Multiscale Autoregressive Models and Wavelets

Khalid Daoudi, Austin B. Frakt, Alan S. Willsky

Multiscale Autoregressive Models and Wavelets

Khalid Daoudi, Austin B. Frakt, *Student Member, IEEE*, and Alan S. Willsky, *Fellow, IEEE*

***Abstract*— The multiscale autoregressive (MAR) framework was introduced to support the development of optimal multiscale statistical signal processing. Its power resides in the fast and flexible algorithms to which it leads. While the MAR framework was originally motivated by wavelets, the link between these two worlds has been previously established only in the simple case of the Haar wavelet. The first contribution of this paper is to provide a unification of the MAR framework and all compactly supported wavelets as well as a new view of the multiscale stochastic realization problem. The second contribution of this paper is to develop wavelet-based approximate internal MAR models for stochastic processes. This will be done by incorporating a powerful synthesis algorithm for the detail coefficients which complements the usual wavelet reconstruction algorithm for the scaling coefficients. Taking advantage of the statistical machinery provided by the MAR framework, we will illustrate the application of our models to sample-path generation and estimation from noisy, irregular, and sparse measurements.**

***Index Terms*—Fractional Brownian motion, graphical models, internal models, multiscale estimation, multiscale models, stochastic realization, wavelets.**

I. Introduction

In THE last two decades, multiresolution decomposition techniques, such as the wavelet transform, have been widely and successfully applied in signal processing. This is due both to their ability to compactly capture the salient scale-to-scale properties that many signals exhibit and to the efficiency of the algorithms to which they lead. With both of these attractive features in mind, the multiscale autoregressive (MAR) framework was introduced [5] to support the development of optimal multiscale statistical signal processing. Like the wavelet transform, MAR processes are recursive in scale and, due to the nature of the scale recursion, fast statistical signal processing algorithms (sample-path generation, linear least squares estimation, likelihood calculation) for MAR processes exist. The power of the MAR framework resides in its ability to simultaneously address several complications which arise in a variety of signal processing applications. For instance, the data sets can be large, the processes to be estimated can be nonstationary, the measurements may be irregularly spaced, nonlocal, and corrupted by nonstationary noise.

Despite the apparent similarities between wavelets and the MAR framework, it seemed that the two could not be easily reconciled except in the simplest case of the Haar wavelet. The crux of the problem, which this paper resolves, hinges on the particular properties that useful MAR processes possess.

One such property, which is very similar to wavelet reconstruction, is that the finest-scale signal of an MAR process is formed from a coarse signal by successively adding detail. Specifically, the MAR framework provides an *implicit* second-order statistical model for a fine-scale signal by creating a dynamically coupled hierarchy of vector-valued signals above the fine-scale signal (see Fig. 1).

The coarse-to-fine dynamics of an MAR process[1] $x(\cdot) \triangleq \{x_j(n) \mid j = 0,1,\cdots,J,\ n = 0,1,\cdots,2^j - 1\}$, for some integer J, are analogous to those of a state-space model

$$x_j(n) = A_j(n)x_{j-1}([n/2]) + w_j(n) \tag{1}$$

where $[\cdot]$ indicates the integer part and $w_j(n)$ is white noise with auto-covariance $Q_j(n)$ (see Fig. 1 for a concise summary of our indexing notation). Note, however, that (1) is more general than a classical state-space model because here, $x(\cdot)$ is indexed by the nodes of *any* tree rather than by the integers which form a monadic tree. We call the vector $x_j(n)$ an MAR state.

The form of (1) suggests that MAR dynamics can be made to mimic those of the wavelet reconstruction algorithm. This has been previously shown only in the case of the Haar wavelet [21]. We will show that through a particular definition of the state vector $x_j(n)$, the MAR dynamics can be chosen to match the reconstruction algorithm associated with *any* compactly supported orthogonal or biorthogonal wavelet. We emphasize, however, that signal synthesis is not our purpose. Instead, *modeling* is our objective. Specifically, given the statistics of a random signal, which we view as indexed by the leaf nodes of a tree, we focus on building an MAR model to capture those given statistics with high fidelity.

A particular class of MAR processes we are going to focus on are *internal* MAR processes. While we will elaborate on the notion of internality later, briefly an internal MAR process is one for which the state at every node[2] (j,n) is a linear functional of the states which reside at the fine-scale nodes which descend from (j,n). In general, there are no constraints on how the MAR states should be defined, but internality is a property which is very useful for many reasons. First, internal models are intellectually important in the context of statistical modeling and are widely used in stochastic realization theory for time-series [30]. It is, therefore, natural to consider the extension of this idea to the context of MAR processes as is done in [24]–[28].

Manuscript received March 1, 1998. This work was supported by the Office of Naval Research under Grant N00014-91-J-1004, the Air Force Office of Scientific Research under Grant F49620-98-1-0349, and by INRIA Rocquencourt, France.

The authors are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: daoudi@lids.mit.edu).

Publisher Item Identifier S 0018-9448(99)02268-3.

[1] In this paper, we will assume without loss of generality that all processes we consider are zero-mean.

[2] We will use the notation (j,n) to refer to the node of a dyadic tree at scale j and shift n.

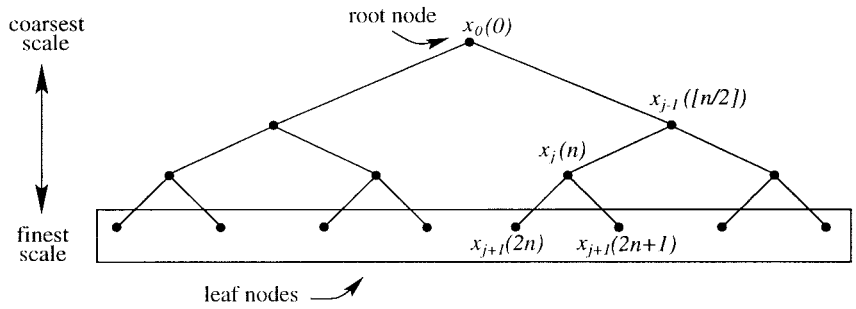


Fig. 1. An MAR process on a dyadic tree. The root node state is $x_0(0)$. The parent of state $x_j(n)$ is $x_{j-1}([n/2])$, while its children are $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$.

Second, in many applications there is a need to fuse measurements taken at different scales [15]. Frequently, the nonlocal coarse-scale measurements in a multiscale measurement set are linear functionals of the finest scale variables (e.g., tomographic measurements). Since internal MAR models include as coarse states nonlocal functions of the finest scale, they allow efficient fusion of nonlocal and local measurements with no increase in complexity as compared to the case of fusing only fine-scale data.

Finally, as we will see, internality provides a convenient parameterization of the information content of the MAR states. Using this parameterization leads to MAR dynamics that incorporate a powerful optimal prediction of a child state from its parent. This optimal prediction will have important and significant consequences for our ability to accurately model signals using MAR processes based on wavelets.

In the early attempts to marry MAR processes and wavelets, it was incorrectly thought that the internal property doomed the union in all but the Haar case. This is because for all but the Haar wavelet, the supports of the wavelet functions overlap. We will show the connection between the overlapping of the wavelet functions and the internal property and illustrate how the nonoverlapping property of the Haar wavelet permits its simple union with the MAR framework. After proving some particular relationships between wavelet coefficients and through appropriate state augmentation we show how to build internal MAR processes based on any compactly supported wavelet.

An important property that MAR processes possess is wide-sense Markovianity. This MAR "Markov Property," as we shall call it, is a generalization of the wide-sense Markovianity of state-space processes. For a state-space process, the present is a boundary between the past and future in that it conditionally decorrelates them. Analogously, for an MAR process, the node (j,n) is a boundary between the subtrees extending away from it; the values of the MAR process indexed by these subtrees are mutually conditionally decorrelated by $x_j(n)$. The Markov Property means that $x_j(n)$ summarizes all the relevant stochastic properties of one subtree leading from (j,n) for optimal statistical reasoning about the other subtrees leading from (j,n) and therefore justifies our terminology in calling $x_j(n)$ an MAR state. For internal MAR processes, the Markov Property is equivalent to the fact that the linear functionals which define the states fulfill the conditional decorrelation role just described. This provides part of our motivation for choosing the linear functionals from a wavelet basis since they have nice decorrelation properties.

Our main objective, after showing how to unify the MAR and wavelet frameworks, is to build approximate internal MAR models for stochastic processes. To do so, we use the statistics of the process to be modeled to derive the dynamics of our internal MAR-wavelet models. While wavelets have nice decorrelation properties, the decorrelation they provide is not exact in general. Therefore, our MAR models based on wavelets are approximate. This does *not* mean that we assume in our internal models that the detail coefficients are white. In fact, while (for comparison purposes) we do make this assumption for what we shall call the standard MAR-wavelet models, our internal MAR-wavelet models are more sophisticated. In particular, they incorporate the powerful property of optimal stochastic prediction for the detail coefficients at a given scale from both detail *and* scaling coefficients at coarser scales. This is different from the common wavelet-based modeling in which the detail coefficients are assumed to be white. In our internal models, we make the weaker assumption that the prediction *errors* are white. This assumption is the reason why our models are approximate.

Another property that an MAR process must posses is low state dimensionality. This is because sample-path generation, linear least squares estimation [7], [8], and likelihood calculation [31] for MAR processes all have complexity which is polynomial in state dimension (and linear in the number of fine-scale nodes).

We will see that the state dimension of our MAR-wavelet models grows only linearly with the lengths of the support of the scaling functions, which are related, in some cases, such as orthogonal wavelets, to the number of vanishing moments of the analyzing wavelet. However, the fact that wavelets with a large number of vanishing moments do a good job of whitening and stationarizing a large class of processes [1], [22], [29], [35], [37], [39], does *not* imply that the degree of statistical fidelity of our internal models necessarily increases with the number of vanishing moments. This is because we are not exclusively concerned with the correlations between wavelet coefficients, but rather with the *conditional* correlation between them (in order to approximately satisfy the Markov Property). Therefore, as will be illustrated, with internal MAR-wavelet models, it is possible to build accurate models using wavelets with fairly short supports and, thus, without dramatically increasing the state dimension.

Throughout this paper, we will use fractional Brownian motion (fBm) as a vehicle to support our results. In particular, we will apply our internal MAR-wavelet models to generating fBm sample-paths and, more importantly, to the problem of estimating an fBm from noisy, irregular, and sparse measurements.

In the context of the previous work on MAR processes, this paper provides a unification of wavelets and MAR processes as well as a new view of the *multiscale stochastic realization problem*. This problem is to design coarse-scale states to match the given fine-scale statistics. Previous approaches to the multiscale stochastic realization problem focussed on designing MAR states to, in some sense, optimally match the statistics of the finest scale process being modeled [23]–[28]. As a consequence, the resulting states typically have no discernible structure beyond the fact that they represent solutions to specific optimization problems.

Our approach differs in that the design of our MAR states is not closely tied to the intricate details of the fine-scale statistics. Our philosophy which, in part, motivated this work is to restrict the class of linear functionals that define the states of an internal MAR process to the small but rich class of wavelet bases. We thus force the states to contain meaningful multiscale representations of the fine-scale process and avoid the computationally burdensome search over all possible linear functionals.

On the other hand, our approach is similar to the previous work on the multiscale stochastic realization problem in that we use the fine-scale statistics of the process to be modeled to derive the dynamics of the MAR model.

There are at least two other frameworks [9], [20] that propose unifications of multiscale stochastic processes and wavelets, both of which are different from ours in significant ways. To fully contrast [9] and [20] with our work requires a longer discussion than space constraints permit. Therefore, we focus on the main points and refer the reader to [9] and [20] for details. The first [9] develops decoupled dynamic wavelet systems that support a fast estimation algorithm whose structure takes the form of a set of monadic trees in scale. The second [20] generalizes the work in [9] to wavelet packets and develops a fast estimation framework whose structure takes the form of a set of dyadic trees. There are several differences between these approaches and ours. The first is that in [9] and [20] it is assumed that the detail coefficients are white.[3] In our internal models, we make no such assumption. The second is that to perform estimation in the frameworks of [9] and [20], the data must be transformed into the wavelet domain. Therefore, there is no way to handle sparse or irregular measurements which our framework can handle easily. Lastly, the models of [9], [20] are only interesting in the noninternal context. Under an internality assumption, they collapse to a trivial case.

There exist other wavelet-based tree models such as those in [6], [12], and [13]. In these models, the key wavelet coefficient dependencies are captured using hidden Markov models (HMM's). In particular, these HMM's are used to capture contexts describing statistical features. These models differ from ours in that they include hidden variables and are, therefore, necessarily not internal. Nevertheless, these models also lead to very powerful and efficient algorithms for signal and image processing.

The remainder of this paper is organized as follows. In Section II we review the MAR framework. In Section III we review the essential elements of wavelet theory and introduce the MAR–Haar process. In Section IV we show how to build an internal MAR-wavelet process based on any compactly supported wavelet. In Section V we build approximate MAR-wavelet models for random signals. The conclusion and a discussion of open questions for future work are in Section VI.

II. MAR Processes Background

MAR processes are tree-indexed stochastic processes. For the purposes of this paper it is sufficient to consider only dyadic trees. Our notation for referring to nodes of a dyadic tree is indicated in Fig. 1. There is a natural notion of scale associated with dyadic trees. The root node represents the coarsest scale which we denote as scale zero. The children of the root node represent the first scale. Continuing, leaf nodes constitute the finest scale which we denote as the Jth scale. We will denote by x_J the stacked vector consisting of $x_J(n)$ for $n = 0, 1, \cdots, 2^J - 1$, i.e., the finest scale subprocess.

As mentioned, an MAR process has a Markov Property: it is a (wide-sense) Markov random field on a tree [7]. This Markov Property leads to fast statistical signal processing algorithms. For instance, sample-path generation (with complexity quadratic in state dimension and linear in the number of finest-scale nodes) is accomplished using (1). Also, a linear least squares estimator [7], [8] and likelihood calculator [31] have been developed based on a measurement model

$$y_j(n) = C_j(n)x_j(n) + v_j(n). \tag{2}$$

The zero-mean white noise, $v_j(n)$ has auto-covariance $R_j(n)$ and is uncorrelated with the MAR process[4] $x(\cdot)$ and the process noise $w(\cdot)$. The estimator and likelihood calculator have computational complexity cubic in state dimension and linear in the number of finest-scale nodes.

An important subclass of the class of MAR processes are *internal* MAR processes. As stated in Section I, an internal MAR process is one for which each state $x_j(n)$ is a linear function of the portion of x_J indexed by finest scale nodes which descend from (j, n). Therefore, if $x(\cdot)$ is an internal MAR process we can write

$$x_j(n) = W_j(n)x_J. \tag{3}$$

We will call $W_j(n)$ an *internal matrix*. As shown in [23] and [24], a necessary and sufficient condition to fulfill (3) is for each state $x_j(n)$ to be a linear function of $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$, the states which reside at the children nodes of (j, n). Notice that this fact imposes some constraints on

[3] Moreover, the approaches in [9] and [20] assume a particular model while we construct one.

[4] We will use the notation $x(\cdot)$ and $w(\cdot)$ to refer to $x_j(n)$ and $w_j(n)$, respectively, for all j and n.

the form of $W_j(n)$ and thus on $x_j(n)$. The internal matrices cannot be arbitrarily and independently defined but are coupled so as to arrive at an internal process [23], [24].

The concept of an internal process was first introduced in the state space time-series modeling literature [2]–[4], [19]. In the time-series modeling context, internal models are of intense interest because they constitute a class of processes which is rich enough to include *minimal* models (i.e., those with the smallest possible state dimensions). While this is not the case for internal MAR models (i.e., a minimal MAR model may not belong to the class of internal models), there are, nonetheless, several good reasons why we are interested in internal MAR processes. As mentioned in Section I, internal processes are of both intellectual and practical interest. First, the theory for internal MAR processes is well-developed and provides the firm foundation upon which our work is built. Indeed, many of the concepts relating to internal processes developed in the state-space context have been generalized to MAR processes in [23]–[28].

Second, while noninternal MAR processes can be constructed, their states have exogenous random components, a property that is not suitable in many problems. In contrast, internal processes consist of states which are linearly related, a property that is essential to address certain problems, for instance when one wishes to estimate coarse-scale states which are, in fact, local averages of a fine-scale signal. In addition, with internality we can make clear the tie of the MAR framework to wavelet analysis.

Finally, internal processes provide a convenient parameterization for the MAR dynamics. From the internal matrices, the dynamics are related in a simple way to the covariance of the fine-scale process[5] $P_J \triangleq E[x_J x_J^T]$. The key to seeing this is recognizing that (1) represents the linear least squares estimate of $x_j(n)$ from $x_{j-1}([n/2])$ plus the estimation error $w_j(n)$. Therefore, the standard linear least squares formulas [38] dictate that

$$A_j(n) = P_{x_j(n)x_{j-1}([n/2])} P^{-1}_{x_{j-1}([n/2])} \tag{4}$$

$$Q_j(n) = P_{x_j(n)} - P_{x_j(n)x_{j-1}([n/2])} P^{-1}_{x_{j-1}([n/2])} \times P_{x_{j-1}([n/2])x_j(n)} \tag{5}$$

where P_z is our notation for the covariance matrix for random vector z and P_{uv} is the cross covariance matrix for random vectors u and v. The state covariances and cross-covariances follow trivially from (3) as

$$P_{x_j(n)} = W_j(n) P_J W_j(n)^T \tag{6}$$

$$P_{x_j(n)x_{j-1}([n/2])} = P^T_{x_{j-1}([n/2])x_j(n)} = W_j(n) P_J W_{j-1}([n/2])^T. \tag{7}$$

While a full theoretical development of internal MAR processes may be found in [24], we point out one aspect of this theory so as to avoid confusion. An MAR process has white noise-driven coarse-to-fine dynamics [cf. (1)]. Yet, for an internal MAR process, the parent state $x_j(n)$ is a linear function of its children states, a deterministic fine-to-coarse relationship. That is, for some matrix $V_j(n)$ we have $x_j(n) = V_j(n)\left[\begin{smallmatrix} x_{j+1}(2n) \\ x_{j+1}(2n+1) \end{smallmatrix}\right]$. That the driving noise $w_{j+1}(2n)$ and $w_{j+1}(2n+1)$ are uncorrelated and that $x_j(n)$ is deterministically related to $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$ is not contradictory because by construction [i.e., using (4)–(7)], the noise vector $\left[\begin{smallmatrix} w_{j+1}(2n) \\ w_{j+1}(2n+1) \end{smallmatrix}\right]$ is in the null-space of $V_j(n)$ [24].

In this paper, our primary interest is the *multiscale stochastic realization problem* (hereafter called simply the realization problem). The realization problem is to build an MAR process such that the fine-scale realized covariance matrix, P_J, matches a specific desired covariance matrix which we will denote by P_f. Of particular interest to us is the decorrelation role that the linear functionals [the rows of $W_j(n)$] play in the realization problem. To achieve $P_J = P_f$, the linear functionals which define the states of the MAR process must represent enough information to satisfy the Markov Property. However, for many problems this requires impractically large state dimensions, resulting in models which are not useful because of the high computational burden of sample-path generation, estimation, or likelihood calculation.

There are several ways to overcome this computational bottleneck. One way is to start with an exact model and then to reduce the state dimensions by throwing away some of the linear functionals which define the states [23]–[28]. One can then compute the dynamics of an MAR model by taking the resulting reduced row-dimension internal matrices and using them in (4)–(7) in which P_f is substituted for P_J. A consequence of reducing the state dimensions is that the Markov Property becomes an approximation and the resulting model is not exact $(P_J \neq P_f)$.

For exact internal models, throwing away elements of states will, in general, also destroy internality. This will occur if, for any node (j,n), the state $x_j(n)$ cannot be written as a linear function of its children states because the necessary information has been discarded in the process of state dimension reduction. This raises the following interesting question: how can low-order approximate models (i.e., those for which $P_J \approx P_f$) also be made internal?

Our approach to this question, in particular, and to the stochastic realization problem, in general, differs considerably from previous work [23]–[28] in several ways. In this previous work, the states of exact models are closely tied to the process being modeled in that the internal matrices are determined by solving optimization problems, the parameters of which are governed by the intricacies of fine-scale desired covariance P_f. Therefore, these internal matrices have no meaningful structure beyond the fact that they are solutions to specific optimization problems.

In contrast to these previous approaches, our internal matrices are not the solutions to optimization problems. Instead, they are selected from the small but rich class of wavelet bases. Therefore, they take no work to compute and have the intuitively pleasing and mathematically rich structure given by wavelets. Moreover, we directly design approximate internal models without first constructing an exact model and then discarding state information.

[5] $E[\cdot]$ is the expectation operator.

III. Wavelet Background

In this section, we give a brief review of wavelet decompositions. For more details see [17], [32], [34], and [36]. The wavelet representation of a continuous signal F consists of a sequence of approximations of F at coarser and coarser scales where the approximation at the jth scale consists of a weighted sum of shifted and dilated versions of a basic function ϕ called the scaling function. By considering the incremental details added in obtaining the $(j+1)$st scale approximation from the jth, one arrives at the wavelet transform based on a single function ψ called the analyzing wavelet.

The reconstruction is performed using the function $\tilde{\psi}$, called the synthesis wavelet, such that the two families $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ and $\{\tilde{\psi}_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ are a biorthogonal Riesz basis of $L^2(\mathbb{R})$ where

$$\psi_{j,n}(t) \triangleq \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^j n}{2^j}\right)$$

and similarly for $\tilde{\psi}_{j,n}$. The synthesis wavelet $\tilde{\psi}$ is obtained from the function $\tilde{\phi}$ which is dual to ϕ, i.e., which satisfies $\langle \phi(t), \tilde{\phi}(t-n)\rangle = \delta(n)$, where $\langle\cdot\rangle$ is the standard inner product in $L^2(\mathbb{R})$ and $\delta(\cdot)$ is the Dirac function. The scaling functions ϕ and $\tilde{\phi}$ must satisfy

$$\phi(t) = \sqrt{2}\sum_n h(n)\phi(2t-n)$$

and

$$\tilde{\phi}(t) = \sqrt{2}\sum_n \tilde{h}(n)\tilde{\phi}(2t-n)$$

where h and $\tilde{h}$ are discrete filters satisfying the biorthogonality condition in $\ell^2(\mathbb{Z})$

$$\sum_k h(k)\tilde{h}(k-2n) = \sum_k \tilde{h}(k)h(k-2n) = \delta(n). \quad (8)$$

ψ and $\tilde{\psi}$ are given by

$$\psi(t) = \sqrt{2}\sum_n g(n)\phi(2t-n)$$
$$\tilde{\psi}(t) = \sqrt{2}\sum_n \tilde{g}(n)\tilde{\phi}(2t-n)$$

where

$$\begin{cases} g(n) = (-1)^{1-n}\tilde{h}(1-n) \\ \tilde{g}(n) = (-1)^{1-n}h(1-n). \end{cases} \quad (9)$$

The discrete filters h, g, $\tilde{h}$, and $\tilde{g}$ must satisfy the perfect reconstruction condition which can be found in [32]. When $h = \tilde{h}$ and $g = \tilde{g}$, then h is a conjugate mirror filter (CMF) and the family $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ constitutes an orthonormal wavelet basis of $L^2(\mathbb{R})$.

The fast wavelet transform computes the wavelet coefficients of a discrete signal $a_J(\cdot)$. The fast wavelet decomposition algorithm is

$$\begin{cases} a_j(n) = \sum_p h(p-2n)a_{j+1}(p) \\ d_j(n) = \sum_p g(p-2n)a_{j+1}(p). \end{cases} \quad (10)$$

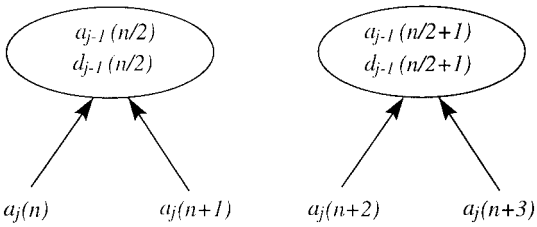


Fig. 2. The Haar dependency graph is a dyadic tree. Here, n is even.

The reconstruction algorithm is

$$a_{j+1}(n) = \sum_p \tilde{h}(n-2p)a_j(p) + \sum_p \tilde{g}(n-2p)d_j(p). \quad (11)$$

The coefficients $a_j(n)$ and $d_j(n)$ are called, respectively, the scaling and detail coefficients at the jth scale and nth shift.

In the remainder of this paper, we consider only the case when h and $\tilde{h}$ are finite impulse response (FIR) filters, i.e., when they have a finite number of nonzero coefficients. For the sake of notational simplicity, we assume that the lengths of h and $\tilde{h}$ are both even. Without loss of generality, we choose $\mathrm{supp}(h) = [-R+1, R]$ and $\mathrm{supp}(\tilde{h}) = [-\tilde{R}+1, \tilde{R}]$ for some integers R and $\tilde{R}$ such that $\tilde{R} \geq R \geq 1$. Thus, using (9) we have $\mathrm{supp}(\tilde{g}) = [-R+1, R]$ and $\mathrm{supp}(g) = [-\tilde{R}+1, \tilde{R}]$. We also assume that R and $\tilde{R}$ have the same parity. We point out, however, that all the results in this paper hold for all perfect reconstruction FIR filters with minor modifications.

The wavelet reconstruction algorithm (11) defines a dynamical relationship between the scaling coefficients $a_j(n)$ at one scale and those at the next finer scale, with the detail coefficients $d_j(n)$ acting as the input. Note that these dynamics are with respect to *scale* rather than time. This suggests that it is natural to think of constructing MAR processes within the wavelet framework. This construction is, in fact, obvious in the case of the Haar wavelet [21] because the dependency structure of the wavelet coefficients is a dyadic tree as shown in Fig. 2. Indeed, the wavelet reconstruction algorithm using the Haar system states that, for each $j = 1, \cdots, J$

$$a_j(n) = \frac{1}{\sqrt{2}}a_{j-1}([n/2]) + \frac{(-1)^{n-1}}{\sqrt{2}}d_{j-1}([n/2]). \quad (12)$$

Assuming that the detail coefficients are white noise, (12) suggests that one can build an MAR process [21] by defining each $x_j(n)$ as containing a scaling *and* detail coefficient at scale j and position n, i.e., $x_j(n) = [a_j(n)\ d_j(n)]^T$. For $j = 1, \cdots, J-1$, the auto-regression for this model is

$$x_j(n) = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & (-1)^{n+1} \\ 0 & 0 \end{bmatrix} x_{j-1}([n/2]) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_j(n) \quad (13)$$

where $w_j(n)$ represents $d_j(n)$. The auto-regression at the finest scale J is simply given by

$$x_J(n) = \frac{1}{\sqrt{2}}[1 \quad (-1)^{n+1}]x_{J-1}([n/2]). \quad (14)$$

The link between MAR processes and wavelets is not obvious if one considers wavelets other than the Haar system. This is due to the overlapping supports of such wavelets (which does not occur in the Haar case). Indeed, to compute a scaling

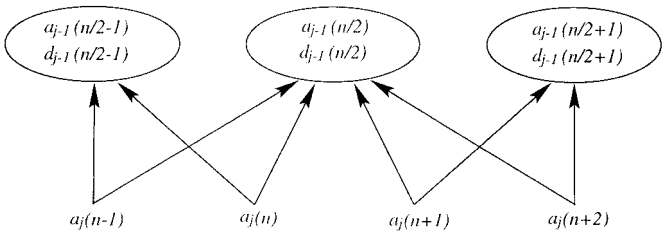


Fig. 3. Dependency graph for the Daubechies four-tap filter. Here n is even.

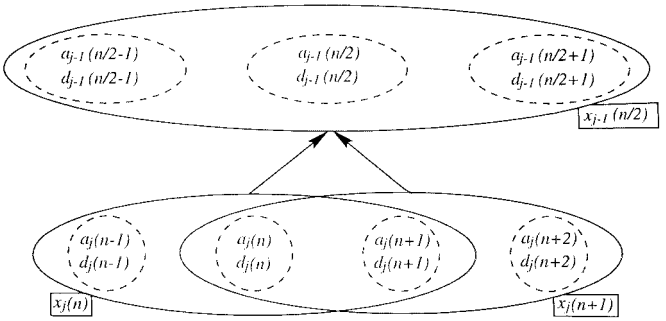


Fig. 4. Through state augmentation, the dependency graph for the Daubechies four-tap filter can be made into a tree. Here n is even.

coefficient $a_j(n)$ one does not need only the time-synchronous parents[6] of $a_j(n)$ but also a number of neighboring coefficients depending on the supports of the analysis and synthesis wavelet. Thus, if we build a multiscale process where the states are defined as in the Haar case, i.e., $x_j(n) = [a_j(n)\ \ d_j(n)]^T$, but where we consider that the scaling and detail coefficients are computed using more regular wavelets, we will end up with a more complex graph structure of the scale-to-scale autoregression instead of a tree one as imposed by (1). This is illustrated in Fig. 3 in the case of the Daubechies four-tap filter [17].

The first issue, then, is to redefine the states in such a way as to arrive at a tree dependency structure rather than a more complex graph. We will see that this can be done easily using state augmentation. The second issue we must address, which is the crucial issue and the most difficult one, is how to provide internality. As mentioned in Section I, this is, in fact, one of our main objectives since we want to use wavelets as the linear functionals that define the internal states. These two issues will be the focus of the next section.

IV. Wavelet-Based Internal MAR Processes

In this section, we first address (in subsection A) the issue of defining the states in such way to obtain a tree dependency structure. We then address (in subsection B) the issue of internality.

[6]The time-synchronous parents of $a_j(n)$ [or $d_j(n)$] are $a_{j-1}([n/2])$ and $d_{j-1}([n/2])$.

A. Tree Structure

To see the intuition behind how to define the states in order to arrive at a tree structure, let us consider the simple case where h is the Daubechies four-tap filter [17]. In this case, we have $\mathrm{supp}(h) = [-1, 2]$. Then, the wavelet reconstruction algorithm (11) implies that for every even integer n

$$\begin{cases} a_j(n-1) = \displaystyle\sum_{p=\frac{n}{2}-1}^{\frac{n}{2}} h(n-2p-1)a_{j-1}(p) \\ \qquad\qquad + \displaystyle\sum_{p=\frac{n}{2}-1}^{\frac{n}{2}} g(n-2p-1)d_{j-1}(p) \\ a_j(n) = \displaystyle\sum_{p=\frac{n}{2}-1}^{\frac{n}{2}} h(n-2p)a_{j-1}(p) \\ \qquad\qquad + \displaystyle\sum_{p=\frac{n}{2}-1}^{\frac{n}{2}} g(n-2p)d_{j-1}(p) \\ a_j(n+1) = \displaystyle\sum_{p=\frac{n}{2}}^{\frac{n}{2}+1} h(n-2p+1)a_{j-1}(p) \\ \qquad\qquad + \displaystyle\sum_{p=\frac{n}{2}}^{\frac{n}{2}+1} g(n-2p+1)d_{j-1}(p) \\ a_j(n+2) = \displaystyle\sum_{p=\frac{n}{2}}^{\frac{n}{2}+1} h(n-2p+2)a_{j-1}(p) \\ \qquad\qquad + \displaystyle\sum_{p=\frac{n}{2}}^{\frac{n}{2}+1} g(n-2p+2)d_{j-1}(p). \end{cases} \tag{15}$$

Therefore, for every $j = 0, \cdots, J$ and for every $n = 0, \cdots, 2^j - 1$, if we chose each state $x_j(n)$ to be

$$x_j(n) = [a_j(n-1), a_j(n), a_j(n+1), d_j(n-1), d_j(n), d_j(n+1)]^T$$

it is clear from (15) that the scaling coefficients carried within each $x_j(n)$ depend only on the parent $x_{j-1}([n/2])$ of $x_j(n)$ (see Fig. 4).

In the general case, for every $j = 0, \cdots, J$ and for every $n = 0, \cdots, 2^j - 1$, the state at scale j and shift n is defined as shown in (16), at the bottom of this page. The details showing that (16) implies that each state depends only on its parent can be found in Appendix A. We then can show that

$$\begin{cases} x_j(n) = A_j(n)x_{j-1}([n/2]) + w_j(n), & \text{if } 0 \le j < J \\ x_j(n) = A_j(n)x_{j-1}([n/2]), & \text{if } j = J \end{cases} \tag{17}$$

$$\begin{cases} x_j(n) = \left[a_j(n-\tilde{R}+1), \cdots, a_j(n+\tilde{R}-1), d_j\left(n-\frac{\tilde{R}+R}{2}+1\right), \cdots, d_j\left(n+\frac{\tilde{R}+R}{2}-1\right)\right]^T, & \text{if } 0 \le j < J \\ x_j(n) = a_j(n), & \text{if } j = J \end{cases} \tag{16}$$

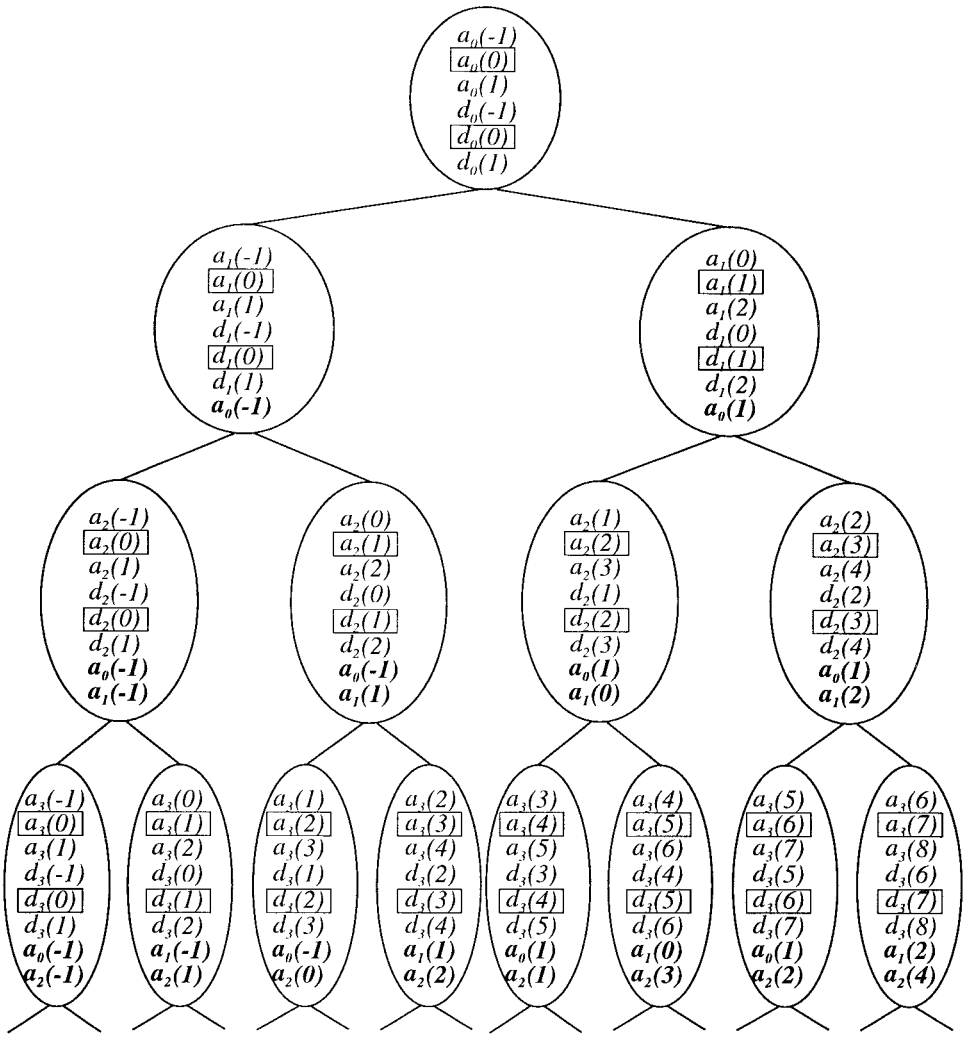


Fig. 5. Example of the internal MAR-wavelet process with the Daubechies four-tap filter. Scaling coefficients in bold illustrate the necessary information transmitted from one scale to the next. The boxed coefficients are a linear function of the coefficients of their children by the wavelet decomposition algorithm.

where

$$w_j(n) = \Bigg[\underbrace{0, \cdots, 0}_{2\tilde{R}-1 \text{ times}}, d_j\left(n - \frac{\tilde{R}+R}{2} + 1\right), \cdots, d_j\left(n + \frac{\tilde{R}+R}{2} - 1\right) \Bigg]^T. \tag{18}$$

The proof of (17) and the expression for the matrices $A_j(n)$ can also be found in Appendix A.

Assuming that $w(\cdot)$ is a white noise process uncorrelated with the root node state $x_0(0)$, (17) represents an MAR process with dynamics matching the reconstruction algorithm associated with any compactly supported orthogonal or biorthogonal wavelet. In the sequel, we refer to this process as *the standard MAR-wavelet process*.

B. The Internal MAR-Wavelet Process

If the coefficients $d_j(n)$ are considered as the detail coefficients computed using the wavelet decomposition algorithm and thus are deterministic inputs, then (17), (18) is just a rewriting of the wavelet reconstruction algorithm (11). But if the coefficients $d_j(n)$ are generated as random variables, then (17) and (18) constitute a statistical *model* for a fine-scale process. However, almost surely, the states generated by this model do not consist of scaling and detail coefficients[7] of the realized fine-scale process x_J. This is because the standard MAR-wavelet process is *not* internal. Indeed, as mentioned in Section II, a necessary and sufficient condition for internality is that each state has to be a linear functional of its immediate children [23]–[25]. From the wavelet decomposition algorithm, one can easily see that, when $\tilde{R} > 1$, each state of the standard MAR-wavelet process is not a linear functional of just its immediate children but of the range of states at the next finer scale depending on the supports of the scaling functions.

The question now is how to build an internal MAR-wavelet process in order to ensure that the states consist of scaling and detail coefficients of the realized fine-scale process. This issue is, in fact, the one which seemed to doom the union between MAR process and wavelets, and the central contribution of

[7] For each state $x_j(n)$, only $a_j(n)$ and $d_j(n)$ can be expressed as linear functions of the children states.

this paper is to solve this problem. This will be done by exhibiting and exploiting some relationships between scaling and detail coefficients and by appropriately modifying the state definition. We emphasize that this is purely deterministic analysis.

The intuition that leads us to construct internal states is essentially state augmentation. For the coarse-to-fine synthesis, each state needs to have all the necessary information. This was done in the standard MAR-wavelet process. We need now to make sure that each state has all the information for *analysis*, so that we have internality. As explained above, the states of the standard MAR-wavelet process do not contain enough information. We thus need to augment the states in order to achieve internality. Before showing how we augment the states, we need the following intermediate result which allows us to add only a few coefficients in the process of defining internal states.

Proposition 1: There exists four matrices I_1, J_1, I_2, J_2 such that

$$\begin{pmatrix} d_j\left(n-\frac{\tilde{R}+R}{2}+1\right) \\ \vdots \\ d_j(n-1) \end{pmatrix} = I_1 \begin{pmatrix} a_j(n-\tilde{R}+1) \\ \vdots \\ a_j\left(n-\frac{\tilde{R}-R}{2}-1\right) \end{pmatrix} + J_1 \begin{pmatrix} a_{j+1}(2n-\tilde{R}+1) \\ \vdots \\ a_{j+1}(2n+\tilde{R}-2) \end{pmatrix} \tag{19}$$

$$\begin{pmatrix} d_j(n+1) \\ \vdots \\ d_j\left(n+\frac{\tilde{R}+R}{2}-1\right) \end{pmatrix} = I_2 \begin{pmatrix} a_j\left(n+\frac{\tilde{R}-R}{2}+1\right) \\ \vdots \\ a_j(n+\tilde{R}-1) \end{pmatrix} + J_2 \begin{pmatrix} a_{j+1}(2n-\tilde{R}+3) \\ \vdots \\ a_{j+1}(2n+\tilde{R}) \end{pmatrix}. \tag{20}$$

Proof: See Appendix B. □

The idea behind constructing internal states is to define new states $x_j(n)$ in such a way so that the left (respectively, right) child of $x_j(n)$ contains $\underline{a}_j^l \triangleq [a_j(n-\tilde{R}+1), \cdots, a_j(n-\frac{\tilde{R}-R}{2}-1)]$ (respecticely, $\underline{a}_j^r \triangleq [a_j(n+\frac{\tilde{R}-R}{2}+1), \cdots, a_j(n+\tilde{R}-1)]$. However, having copied $\underline{a}_j^l$ and $\underline{a}_j^r$ from $x_j(n)$ to its children $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$, we must continue to pass $\underline{a}_j^l$ and $\underline{a}_j^r$ down to the children (and grandchildren and so on) of $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$ to maintain internality. Of course, we must do this for all j and n. This seems to suggest that the state dimensions will explode. However, by simply splitting at each step the necessary information between the two children, the state dimension remains bounded. The construction of the states is depicted in Fig. 5 in the simple case of Daubechies four-tap filter. To define rigorously the internal states in the general case, for $j = 0, \cdots, 2^j - 1$, we define recursively the sequence of vectors $U_j(n)$ as shown in (21) at the bottom of this page. We then have the final result in which we show in bold the augmentation of the states with respect to the standard MAR-wavelet model. This is shown in (22)

Proposition 2: The MAR process for which the states are defined by

$$\begin{aligned} x_0(0) = \Bigg(& a_0(-\tilde{R}+1), \cdots, a_0(\tilde{R}-1), \\ & d_0\left(-\frac{\tilde{R}+R}{2}+1\right), \cdots, d_0\left(\frac{\tilde{R}+R}{2}-1\right)\Bigg)^T \\ x_1(n) = \Bigg(& a_1(n-\tilde{R}+1), \cdots, a_1(n+\tilde{R}-1), \\ & d_1\left(n-\frac{\tilde{R}+R}{2}+1\right), \cdots, \\ & d_1\left(n+\frac{\tilde{R}+R}{2}-1\right), \boldsymbol{U_0^{n+1}(0)}\Bigg)^T \end{aligned}$$

$$U_0(0) = \Bigg(\underbrace{a_0(-\tilde{R}+1), \cdots, a_0\left(-\frac{\tilde{R}-R}{2}-1\right)}_{U_0^1(0)}, \underbrace{a_0\left(\frac{\tilde{R}-R}{2}+1\right), \cdots, a_0(\tilde{R}-1)}_{U_0^2(0)} \Bigg)$$

$$U_j(n) = \begin{cases} \Bigg(\underbrace{U_{j-1}^1(n/2)}_{U_j^1(n)}, \underbrace{a_{j-1}(n/2-\tilde{R}+1), \cdots, a_{j-1}\left(n/2-\frac{\tilde{R}-R}{2}-1\right)}_{U_j^2(n)} \Bigg), & \text{if } n \text{ is even} \\ \Bigg(\underbrace{U_{j-1}^2([n/2])}_{U_j^1(n)}, \underbrace{a_{j-1}\left([n/2]+\frac{\tilde{R}-R}{2}+1\right), \cdots, a_{j-1}([n/2]+\tilde{R}-1)}_{U_j^2(n)} \Bigg), & \text{if } n \text{ is odd} \end{cases} \tag{21}$$

$$x_j(n) = \Bigg(a_j(n-\tilde{R}+1), \cdots, a_j(n+\tilde{R}-1), d_j\left(n - \frac{\tilde{R}+R}{2}+1\right), \cdots, d_j\left(n+\frac{\tilde{R}+R}{2}-1\right), \boldsymbol{U_j(n)} \Bigg)^T \tag{22}$$

is internal.

Proof: See Appendix C. □

We refer to this new process as *the internal MAR-wavelet process*. Notice that the size of each $U_j(n)$ is $\tilde{R}+R-2$. Thus the maximal state dimension of the internal MAR-wavelet process is $4\tilde{R}+2R-4$.

With Proposition 2, we have shown how to build internal MAR processes based on any compactly supported orthogonal or biorthogonal wavelet. This completes our unification of wavelets with MAR processes.

V. Approximate MAR-Wavelet Models for Stochastic Processes

In this section, we focus on the construction of approximate MAR-wavelet models for stochastic processes in order to take advantage of the fast statistical signal processing algorithms associated with the MAR framework. In our examples, we will use the fast sample-path generator and the fast linear least squares estimator for estimating a signal from irregularly spaced, sparse measurements corrupted by nonstationary noise.

The standard MAR-wavelet process, defined by (17) and (18), can be used as an approximate model for a stochastic process by assuming that the detail coefficients are white. We call this model the *standard MAR-wavelet model*. However, the states realized using this model are not consistent with the fine-scale realized process x_J in that they do not represent, with probability one, scaling and detail coefficients of x_J. This is because of the lack of internality, as discussed in Section IV-B. Notice that the assumption of the whiteness of $w(\cdot)$ [defined by (18)] is never fulfilled if $\tilde{R} > 1$. Indeed, for n and m such that $0 < |n-m| \leq R+\tilde{R}-2$, it is clear that, at a given scale j, $w_j(n)$ and $w_j(m)$ are correlated since they share at least one detail coefficient.

By achieving internality, the states of the internal MAR-wavelet process (22) are forced to be consistent with the fine-scale realized process. We can then examine the problem of building internal stochastic models that are consistent with the graph structure of the tree (i.e., we insist on models in which we have the Markov Property) and approximate the given statistics of a fine-scale process. Given these fine-scale statistics, internality provides immediately the statistics of any MAR state and the statistics between each state and its parent, as each state is a set of linear functionals of the fine scale. As a result, we can immediately define the linear dynamics of an MAR model of the form (1). The dynamics of this model incorporate optimal prediction from parent to child using the implied statistics, and we then model the errors in this prediction as white (the requirement for the Markov Property to be satisfied).

The implications of this are twofold. First, the resulting internal model in general produces fine-scale statistics that only approximate the desired ones (because of our insistence that the coarse-to-fine prediction errors be white). To be sure, our internal MAR-wavelet model *does* produce the correct marginal statistics at each node and the correct joint statistics for each state and its parent, but other statistics (e.g., cross-covariance for two nodes at the same scale) are only captured approximately. The second point is that the coarse-to-fine dynamics so defined are in general *very* different from standard wavelet modeling. In particular, these dynamics exploit correlation between detail coefficients *and* coarser scale scaling and detail coefficients by performing optimal prediction and then assuming that the errors in these predictions are white. This is in contrast to one common approach in using wavelets for modeling stochastic processes in which the detail coefficients are themselves modeled as white (i.e., the wavelet representation is *assumed* to be the Karhunen–Loeve (K–L) decomposition). In our case, since we allow MAR dynamics, we do not *need* to have K–L diagonalization. Rather, the success of our method in approximating stochastic processes relies only on the weaker requirement that the *errors* in predicting finer scale detail coefficients from coarser scale coefficients are white. As we illustrate, an implication of this is that we can use fairly short wavelets, implying lower state dimensions, which certainly do *not* do a good job of whitening the details (as evidenced by our results using the noninternal standard MAR-wavelet models), but which do remarkably well for our internal models.

Using this optimal prediction procedure, we incorporate a synthesis algorithm for the detail coefficients in addition to the usual wavelet reconstruction algorithm for the scaling coefficients. The initialization for this new synthesis algorithm is given by the statistics of the scaling and detail coefficients at the coarsest scale. Those statistics are given by the covariance matrix $P_{x_0(0)}$ of the root node state which is computed using (6) (in which P_J is replaced by P_f).

More precisely, the optimal prediction is performed as follows: if $D_j(n)$ represents the detail coefficients carried by the state $x_j(n)$ defined by (22), then

$$D_j(n) = P_{D_j(n)x_{j-1}([n/2])} P^{-1}_{x_{j-1}([n/2])} x_{j-1}([n/2]) + \tilde{w}_j(n) \tag{23}$$

where the covariance matrix for $\tilde{w}_j(n)$ is

$$P_{\tilde{w}_j(n)} = P_{D_j(n)} - P_{D_j(n)x_{j-1}([n/2])} P^{-1}_{x_{j-1}([n/2])} \times P^T_{D_j(n)x_{j-1}([n/2])}$$

and where $P_{D_j(n)}$, $P_{D_j(n)x_{j-1}([n/2])}$, and $P_{x_{j-1}([n/2])}$ are submatrices of $P_{x_j(n)}$ and $P_{x_j(n)x_{j-1}([n/2])}$ which are computed using (6) and (7) in which P_J is replaced by P_f, the covariance matrix for the signal being modeled.[8] The prediction errors $\tilde{w}_j(n)$ are not white in general. This can be easily seen from the fact that the states of the internal MAR-wavelet model

[8]The internal matrices required in (6) and (7) are implicitly given by Proposition 2.

contain duplicated detail coefficients. Yet, we assume that the prediction errors $\tilde{w}_j(n)$ in (22) are white noise to legitimately apply the signal processing algorithms provided by the MAR framework. We thus have an approximate MAR model for any process whose second-order statistics, P_f, are known. We will refer to this model as the *internal MAR-wavelet model*. Note that an advantage of the internal MAR-wavelet model and, in fact, of any internal model is that it achieves the correct variances (i.e., the diagonal elements of P_J match *exactly* those of P_f).

We emphasize another important point. In real world problems, the user may not know how to choose the appropriate wavelet to do a good job in decorrelating the process under study. Thus, the resulting detail and scaling coefficients may be strongly correlated. We point out that, even in this case, our internal MAR-wavelet model can be an accurate one for the underlying process by exploiting these potential correlations, as well as the correlations between detail and scaling coefficients, in the optimal prediction scheme. We will illustrate this later in the case of an fBm.

To be more precise, let us discuss this optimal prediction algorithm in the simple case of the Haar wavelet and support this discussion by showing some examples for fBm. We will consider fBm defined on the time interval $(0, 1]$ and we normalize the fBm statistics to have unit variance at time one. The covariance function for fBm [33] is

$$r_H(t_1, t_2) = \frac{1}{2}(|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}). \qquad (24)$$

Fig. 6(a) and (b) shows P_f for 64 samples of fBm on the interval $(0, 1]$ with Hurst parameter $H = 0.3$ and $H = 0.7$, respectively.

A. The Haar Case

The MAR-Haar model defined in Section III by (13) and (14) has many drawbacks. First, the assumption that the detail coefficients are white is very poor in general. Indeed, since the Haar wavelet has only one vanishing moment, for most processes the resulting detail coefficients are strongly correlated both in space and scale. Second, due to the piecewise constant shape of the Haar wavelet, any realized covariance matrix (i.e., the covariance matrix of x_J) with this model will have "blockiness" in general. Thus, any sample-path generated using this model will have distracting artifacts and will be blocky. Therefore, this model is not appropriate for synthesizing or estimating stochastic processes in general. An illustration of this phenomenon is shown in Fig. 7(a) which displays P_J for an fBm with Hurst parameter $H = 0.3$. It is clear by comparing Fig. 6(a) and Fig. 7(a) that P_J is a poor approximation to P_f. We point out, however, that this model has been successfully used for hypothesis discrimination. In [21] the authors applied the MAR likelihood calculator to accurately estimate the Hurst parameter of fBm.

Now, notice that this model is internal, which is clear from the wavelet decomposition algorithm associated with the Haar system. Therefore, instead of using the dynamics defined by (13) and (14), one can build a more accurate model [14] by computing the auto-regression parameters so that the multiscale auto-regression is the optimal prediction of $x_j(n)$ from $x_{j-1}([n/2])$. Thus, while $w_j(n)$ in the model defined by (13) and (14) represents the detail coefficient $d_j(n)$, the process noise in the internal model based on optimal prediction will represent the prediction error in the estimation of $d_j(n)$ *conditioned* on the detail and scaling coefficient represented by $x_{j-1}([n/2])$, i.e.,

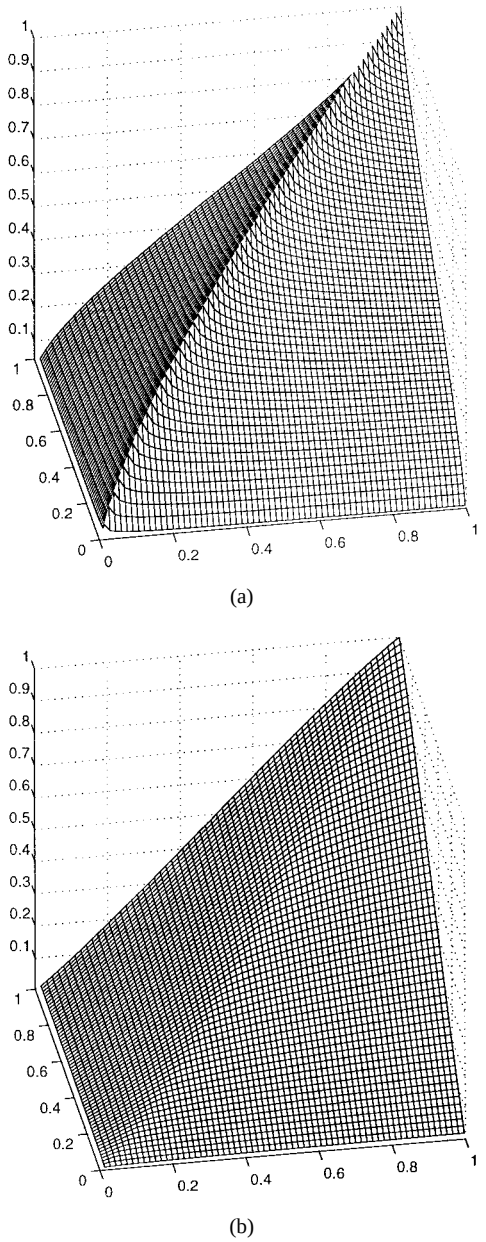
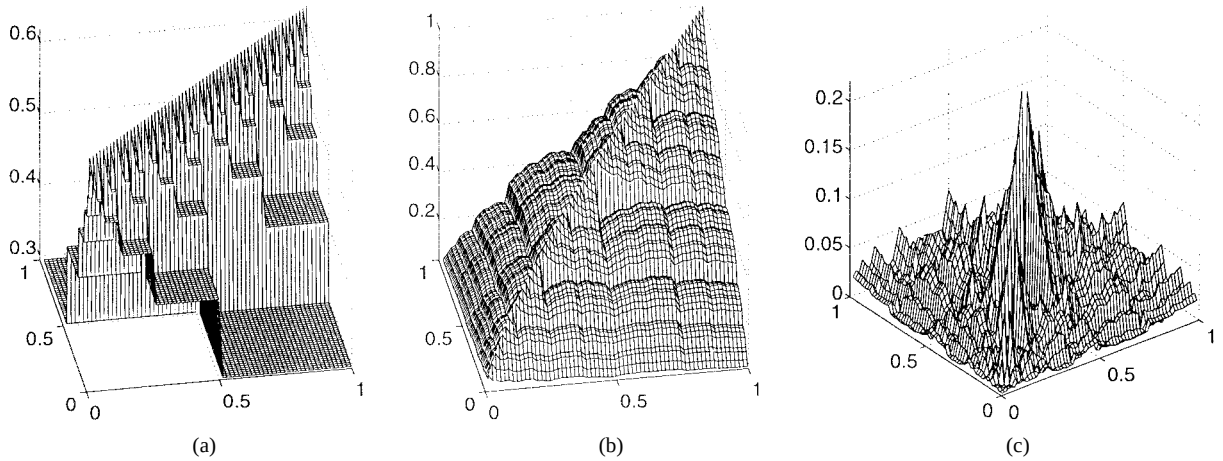


Fig. 6. Exact covariance matrices for fBm. (a) $H = 0.3$. (b) $H = 0.7$.

$$Q_j(n) = \begin{bmatrix} 0 & 0 \\ 0 & \mathrm{var}[\tilde{w}_j(n)] \end{bmatrix}$$

where, from the linear least squares estimation error formula, we have

$$\mathrm{var}[\tilde{w}_j(n)] = \mathrm{var}[d_j(n)] - P_{d_j(n)x_{j-1}([n/2])} P^{-1}_{x_{j-1}([n/2])} \times P_{x_{j-1}([n/2])d_j(n)}.$$



(a) (b) (c)

Fig. 7. Realized covariance matrices P_J for fBm with $H = 0.3$. (a) P_J using the standard MAR-Haar model. (b) P_J using the internal MAR-Haar model. (c) $|P_f - P_J|$ where P_J is from (b).

Therefore, this multiscale model will capture correlations in scale among the detail coefficients represented by states at neighboring nodes in the tree. Thus, this optimal prediction model will do a better job of approximating the statistics of the underlying process than does the model defined by (13) and (14). The improvement is illustrated in the case of fBm in Fig. 7(b) which displays the realized covariance matrix of an fBm with Hurst parameter $H = 0.3$ and should be compared to Fig. 7(a). This shows the power of the optimal prediction procedure in the simple case of the Haar wavelet.

B. The General Case

It is worth noting that, with the Haar wavelet, the detail coefficients which are not neighbors (in space and scale) are in general also strongly correlated. Therefore, even with this optimal prediction procedure, the internal MAR–Haar model remains very crude since it captures only the correlations between a detail coefficient at a given scale and the time-synchronous detail and scaling coefficients at the previous coarser scale. Fig. 7(b) and (c) illustrates the limitation of this model in the case of fBm with $H = 0.3$. One sees that the realized covariance matrix is still a poor approximation of the true one.

One way to overcome the limitations of the Haar wavelet is to build an internal MAR-wavelet model using an analyzing wavelet with a large number of vanishing moments. With such a wavelet, the detail coefficients which are not neighbors in space and scale will, in general, be better decorrelated and the potential correlations will reside only between neighboring coefficients. Then, our optimal prediction procedure will exploit these residual correlations between detail and scaling coefficients and do the best job in linearly predicting the detail coefficients.

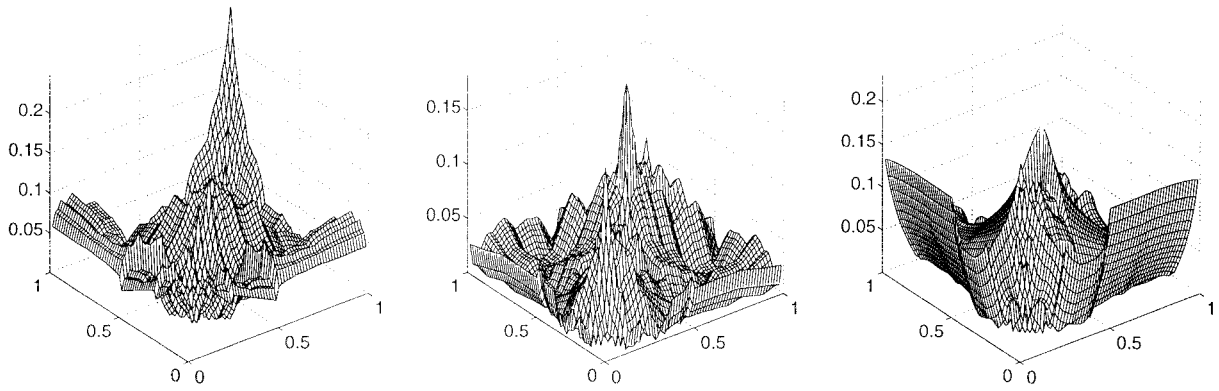
However, this is not the only solution. One can still build accurate models without using an analyzing wavelet with large number of vanishing moments. Indeed, all we need to have accurate models is to provide a good approximation to the Markov Property. Therefore, accurate models will be provided using *any* wavelet yielding scaling and detail coefficients such that the states they form approximatively fulfill the *conditional* decorrelation role of the Markov Property.

To support these arguments and to illustrate the performance of the optimal prediction procedure, we apply our MAR-wavelet models to approximate the statistics of fBm using different wavelets. Specifically, we compare the internal MAR-wavelet model to the standard MAR-wavelet model which assumes the whiteness of the detail coefficients. We use the Daubechies orthogonal wavelet with two vanishing moments (Daub4), the Daubechies orthogonal wavelet with three vanishing moments (Daub6), the spline biorthogonal wavelet (Spline13) such that[9] $\tilde{H}(z)$ [respectively, $H(z)$] has three (respectively, 1) zeros at $z = -1$, and the spline biorthogonal wavelet (Spline31) such that $\tilde{H}(z)$ [respectively, $H(z)$] has 1 (respectively, three) zeros at $z = -1$.

Fig. 8(a)–(c) displays the element-wise absolute value of the difference between P_f and P_J obtained by the standard MAR-wavelet model for an fBm with $H = 0.3$ using, respectively, Daub4, Daub6, and Spline13. The improvement with respect to the standard MAR–Haar model is clear as expected, since we are using analyzing wavelets with more than one vanishing moment. However, the approximation is not satisfactory which is not surprising since the detail coefficients are not exactly decorrelated using these wavelets. Finally, note that Daub6 does better than Spline13 because Daub6 is an orthogonal wavelet and is smoother than the analyzing wavelet of Spline13.

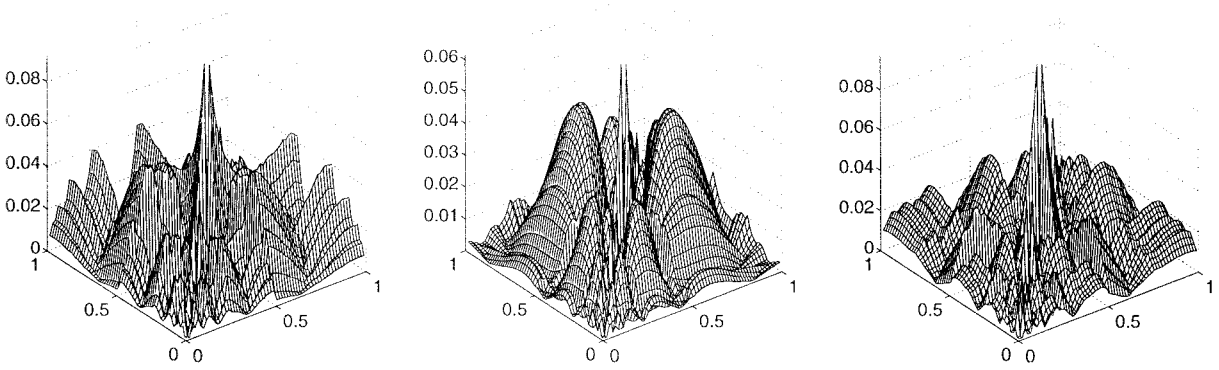
Now, with the internal MAR-wavelet model, the detail coefficients are no longer assumed to be white noise. Instead, they are computed using the optimal prediction procedure described above. Therefore, the internal MAR-wavelet model will better approximate the statistics of fBm. Fig. 9(a)–(c) displays the element-wise absolute value of the difference between P_f and P_J for an fBm with $H = 0.3$ using, respectively, Daub4, Daub6, and Spline13. The improvement with respect to the standard MAR-wavelet model is clear. Fig. 10(a)–(c) displays the same element-wise absolute value

[9] $H(z)$ (respectively, $\bar{H}(z)$) is the z-transform of $h(n)$ (respectively, $\bar{h}(n)$).



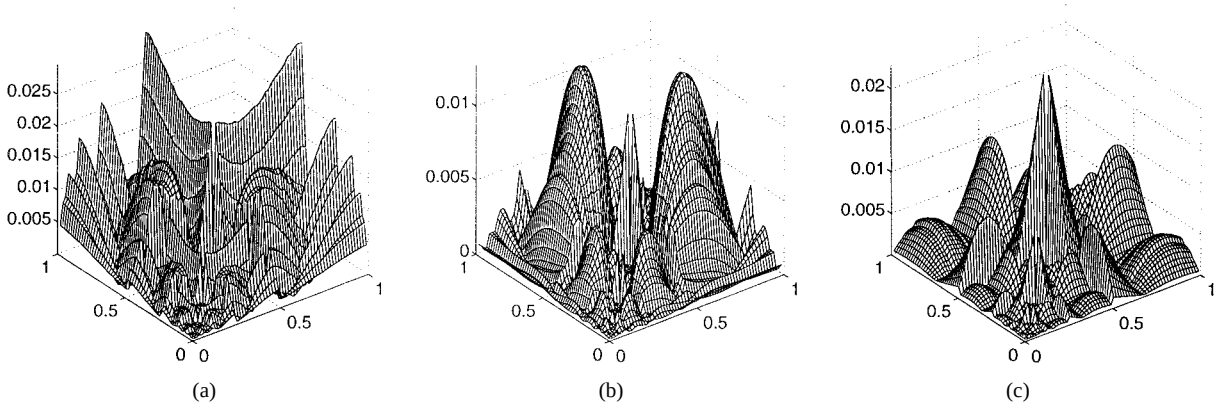
(a) (b) (c)

Fig. 8. $|P_f - P_J|$ for fBm with $H = 0.3$ using the standard MAR-wavelet model. (a) Daub4 (state dimension six). (b) Daub6 (state dimension ten). (c) Spline13 (state dimension eight).



(a) (b) (c)

Fig. 9. $|P_f - P_J|$ for fBm with $H = 0.3$ using the internal MAR-wavelet model. (a) Daub4 (state dimension eight). (b) Daub6 (state dimension 14). (c) Spline13 (state dimension ten).



(a) (b) (c)

Fig. 10. $|P_f - P_J|$ for fBm with $H = 0.7$ using the internal wavelet model. (a) Daub4 (state dimension eight). (b) Daub6 (state dimension 14). (c) Spline13 (state dimension ten).

of the difference obtained using the internal MAR-wavelet model for fBm with $H = 0.7$.

To illustrate, in the case of fBm, the fact that even with relatively nonregular wavelets our internal MAR-wavelet model can provide very accurate models, we use the biorthogonal wavelet Spline31. The analyzing wavelet for Spline31 has only one vanishing moment and the synthesis wavelet is extremely singular (see Fig. 11). Fig. 12(a) displays the element-wise absolute value of the difference between P_f and P_J using the standard MAR-wavelet model. One sees that the approximation is extremely bad, which is not surprising given the properties of Spline31 and the weakness of the assumption that the detail coefficients are white. However, using the internal MAR-wavelet model, the approximation is very accurate as displayed in Fig. 12(b). Furthermore, notice that this approximation is more accurate than the one illustrated in Fig. 10(c) in which the state dimension is larger. Indeed, in Fig. 12(b) we have $R = \tilde{R} = 2$ and thus the

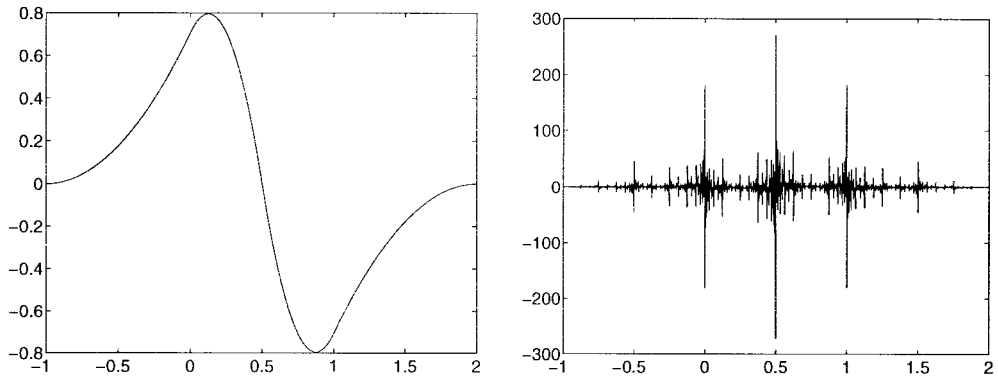


Fig. 11. Spline biorthogonal wavelet (Spline31). (a) Analyzing wavelet. (b) Synthesis wavelet.

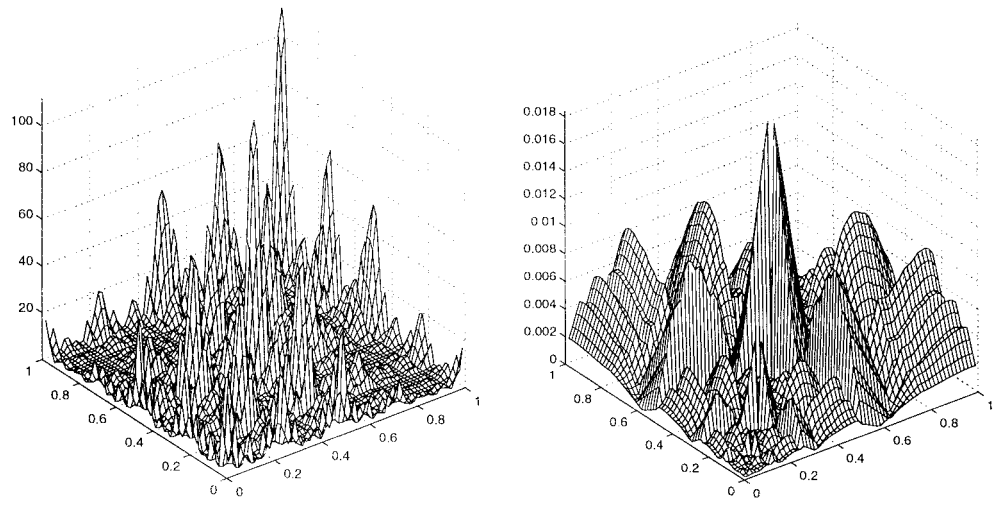


Fig. 12. $|P_f - P_J|$ for fBm with $H = 0.7$ using (a) the standard MAR-wavelet model with Spline31 (state dimension six) and (b) the internal MAR-wavelet model with Spline31 (state dimension eight).

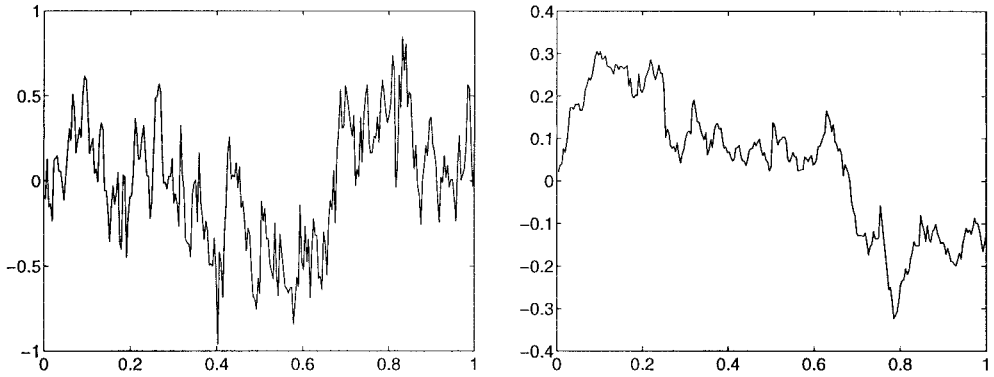
maximum state dimension is eight while in Fig. 10(c) we have $R = \tilde{R} = 3$ and thus the maximum state dimension is ten. This shows the power of the optimal prediction procedure in approximating the Markov Property even without considering analyzing wavelets with a large number of vanishing moments.

Now, we use the fast signal processing algorithms associated with the MAR framework to synthesize fBm sample-paths and to perform estimation from irregularly-spaced and sparse measurements corrupted by nonstationary noise. We emphasize that the latter is a problem which *cannot* be easily and efficiently handled with other estimation techniques due to the nonstationarity of the process to be estimated and the process noise and the irregularity of the measurements. Fig. 13(a) and (b) displays 256-point sample-paths using the internal MAR-wavelet model with Daub6 for an fBm with $H = 0.3$ and $H = 0.7$, respectively. Fig. 14(a) displays an exact 64-point realization[10] of fBm with $H = 0.3$. Fig. 14(b) displays noisy observations of (a) where observations are only available on $(0, 1/3]$ (over which the white measurement noise has variance 0.3) and $(2/3, 1]$ (over which the white measurement noise has variance 0.5). Fig. 14(c) displays the MAR estimates based on 14(b) using the internal wavelet model with Daub6. The MAR estimates are the solid line and the optimal estimates based on the exact statistics[11] are the dash–dot line. The plus/minus one standard deviation error bars are the dashed line. Fig. 15 illustrates the same processing but for fBm with $H = 0.7$. Notice that in both Fig. 14(c) and in Fig. 15(c) the optimal estimate based on the exact statistics is not easily distinguishable from the MAR estimate since the two nearly coincide. Also, the estimation error standard deviations that the MAR estimator provides are very close to the ones based on the exact statistics (although we have not plotted the latter in our examples). More importantly, the difference between the optimal estimate and the MAR estimate is well within the one standard deviation error bars. This demonstrates that the degree to which our internal MAR-wavelet model deviates from the exact model is statistically irrelevant.

Finally, we point out that in the case of finite length signals, folded discrete transforms [32] can be applied in the case of

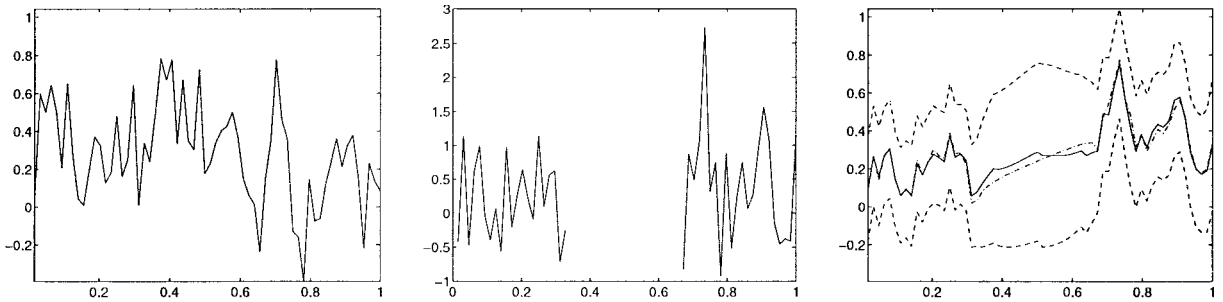
[10] Exact realizations of fBm are obtained by multiplying white Gaussian noise by the matrix square root of P_f. This requires $O(N^3)$ computations if P_f is $N \times N$. In contrast, the MAR sample-path generator is $O(N)$.

[11] That is, the optimal estimates are obtained by solving the normal equations based on the true fBm and measurement statistics. Note that solving the normal equations requires $O(N^3)$ computations while the MAR estimator is $O(N)$ where N is the size of the signal to be estimated [7]–[9].



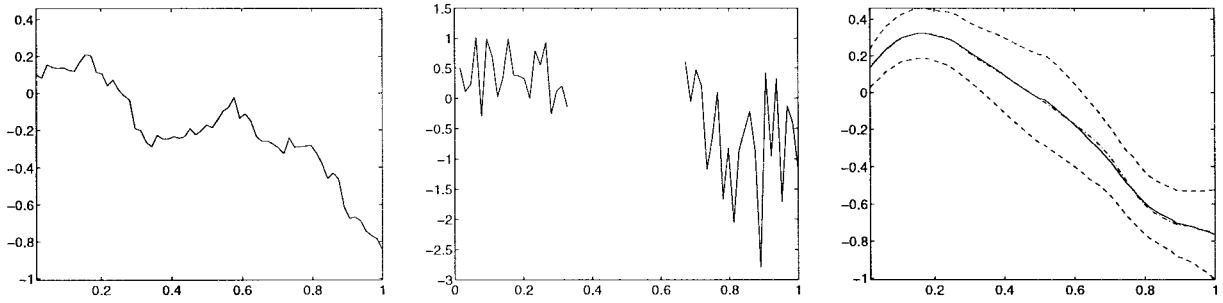
(a) (b)

Fig. 13. fBm sample-paths using the internal MAR-wavelet model with Daub6. (a) $H = 0.3$. (b) $H = 0.7$.



(a) (b) (c)

Fig. 14. MAR estimation for fBm with $H = 0.3$ using the internal MAR-wavelet model with Daub6. (a) Sample-path using exact statistics. (b) Noisy, irregular, and sparse observations of (a). The noise variance over $(0, 1/3]$ is 0.3 and over $(2/3, 1]$ is 0.5. (c) MAR estimates are the solid line and optimal estimates based on the exact statistics are the dash–dot line. The plus/minus one standard deviation error bars are the dashed line.



(a) (b) (c)

Fig. 15. MAR estimation for fBm with $H = 0.7$ using the internal MAR-wavelet model with Daub6. (a) Sample-path using exact statistics. (b) Noisy, irregular, and sparse observations of (a). The noise variance over $(0, 1/3]$ is 0.3 and over $(2/3, 1]$ is 0.5. (c) MAR estimates are the solid line and optimal estimates based on the exact statistics are the dash–dot line. The plus/minus one standard deviation error bars are the dashed line.

symmetric or antisymmetric biorthogonal wavelets. Boundary wavelets [11] can be used in the case of orthogonal wavelets. We used the folded discrete transform in both cases. In all the examples we considered, the results were very close to those obtained by assuming the knowledge of the statistics of the signal over $(-\infty, \infty)$. However, at this point, we have no precise mathematical description for the influence of the boarder treatment techniques on the accuracy of the models.

VI. Conclusion

The primary contribution of this paper has been to provide a unification of the MAR framework with the wavelet framework. We have shown how to construct internal MAR processes based on any compactly supported orthogonal or biorthogonal wavelet. We then used these internal MAR-wavelet processes as approximate models for stochastic processes. The marriage of the MAR and wavelet frameworks led us to incorporate a powerful reconstruction algorithm for the detail coefficients which complements the usual wavelet reconstruction algorithm for the scaling coefficients. While we have used fBm as a vehicle to illustrate the performance of our internal MAR-wavelet models, these models can be applied to *any* process whose second-order statistics exist. While, in this paper, we have assumed that these second-order statistics are known (i.e., P_f is known), this is not a prerequisite: the internal MAR-wavelet dynamics can be efficiently estimated directly from data.

In this way, our approach to modeling represents a new view of the multiscale stochastic realization problem. Rather than selecting internal matrices which are finely tuned according to the particular signal being modeled (as previous approaches do), we select them from a library of linear functionals. Our approach is not only computationally fast, but it provides a systematic way to build MAR states that have meaningful and useful structure and are not tied directly to the intricate details of the process being modeled.

The idea of selecting internal matrices from a library of linear functionals is an extremely important one in many applications. For example, if one is interested in estimating nonlocal variables from data collected at multiple resolutions, the MAR states must be able to represent the coarser variables in addition to playing a conditional decorrelation role. This corresponds to selecting internal matrices from a library of linear functionals which include ones that not only do a good conditional decorrelation job but also include ones which represent nonlocal variables to estimate or to include as measurements. The particular library consisting of wavelet bases represents a natural candidate for this role and might be useful for data fusion problems in which data are available at several resolutions.

While we have shown how to build MAR processes for modeling one-dimensional signals, our approach generalizes to higher dimensions. In two dimensions, there are a number of important and challenging problems associated with the modeling and estimation of random fields. We are currently working on extending the ideas in this paper to applications in image processing.

In addition to extending our work to two-dimensional problems, there are a number of other interesting and challenging problems to consider for future research. First, having states which contain wavelet coefficients suggests ways of doing data-adaptive multiresolution estimation where the resolution of the estimate is governed by the quality of the data. Since the MAR estimator provides estimates of $x_j(n)$ for *all* (j,n), one can use the estimates of the coarser-scale scaling coefficients of an MAR-wavelet model to obtain a coarser-scale signal estimate. Generalizing this idea, one can consider estimating the signal at a fine scale in areas where the data are of relatively high quantity and quality and at a coarser scale where the data are more sparse and of poor quality.

Another challenging problem is to build MAR-wavelet models from incomplete covariance information (i.e., from just parts of P_f). Solutions to problems of this type have been elusive because, before our work, the linear functionals which define the states were based on complete knowledge of P_f and the relaxation of this has proven difficult. However, since we select linear functionals from a library, we have no need for knowledge of P_f until we want to determine the MAR-wavelet dynamics. However, at a given node (j,n), the MAR parameters $A_j(n)$ and $Q_j(n)$ do not rely on all of P_f since they are a function just of $P_{x_j(n)}$ and $P_{x_j(n)x_{j-1}([n/2])}$. These are small matrices and, therefore, the dynamics which rely on unavailable parts of P_f can be efficiently estimated from data (if available) or chosen to be consistent with the pieces of P_f which are known.

Finally, there are a number of open questions associated with the work presented in this paper. One is how to choose the right wavelet which would provide the most accurate model for a particular class of processes. While there are many ways to address this issue, the one which is of most interest to us is to use the lifting scheme concept [10], [18] to perform multiscale decompositions which are adapted to the statistics of the process under study. Another interesting question is: how does modeling performance translate to performance in a particular estimation problem? These are questions which will motivate our future work on MAR-wavelet models.

APPENDIX A
PROOF OF (17)

To see that (16) implies that each state depends only on its parent, consider two states $x_j(n)$ and $x_j(n+1)$ at scale j, for some even integer $n \in \{0,\cdots,2^j-2\}$. The parent of these two states is

$$x_{j-1}(n/2) = \Bigg[a_j(n/2-\tilde{R}+1),\cdots,a_j(n/2+\tilde{R}-1), d_j(n/2-\frac{\tilde{R}+R}{2}+1),\cdots,d_j(n/2+\frac{\tilde{R}+R}{2}-1)\Bigg]^T.$$

Then, for every integer $i \in \{-\tilde{R}+1,\cdots,\tilde{R}-1\}$, we have

$$a_j(n+i) = \sum_{\frac{n}{2}+\lceil\frac{i-\tilde{R}}{2}\rceil}^{\frac{n}{2}+[\frac{i+\tilde{R}-1}{2}]} \tilde{h}(n+i-2p)a_{j-1}(p) + \sum_{\frac{n}{2}+\lceil\frac{i-R}{2}\rceil}^{\frac{n}{2}+[\frac{i+R-1}{2}]} \tilde{g}(n+i-2p)d_{j-1}(p) \quad (25)$$

and

$$a_j(n+1+i) = \sum_{\frac{n}{2}+\lceil\frac{i+1-\tilde{R}}{2}\rceil}^{\frac{n}{2}+[\frac{i+\tilde{R}}{2}]} \tilde{h}(n+i+1-2p)a_{j-1}(p) + \sum_{\frac{n}{2}+\lceil\frac{i+1-R}{2}\rceil}^{\frac{n}{2}+[\frac{i+R}{2}]} \tilde{g}(n+i+1-2p)d_{j-1}(p). \quad (26)$$

In order to check that every $a_{j-1}(p)$ and $d_{j-1}(p)$ in (25) and (26) is carried by $x_{j-1}(n/2)$, one can easily check that

$$\left[\left\lceil\frac{i-\tilde{R}}{2}\right\rceil,\left[\frac{i+\tilde{R}}{2}\right]\right] \subseteq [-\tilde{R}+1,\tilde{R}-1] \quad \forall i \in \{-\tilde{R}+1,\cdots,\tilde{R}-1\}$$

and that

$$\left[\left\lceil\frac{i-R}{2}\right\rceil,\left[\frac{i+R}{2}\right]\right] \subset \left[-\frac{\tilde{R}+R}{2}+1,\frac{\tilde{R}+R}{2}-1\right] \quad \forall i \in \{-\tilde{R}+1,\cdots,\tilde{R}-1\}.$$

Then, using (25) and (26), we get (17) where, for every $j \in \{1,\cdots,J-1\}$, the matrices $A_j(n)$ are $(3\tilde{R}+R-$

$2) \times (3\tilde{R}+R-2)$ and follow from (25) and (26) as: for $l \in \{1,\cdots,2\tilde{R}-1\}$, for $p_a \in \{\lceil \frac{n+l-2\tilde{R}}{2} \rceil, \cdots, [\frac{n+l-1}{2}]\}$ and for $p_d \in \{\lceil \frac{n+l-\tilde{R}-R}{2} \rceil, \cdots, [\frac{n+l+R-\tilde{R}-1}{2}]\}$

$$A_j(n)(l, p_a - [n/2] + \tilde{R}) = \tilde{h}(n+l-\tilde{R}-2p_a)$$
$$A_j(n)\left(l, p_d - [n/2] + \frac{5\tilde{R}+1}{2} - 1\right) = \tilde{g}(n+l-\tilde{R}-2p_d).$$

When $j = J$, $A_j(n)$ are vectors of length $3\tilde{R}+R-2$ and follow as: for $p_a \in \{\lceil \frac{n+1-2\tilde{R}}{2} \rceil, \cdots, [\frac{n}{2}]\}$ and for $p_d \in \{\lceil \frac{n+1-\tilde{R}-R}{2} \rceil, \cdots, [\frac{n+R-\tilde{R}}{2}]\}$

$$A_J(n)(1, p_a - [n/2] + \tilde{R}) = \tilde{h}(n+1-\tilde{R}-2p_a).$$
$$A_J(n)\left(1, p_d - [n/2] + \frac{5\tilde{R}+1}{2} - 1\right) = \tilde{g}(n+1-\tilde{R}-2p_d).$$

APPENDIX B
PROOF OF PROPOSITION 1

In this Appendix we provide a proof of Proposition 1. The proof of Proposition 1 requires the following lemma. Due to the lack of space, the proof of this lemma is omitted but can be found in [16].

Lemma 1: Let i be an integer in $\{1,\cdots,\frac{\tilde{R}+R}{2}-1\}$, then

$$\sum_{p=0}^{i-1} \tilde{h}(\tilde{R}-2p) a_j\left(n - \frac{\tilde{R}-R}{2} - i + p\right) + \tilde{g}(R-2p) \times d_j(n-i+p) = \sum_{k=-\tilde{R}+1}^{\tilde{R}-2} \alpha_i(k) a_{j+1}(2n+k) \quad (27)$$

and

$$\sum_{p=0}^{i-1} \tilde{h}(-\tilde{R}+1+2p) a_j\left(n + \frac{\tilde{R}-R}{2} + i - p\right) + \tilde{g}(-R+1+2p) d_j(n+i-p) = \sum_{k=-\tilde{R}+3}^{\tilde{R}} \beta_i(k) a_{j+1}(2n+k) \quad (28)$$

where we have (29), as shown at the bottom of the page, and (30), also shown at the bottom of the page.

We now prove Proposition 1.

Proof: Define the matrices K_1 and K_2 as shown in (31) at the bottom of this page, and the $(\frac{\tilde{R}+R}{2}-1)\times(\frac{\tilde{R}+R}{2}-1)$ triangular matrices $H_1, H_2, G_1,$ and G_2, as also shown at the bottom of this page, then (27) and (28) imply that

$$\alpha_i(k) = \begin{cases} \sum_{p=0}^{i-1} \tilde{h}(\tilde{R}-2p) h(k+2i-2p+\tilde{R}-R) + \tilde{g}(R-2p) g(k+2i-2p), & \text{if } k \le 2R-\tilde{R}-2 \\ \sum_{p=0}^{i-1} \tilde{g}(R-2p) g(k+2i-2p), & \text{if } k > 2R-\tilde{R}-2 \end{cases} \quad (29)$$

$$\beta_i(k) = \begin{cases} \sum_{p=0}^{i-1} \tilde{h}(-\tilde{R}+1+2p) h(k+2p-2i+R-\tilde{R}) + \tilde{g}(-R+1+2p) g(k+2p-2i), & \text{if } k \ge \tilde{R}-2R+3 \\ \sum_{p=0}^{i-1} \tilde{g}(-R+1+2p) g(k+2p-2i), & \text{if } k < \tilde{R}-2R+3 \end{cases} \quad (30)$$

$$K_1 = \begin{pmatrix} \alpha_{\frac{\tilde{R}+R}{2}-1}(-\tilde{R}+1) & \alpha_{\frac{\tilde{R}+R}{2}-1}(-\tilde{R}+2) & \cdots & \alpha_{\frac{\tilde{R}+R}{2}-1}(\tilde{R}-2) \\ \alpha_{\frac{\tilde{R}+R}{2}-2}(-\tilde{R}+1) & \alpha_{\frac{\tilde{R}+R}{2}-2}(-\tilde{R}+2) & \cdots & \alpha_{\frac{\tilde{R}+R}{2}-2}(\tilde{R}-2) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1(-\tilde{R}+1) & \alpha_1(-\tilde{R}+2) & \cdots & \alpha_1(\tilde{R}-2) \end{pmatrix}$$
$$K_2 = \begin{pmatrix} \beta_1(-\tilde{R}+3) & \beta_1(-\tilde{R}+4) & \cdots & \beta_1(\tilde{R}) \\ \beta_2(-\tilde{R}+3) & \beta_2(-\tilde{R}+4) & \cdots & \beta_2(\tilde{R}) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{\frac{\tilde{R}+R}{2}-1}(-\tilde{R}+3) & \beta_{\frac{\tilde{R}+R}{2}-1}(-\tilde{R}+4) & \cdots & \beta_{\frac{\tilde{R}+R}{2}-1}(\tilde{R}) \end{pmatrix} \quad (31)$$

$$H_1(l,c) = \begin{cases} \tilde{h}(\tilde{R}-2(c-l)), & \text{for } l = 1,\cdots,\frac{\tilde{R}+R}{2}-1 \text{ and } c = l,\cdots,\frac{\tilde{R}+R}{2}-1 \\ 0, & \text{otherwise} \end{cases}$$
$$G_1(l,c) = \begin{cases} \tilde{g}(R-2(c-l)), & \text{for } l = 1,\cdots,\frac{\tilde{R}+R}{2}-1 \text{ and } c = l,\cdots,\frac{\tilde{R}+R}{2}-1 \\ 0, & \text{otherwise} \end{cases}$$
$$H_2(l,c) = \begin{cases} \tilde{h}(-\tilde{R}+1+2(l-c)), & \text{for } l = 1,\cdots,\frac{\tilde{R}+R}{2}-1 \text{ and } c = 1,\cdots,l \\ 0, & \text{otherwise} \end{cases}$$
$$G_2(l,c) = \begin{cases} \tilde{g}(-R+1+2(l-c)), & \text{for } l = 1,\cdots,\frac{\tilde{R}+R}{2}-1 \text{ and } c = 1,\cdots,l \\ 0, & \text{otherwise} \end{cases}$$

$$G_1\begin{pmatrix} d_j(n-\frac{\tilde{R}+R}{2}+1) \\ \vdots \\ d_j(n-1) \end{pmatrix} + H_1\begin{pmatrix} a_j(n-\tilde{R}+1) \\ \vdots \\ a_j(n-\frac{\tilde{R}-R}{2}-1) \end{pmatrix} = K_1\begin{pmatrix} a_{j+1}(2n-\tilde{R}+1) \\ \vdots \\ a_{j+1}(2n+\tilde{R}-2) \end{pmatrix}$$

$$G_2\begin{pmatrix} d_j(n+1) \\ \vdots \\ d_j(n+\frac{\tilde{R}+R}{2}-1) \end{pmatrix} + H_2\begin{pmatrix} a_j(n+\frac{\tilde{R}-R}{2}+1) \\ \vdots \\ a_j(n+\tilde{R}-1) \end{pmatrix} = K_2\begin{pmatrix} a_{j+1}(2n-\tilde{R}+3) \\ \vdots \\ a_{j+1}(2n+\tilde{R}) \end{pmatrix}.$$

Since $\tilde{g}(R) \neq 0$ and $\tilde{g}(-R+1) \neq 0$, then G_1 and G_2 are invertible and (19) and (20) follow with $I_1 = -G_1^{-1}H_1$, $J_1 = G_1^{-1}K_1$, $I_2 = -G_2^{-1}H_2$, and $J_2 = G_2^{-1}K_2$. □

APPENDIX C
PROOF OF PROPOSITION 2

In this Appendix we provide a proof of Proposition 2.

Proof: First, let us define for notational simplicity $\hat{R} \triangleq \frac{\tilde{R}+R}{2}$. Now consider the children $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$ of $x_j(n)$. Then we have the following.

- $[a_j(n-\tilde{R}+1),\cdots,a_j(n-\frac{\tilde{R}-R}{2}-1)]$ (respectively, $[a_j(n+\frac{\tilde{R}-R}{2}+1),\cdots,a_j(n+\tilde{R}-1)]$) is obviously a linear function of $x_{j+1}(2n)$ and (respectively, $x_{j+1}(2n+1)$) since it is simply copied in $U_{j+1}(2n)$ (respectively, $U_{j+1}(2n+1)$).
- Using Proposition 1, $[d_j(n-\hat{R}+1),\cdots,d_j(n-1)]$ and $[d_j(n+1),\cdots,d_j(n+\hat{R}-1)]$ are a linear function of $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$, respectively.
- The wavelet decomposition formulas (10) imply that $d_j(n)$ is a linear function of $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$ since they contain $a_{j+1}(m)$ for $m \in \{2n-\tilde{R}+1,\cdots,2n+\tilde{R}\}$.
- The wavelet decomposition formulas (10) also imply that $[a_j(n-\frac{\tilde{R}-R}{2}),\cdots,a_j(n+\frac{\tilde{R}-R}{2})]$ is a linear function of $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$. Indeed, for $i \in \mathcal{I} \triangleq \{-\frac{\tilde{R}-R}{2},\cdots,\frac{\tilde{R}-R}{2}\}$ we have

$$a_j(n+i) = \sum_{p=2n+2i-R+1}^{2n+2i+R} h(p-2n-2i)a_{j+1}(p).$$

Since

$$\{2n+2i-R+1,\cdots,2n+2i+R\}_{i\in\mathcal{I}} = \{2n-\tilde{R}+1,\cdots,2n+\tilde{R}\}$$

it follows that $[a_j(n-\frac{\tilde{R}-R}{2}),\cdots,a_j(n+\frac{\tilde{R}-R}{2})]$ is a linear function of $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$.

- Finally, $U_j(n)$ is a linear function of $x_{j+1}(2n)$ and $x_{j+1}(2n+1)$ since the two parts $U_j^1(n)$ and $U_j^2(n)$ that compose $U_j(n)$ are carried by $U_{j+1}(2n)$ and $U_{j+1}(2n+1)$, respectively. □

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their useful and insightful comments which helped to vastly improve the presentation of the paper.

REFERENCES

[1] P. Abry and F. Sellan, "The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation," *Appl. Computational Harmonic Anal.*, vol. 3, pp. 377–383, 1996.
[2] H. Akaike, "Stochastic theory of minimal realizations," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 667–674, Dec. 1974.
[3] ———, "Markovian representation of stochastic processes by canonical variables," *SIAM J. Contr.*, vol. 13, no. 1, pp. 162–173, Jan. 1975.
[4] K. S. Arun and S. Y. Kung, "Balanced approximation of stochastic systems," *SIAM J. Matrix Anal. Appl.*, vol. 11, no. 1, pp. 42–68, Jan. 1990.
[5] 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.
[6] H. Cheng and C. A. Bouman, "Trainable context model for multiscale segmentation," in *Proc. IEEE Int. Conf. Image Processing*, Chicago, IL, Oct. 1998.
[7] 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.
[8] K. C. Chou, A. S. Willsky, and R. Nikoukhah, "Multiscale systems, Kalman filters, and Riccati equations," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 479–492, Mar. 1994.
[9] K. Chou, "A stochastic modeling approach to multiscale signal processing," Ph.D. dissertation, Massachusetts Institute of Technology, May 1991.
[10] R. L. Claypoole, R. G. Baraniuk, and R. D. Nowak, "Adaptive wavelet transforms via lifting," in *Trans. Int. Conf. Acoustics, Speech and Signal Processing*, Seattle, WA, May 1998, pp. 1513–1516.
[11] A. Cohen, I. Daubechies, and P. Vial, "Wavelet bases on the interval and fast algorithms," *J. Appl. Computational Harmonic Anal.*, vol. 1, pp. 54–81, 1993.
[12] M. S. Crouse and R. G. Baraniuk, "Simplified wavelet-domain hidden Markov models using contexts," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Seattle, WA, May 1998, pp. 2277–2280.
[13] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
[14] M. Daniel and A. S. Willsky, "Modeling and estimation of fractional Brownian motion using multiresolution stochastic processes," in *Fractals in Engineering*, J. L. Vehel, E. Lutton, and C. Tricot, Eds. New York: Springer, 1997, pp. 124–137.
[15] ———, "A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing," *Proc. IEEE*, vol. 85, pp. 164–180, Jan. 1997.
[16] K. Daoudi, A. B. Frakt, and A. S. Willsky, "Multiscale autoregressive models and wavelets: Extended version," Tech. Rep. LIDS-P-2437, Massachusetts Institute of Technology, Nov. 1998.
[17] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.
[18] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Analysis Appl.*, vol. 4, no. 3, pp. 245–267, 1998.
[19] U. B. Desai and D. Pal, "A realization approach to stochastic model reduction and balanced stochastic realizations," in *IEEE Conf. Decision and Control*, 1982, pp. 1105–1112.
[20] E. Fabre, "New fast smoothers for multiscale systems," *IEEE Trans. Signal Processing*, vol. 44, no. 8, pp. 1893–1911, Aug. 1996.
[21] P. W. Fieguth and A. S. Willsky, "Fractal estimation using models on multiscale trees," *IEEE Trans. Signal Processing*, vol. 44, pp. 1297–1300, May 1996.
[22] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–917, Mar. 1992.
[23] A. B. Frakt, "A realization theory for multiscale autoregressive stochastic models," Ph.D. dissertation proposal, Nov. 1997.
[24] A. B. Frakt and A. S. Willsky, *Computationally Efficient Stochastic Realization for Internal Multiscale Autoregressive Models*, to be published.
[25] ———, "Efficient multiscale stochastic realization," in *Int. Conf. Acoustics, Speech, and Signal Processing*, Seattle, WA, May 1998.

[26] W. W. Irving, W. C. Karl, and A. S. Willsky, "A theory for multiscale stochastic realization," in *Proc. 33rd IEEE Conf. Decision and Control*, Lake Buena Vista, FL, Dec. 1994, vol. 1, pp. 655–662.
[27] W. W. Irving, "Multiscale stochastic realization and model identification with applications to large-scale estimation problems," Ph.D. dissertation, Massachusetts Institute of Technology, Sept. 1995.
[28] W. W. Irving and A. S. Willsky, "A canonical correlations approach to multiscale stochastic realization," *IEEE Trans. Automat. Contr.*, to be published.
[29] H. Krim and J.-C. Pesquet, "Multiresolution analysis of a class of nonstationary processes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1010–1020, July 1995.
[30] A. Lindquist and G. Picci, "On the stochastic realization problem," *Siam J. Contr. Optim.*, vol. 17, no. 3, pp. 365–389, May 1979.
[31] M. R. Luettgen and A. S. Willsky, "Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination," *IEEE Trans. Image Processing*, vol. 4, no. 2, pp. 194–207, Feb. 1995.
[32] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.
[33] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, vol. 10, pp. 422–437, 1968.
[34] Y. Meyer, *Ondelettes et Opérateurs, I: Ondeletts, II: Opérateurs de Calderón-Zygmund, III: Opérateurs Multilinéaires* (with R. Coifman). Paris, Hermann, 1993.
[35] F. Sellan, "Synthèse de mouvements Browniens fractionnaires à l'aide de la transformation par ondelettes," *C.R.A.S. Paris Sér. I Math*, vol. 321, pp. 351–358, 1995.
[36] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.
[37] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 904–909, Mar. 1992.
[38] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1968.
[39] J. Zhang and G. Walter, "A wavelet-based KL-like expansion for wide-sense stationary random processes," *IEEE Trans. Signal Processing*, vol. 42, pp. 1737–1745, July 1994.