamma}}}W_{s\bar{\gamma}}^T A^T(s) + Q(s) \\ &= W_s P_{\chi_s} W_s^T. \end{aligned} \tag{37}$$

Here, the first equality follows from the structure (1) of the multiscale dynamics, the second equality follows from the induction hypothesis, and the third equality follows by substitution of the expressions (14a) and (14b) for $A(s)$ and $Q(s)$. Turning to the other possibility, if $s \neq t$, then

$$\begin{aligned} P_{x(s)x(t)} &= A(s)P_{x(s\bar{\gamma})x(t\bar{\gamma})}A^T(t) \\ &= A(s)W_{s\bar{\gamma}}P_{\chi_{s\bar{\gamma}}\chi_{t\bar{\gamma}}}W_{t\bar{\gamma}}^T A^T(t) \\ &= W_s P_{\chi_s\chi_t} W_t^T. \end{aligned} \tag{38}$$

Here, the first equality follows from the structure (1) of the multiscale dynamics, the second equality follows from the induction hypothesis, and the third equality follows by substitution of the expression (14a) for $A(s)$ together with the foregoing Lemma. The proof of the proposition is now complete. **QED**

APPENDIX B
PROOF OF PROPOSITION 3

In this appendix, we complete the proof of Proposition 3, by establishing the validity of (20a) and (20b). For this purpose, we continue to use the notation established in Section IV-A.

We begin by making explicit the connection between the value of $\bar{\rho}(\mu_1,\mu_2|W_1\mu_1)$ and the cross-covariance D between μ_1 and μ_2. This connection is useful for proving both Propositions 3 and 4.

Lemma 3: Let $W_1 \in \mathcal{N}_\lambda$, $0 \leq \lambda < m_1$, and let $W_1^\perp$ be a matrix whose rows form an orthonormal basis for the nullspace of W_1. Then,

$$\bar{\rho}(\mu_1,\mu_2|W_1\mu_1) = \max_{g_1\in G_1, g_2\in G_2} \left\{g_1^T W_1^\perp D g_2\right\} \tag{39a}$$

$$= \max_{g_2\in G_2} \|W_1^\perp D g_2\|_2 \tag{39b}$$

where G_1 and G_2 denote the following sets:

$$G_1 = \left\{g \in \mathcal{R}^{m_1-\lambda}; g^T g = 1\right\}$$
$$G_2 = \left\{g \in \mathcal{R}^{m_2} g^T\left(I_{m_2} - D^T W_1^T W_1 D\right) g = 1\right\}.$$

Proof: It is sufficient to establish (39a); then, (39b) follows by simple optimization theory. Proceeding, we let $W_1 \in \mathcal{N}_\lambda$, $0 \leq \lambda < m_1$. It is easy to see that

$$\bar{\rho}(\mu_1,\mu_2|W_1\mu_1) = \bar{\rho}(\mu_1',\mu_2|W_1'\mu_1') \tag{40}$$

where μ_1' and W_1' are related respectively to μ_1 and W_1 via unitary (but otherwise arbitrary) matrix U

$$\mu_1' \equiv U^T\mu_1, \quad W_1' \equiv W_1 U \tag{41}$$

To exploit this fact, we first let $W_1^\perp$ be a matrix whose rows form an orthonormal basis for the nullspace of W_1, and set $U = \left(W_1^T\left(W_1^\perp\right)^T\right)$, so that $W_1' = (I_\lambda \quad \mathbf{0})$. Also, we define

$$\tilde{\mu}_1' \equiv \mu_1' - \hat{E}\left(\mu_1'|W_1\mu_1\right), \quad \tilde{\mu}_2 \equiv \mu_2 - \hat{E}\left(\mu_2|W_1\mu_1\right)$$

with

$$\begin{pmatrix} P_{\tilde{\mu}_1'} & P_{\tilde{\mu}_1'\tilde{\mu}_2} \\ P_{\tilde{\mu}'\mu_2}^T & P_{\tilde{\mu}_2} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I_{m_1-\lambda} \end{pmatrix} & \begin{pmatrix} 0 \\ W_1^\perp D \end{pmatrix} \\ \begin{pmatrix} 0 \\ W_1^\perp D \end{pmatrix}^T & I_{m_2} - D^T W_1^T W_1 D \end{pmatrix}. \tag{42}$$

Finally, we note that by elementary theory of least-squares estimation

$$\bar{\rho}(\mu_1',\mu_2|W_1'\mu_1') = \max_{f_1\in F_1, f_2\in F_2} f_1^T P_{\tilde{\mu}_1'\tilde{\mu}_2} f_2$$

where

$$F_1 \equiv \left\{f; f^T P_{\tilde{\mu}_1'} f = 1\right\}, \quad F_2 \equiv \left\{f; f^T P_{\tilde{\mu}_2} f = 1\right\}. \quad \textbf{QED}$$

As a direct consequence of this lemma

$$\bar{\rho}\left(\mu_1,\mu_2|\begin{pmatrix} I_\lambda & \mathbf{0}\end{pmatrix}\mu_1\right)=\begin{cases} d_{\lambda+1}, & 0\le\lambda<m_{12}\\ 0 & \lambda\ge m_{12}\end{cases} \tag{43}$$

This fact establishes (20b).

What remains is to establish (20a). The following lemma provides the needed argument.

Lemma 4: For $\lambda < m_{12}$,

$$\min_{W_1\in\mathcal{N}_\lambda}\bar{\rho}(\mu_1,\mu_2|W_1\mu_1)=\bar{\rho}\left(\mu_1,\mu_2|\begin{pmatrix} I_\lambda & \mathbf{0}\end{pmatrix}\mu_1\right)=d_{\lambda+1}.$$

Proof: Thanks to (43), it is sufficient to show that for all $W_1\in\mathcal{N}_\lambda$

$$\bar{\rho}(\mu_1,\mu_2|W_1\mu_1)\ge d_{\lambda+1}. \tag{44}$$

To establish (44), we fix $W_1\in\mathcal{N}_\lambda$. Referring back to Lemma 3, and in particular (39a), we devise particular values for $g_1\in G_1$ and $g_2\in G_2$ for which

$$g_1^T\left(W_1^\perp D\right)g_2\ge d_{\lambda+1} \tag{45}$$

thus implying that $\bar{\rho}(\mu_1,\mu_2|W_1\mu_1)\ge d_{\lambda+1}$.

To establish (45), we first note that at least one of the unit vectors $e_1^T,e_2^T,\ldots,e_{\lambda+1}^T$ must belong to the row space of $W_1^\perp$, which itself has a dimension of $m_1-\lambda$; let us suppose that e_j^T belongs, with $h^TW_1^\perp=e_j^T$ for some $h\in\mathcal{R}^{m_1-\lambda}$. Now, exploiting the orthonormality of the rows of $W_1^\perp$, we see that $h\in G_1$, and hence we let $g_1=h$. Also, we let $g_2=e_j$, where the fact that $De_j=d_je_j=d_j\left(W_1^\perp\right)^Tg_1$, implies that $W_1De_j=\mathbf{0}$, so that indeed $e_j\in G_2$. But for these values for g_1 and g_2, $g_1^T(W_1^\perp D)g_2=d_j\ge d_{\lambda+1}$, thus establishing (45) and completing the proof. **QED**

The proof is now complete. **QED**

Appendix C
Proof of Proposition 4

In this appendix, we complete the proof of Proposition 4, by establishing the validity of (27). For this purpose, we continue to use the notation established in Section IV-A.

We begin by fixing W_1, which we assume without loss of generality to have orthonormal rows. We know from (43) that $\bar{\rho}(\mu_1,\mu_2)=d_1$. Combining this fact with (39b), it follows that the Proposition will be proved if we can show that:

$$\max_{g_2\in G_2}\left\|W_1^\perp Dg_2\right\|_2^2\le d_1^2. \tag{46}$$

To establish (46), we first note that since the rows of $W_1^\perp$ form an orthonormal basis for the null space of W_1, we have that $\forall\, x$

$$\|x\|_2^2=\|W_1x\|_2^2+\left\|W_1^\perp x\right\|_2^2. \tag{47}$$

Since $\forall\, g_2\in G_2$,

$$g_2^T\left(I-D^TW_1^TW_1D\right)g_2=\|g_2\|_2^2-\|W_1Dg_2\|_2^2=1 \tag{48}$$

we can apply (47) in (48) with $x=Dg_2$ to see that $\forall\, g_2\in G_2$

$$\left\|W_1^\perp Dg_2\right\|_2^2=\|Dg_2\|_2^2-\|g_2\|_2^2+1,\quad\forall\, g_2\in G_2. \tag{49}$$

However

$$\begin{aligned}&\min_{g_2\in G_2}\left\{\|g_2\|_2^2-\|Dg_2\|_2^2\right\}\\ &\quad=\min_{g_2\ne\mathbf{0}}\left\{\frac{g_2^T\left(I-D^TD\right)g_2}{g_2^T\left(I-D^TW_1^TW_1D\right)g_2}\right\}\\ &\quad\ge\min_{g_2\ne\mathbf{0}}\left\{\frac{g_2^T\left(I-D^TD\right)g_2}{g_2^Tg_2}\right\}\\ &\qquad\times\min_{g_2\ne\mathbf{0}}\left\{\frac{g_2^Tg_2}{g_2^T(I-D^TW_1^TW_1D)g_2}\right\}\\ &\quad=\mathrm{eig}_{\min}\left(I-D^TD\right)\mathrm{eig}_{\min}\left[\left(I-D^TW_1^TW_1D\right)^{-1}\right]\\ &\quad=\left(1-d_1^2\right)(1)\end{aligned} \tag{50}$$

where $\mathrm{eig}_{\min}(\cdot)$ denotes the smallest eigenvalue of the enclosed matrix expression. In the third line, we have used Rayleigh's principle [22], which asserts that for any pair of symmetric, positive definite matrices A and B

$$\min_{x\ne\mathbf{0}}\frac{x^TBx}{x^TAx}=\mathrm{eig}_{\min}\left(A^{-1}B\right).$$

By combining (49) and (50), the desired result (46) is established. **QED**

Acknowledgment

The authors are grateful to M. Daniel for his detailed comments which have greatly enhanced the manuscript's clarity.

References

[1] H. Akaike, "Markovian representation of stochastic processes by canonical variables," *SIAM J. Control*, vol. 13, no. 1, pp. 162–173, Jan. 1975.
[2] ——, "Stochastic theory of minimal realization," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 667–674, Dec. 1974.
[3] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1958.
[4] M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 766–784, Mar. 1992.
[5] K. C. Chou, A. S. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion and regularization," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 464–478, Mar. 1994.
[6] M. Daniel and A. S. Willsky, "A multiresolution methodology for signal-level fusion and data assimilation with applications in remote sensing," *Proc. IEEE*, vol. 85, pp. 164–180, Jan. 1997.
[7] M. Daniel, "Multiresolution Statistical Modeling With Application to Modeling Groundwater Flow," Ph.D. dissertation, Laboratory for Information and Decision Systems, Massachusetts Insti. Technol., Cambridge, MA, 1997.
[8] H. Derin and P. A. Kelly, "Discrete-index Markov-type random processes," *Proc. IEEE*, vol. 77, pp. 1485–1510, Oct. 1989.
[9] U. B. Desai and D. Pal, "A realization approach to stochastic model reduction and balanced stochastic realizations," in *Proc. 21st Conf. Decision Control*, 1982, pp. 1105–1114.
[10] P. Fieguth and A. Willsky, "Fractal estimation using models on multiscale trees," *IEEE Trans. Signal Processing*, vol. 44, pp. 1297–1300, May 1996.
[11] P. Fieguth, A. Willsky, W. Karl, and C. Wunsch, "Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry," *IEEE Trans. Geosci. Rem. Sensing*, vol. 33, pp. 280–292, Mar. 1995.
[12] T. Ho, P. Fieguth, and A. Willsky, "Multiresolution stochastic models for the efficient solution of large-scale space-time estimation problems," IEEE Trans. Automat. Contr., submitted for publication.
[13] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321–377, 1936.

[14] W. Irving, "Multiresolution stochastic realization and model identification with applications to large-scale estimation problems," Ph.D. dissertation, Laboratory for Information and Decision Systems, Massachusetts Inst. Technol., Cambridge, MA, 1995.
[15] W. Irving, P. Fieguth, and A. Willsky, "An overlapping tree approach to multiscale stochastic modeling and estimation," *IEEE Trans. Image Processing*, vol. 6, pp. 1517–1529, Nov. 1997.
[16] A. Lindquist and G. Picci, "On the stochastic realization problem," *SIAM J. Control Optim.*, vol. 17, no. 3, pp. 365–389, May 1979.
[17] M. Luettgen and A. S. Willsky, "Likelihood calculations for a class of multiscale stochastic models, with applications to texture discrimination," *IEEE Trans. Image Processing*, vol. 4, pp. 194–207, Feb. 1995.
[18] M. Luettgen, W. C. Karl, A. S. Willsky, and R. R. Tenney, "Multiscale representations of Markov random fields," *IEEE Trans. Signal Processing*, vol. 41, pp. 3377–3396, Dec. 1993.
[19] M. Luettgen, W. Karl, and A. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. Image Processing*, vol. 3, pp. 41–64, Jan. 1994.
[20] G. Picci, "Stochastic realization of Gaussian processes," *Proc. IEEE*, vol. 64, no. 1, pp. 112–122, Jan. 1976.
[21] M. A. Sironvalle, "The random coin method: Solution of the problem of simulation of a random function in the plane," *Math. Geology*, vol. 12, no. 1, pp. 29–36, Jan. 1980.
[22] G. Strang, *Linear Algebra and its Applications*. San Diego, CA: Harcourt Brace Jovanovich, 1988.

William W. Irving received the S.B., S.M., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology (M.I.T.), Cambridge, MA, in 1987, 1991, and 1995, respectively.

From 1987 to 1995, he was a member of the technical staff at M.I.T. Lincoln Laboratory, where he participated in the Staff Associate program for graduate study. From 1995 to 1999, he was a member of the technical staff at Alphatech, Inc., Burlington, MA. where he developed algorithms for automatic target recognition with synthetic aperture radar. Since 1999, he has ben a quantitative analyst in the Fixed-Income Division of Fidelity Investments, Merrimack, NH.

Alan S. Willsky (S'70–M'73–SM'82–F'86) received the S.B. and Ph.D degrees in aeronautics and astronautics from the Massachusetts Institute of Technology (M.I.T.), Cambridge, MA, in 1969 and 1973, respectively.

He joined the M.I.T. faculty in 1973, and is currently the Edwin S. Webster Professor of Electrical Engineering. He has held visiting positions in England and France, as well as various leadership positions in the IEEE Control Systems Society (which made him a Distinguished Member in 1988). He has delivered numerous keynote addresses, and is coauthor of the undergraduate text *Signals and Systems* (Upper Saddle River, NJ: Prentice-Hall, 1st ed. 1983, 2nd ed. 1997). His research interests are in the development and application of advanced methods of estimation and statistical signal and image processing. Methods he has developed have been successfully applied in a variety of applications, including failure detection, surveillance systems, biomedical signal and image processing, and remote sensing.

Dr. Willsky is a founder and member of the Board of Directors of Alphatech, Inc., Burlington, MA, and a member of the U.S. Air Force Scientific Advisory Board. He has received several awards, including the 1975 American Automatic Control Council Donal P. Eckman Award, the 1979 ASCE Alfred Noble Prize, and the 1980 IEEE Browder J. Thompson Memorial Award.