*Research Article*

# Optimal Modeling and Filtering of Stochastic Time Series for Geoscience Applications

## J. A. Rod Blais

*Department of Geomatics Engineering, Pacific Institute for the Mathematical Sciences, University of Calgary, Calgary, AB, Canada T2N 1N4*

Correspondence should be addressed to J. A. Rod Blais; blais@ucalgary.ca

Sequences of observations or measurements are often modeled as realizations of stochastic processes with some stationary properties in the first and second moments. However in practice, the noise biases and variances are likely to be different for different epochs in time or regions in space, and hence such stationarity assumptions are often questionable. In the case of strict stationarity with equally spaced data, the Wiener-Hopf equations can readily be solved with fast Fourier transforms (FFTs) with optimal computational efficiency. In more general contexts, covariance matrices can also be diagonalized using the Karhunen-Loève transforms (KLTs), or more generally using empirical orthogonal and biorthogonal expansions, which are unfortunately much more demanding in terms of computational efforts. In cases with increment stationarity, the mathematical modeling can be modified and generalized covariances can be used with some computational advantages. The general nonlinear solution methodology is also briefly overviewed with the practical limitations. These different formulations are discussed with special emphasis on the spectral properties of covariance matrices and illustrated with some numerical examples. General recommendations are included for practical geoscience applications.

## 1. Introduction

Consider a stochastic process or time series $\{X(t), -\infty < t < \infty\}$ which has been observed or measured as $\{y(s), -\infty < s < \infty\}$, most often in practice at equispaced discrete time or space intervals $\Delta s = 1$ for simplicity over some finite domain. In practice, it is also often convenient to assume these processes or time series to be centered, that is, with null mean value or expectation following the appropriate trend modeling in increment stationary situations. When these observations (or measurements) are of Gaussian processes, their second moments or covariances then fully specify the processes. Otherwise, it can only be assumed that the covariance functions are sufficient for the following modeling and filtering operations. Specifically, it is assumed that these covariance functions have positive Fourier transforms or power spectra for realizability.

In general, the optimal estimate of $X(t)$ from observations $y(s)$ is given by the conditional expectation

$$\widehat{X}(t) = \mathbf{E}\left[X(t) \mid y(s)\right] \tag{1}$$

in which the conditional expectation of a random variable $U$ given an observation $v$ is defined by

$$\mathbf{E}[U \mid v] = \int_{-\infty}^{\infty} u\, p(u \mid v)\, du \tag{2}$$

using the conditional density $p(u \mid v)$; that is,

$$p(u \mid v) = \frac{p(u, v)}{p_U(u)} \tag{3}$$

for the joint density $p(u, v)$ with positive marginal density $p_U(u)$; that is,

$$p_U(u) = \int_{-\infty}^{\infty} p(u, v)\, dv. \tag{4}$$

These quantities can readily be generalized to several variables. The previous expression is also called the conditional mean with corresponding conditional variance

$$\mathrm{Var}\,[U \mid v] = \mathbf{E}\left[(U \mid v - \mathbf{E}\,[U \mid v])^2\right]$$
$$\equiv \mathbf{E}\left[(U \mid v - \mathbf{E}\,[U \mid v])(U \mid v - \mathbf{E}[U \mid v])^*\right] \tag{5}$$

which is commonly used in optimization, with the $^*$ corresponding to conjugate transpose, or simply transpose in real cases. For the random variable $V$ corresponding to the observations, $\mathbf{E}[U \mid V]$ and $\mathrm{Var}[U \mid V]$ are then the conditional expectation for the mean and variance of $U$ given $V$.

Such conditional expectation $\mathbf{E}[X(t) \mid Y(s)]$ can easily be seen as an optimal estimate $\widehat{X}(t)$ as

$$\mathbf{E}\left[\widehat{X}(t)\right] = \mathbf{E}\left[\mathbf{E}\left[X(t) \mid Y(s)\right]\right] = \mathbf{E}\left[X(t)\right] \tag{6}$$

and its variance

$$\mathrm{Var}\left[\widehat{X}(t)\right] = \mathrm{Var}\left[\mathbf{E}\left[X(t) \mid Y(s)\right]\right]$$
$$= \mathrm{Var}\left[X(t)\right] - \mathbf{E}\left[\mathrm{Var}\left[X(t) \mid Y(s)\right]\right] \tag{7}$$
$$\leq \mathrm{Var}\left[X(t)\right]$$

as expected. For any other estimator $Z(t)$, that is, any function of $Y(t)$, it is straightforward to verify that

$$\mathrm{MSE}\left[X(t) - \mathbf{E}\left[X(t) \mid Y(s)\right]\right] \leq \mathrm{MSE}\left[X(t) - Z(t)\right] \tag{8}$$

for the mean-square errors (MSEs); that is, the conditional expectation estimator $\mathbf{E}[X(t) \mid Y(s)]$ is optimal in the mean-square sense. For any unbiased estimator $Z(t)$, the relationship is in the least-squares sense with variances replacing the mean-square errors. More discussion of these relationships can be found, for example, in [1].

In the linear context, the conditional expectation $\widehat{X}(t) = \mathbf{E}[X(t) \mid y(s)]$ is a linear function of the observations $y(s)$; that is,

$$\widehat{X}(t) = \int_{-\infty}^{\infty} H(t, s)\, y(s)\, ds \tag{9}$$

for some unknown kernel $H(t, s)$, or with discrete data,

$$\widehat{X}(t) = \sum_{s=-\infty}^{\infty} H(t, s)\, y(s), \tag{10}$$

that is, simply an unknown linear combination of the available data. In practice, only finite data are available so that the previous limits are finite, and furthermore in causal cases, only past observations are available for processing. Furthermore to simplify the presentation in the following, the inner product notation will be used; that is,

$$\widehat{X}(t) = \left\langle H(t, s), y^*(s) \right\rangle \tag{11}$$

for the continuous and discrete applications whenever appropriate. The superscript $^*$ stands for conjugate transpose although only real applications are considered herein.

The objective in the following is to estimate optimally $X(t)$ from the available data $y(s)$ using minimal assumptions. Using the above discussed property of the conditional expectation estimator $\mathbf{E}[X(t) \mid y(s)]$, one can write the implied orthogonality of the residuals to the observations; that is,

$$\mathbf{E}\left[\left(X(t) - \widehat{X}(t)\right) y^*(s)\right] = \mathbf{E}\left[\left\langle \left(X(t) - \widehat{X}(t)\right), y(s) \right\rangle\right] = 0; \tag{12}$$

that is, using the linear transformation of $y(t)$ for $\widehat{X}(t)$,

$$\mathbf{E}\left[\left\langle \left(X(t) - \left\langle H(t, r), y^*(r) \right\rangle\right), y(s) \right\rangle\right] = 0 \tag{13}$$

or

$$\mathbf{E}\left[\left\langle X(t), y(s) \right\rangle\right] = \mathbf{E}\left[\left\langle \left\langle H(t, r), y^*(r) \right\rangle, y(s) \right\rangle\right]$$
$$= \mathbf{E}\left[\left\langle H(t, r), \left\langle y(r), y(s) \right\rangle \right\rangle\right] \tag{14}$$
$$= \left\langle H(t, r), \mathbf{E}\left[\left\langle y(r), y(s) \right\rangle\right] \right\rangle$$

which is usually interpreted in terms of cross-covariance between $X(t)$ and $y(s)$ being equal to the autocovariance of $y(r)$ modified by the unknown kernel $H(t, r)$, assuming $\mathbf{E}[y(r)] = 0$, or equivalently,

$$\left\langle K_{yy}(s, r), H(r, t) \right\rangle = K_{yX}(s, t) \tag{15}$$

using the cross-covariance $K_{yX}(s, t)$ and (symmetric real positive definite) covariance $K_{yy}(s, r)$, often written as $K_y(s, r)$ in practice, respectively. These normal equations are the well-known Wiener-Hopf equations, which are Fredholm integral equations (of the first kind) in the continuous context. In cases of stationarity, this unknown kernel $H(r, t)$ becomes $H(r - t)$ which will be seen to greatly simplify the situation in practice in terms of computations.

It should be mentioned that the previous interchange of the order in the linear expectation operation in the inner products needs to be done with care in the case of continuous applications with infinite limits. However for discrete applications with finite limits, the situation is much simpler and usually carried out without second thoughts. More discussion of those aspects can be found in, for example, the works of Dougherty [1] and Pugachev [2] with references in functional analysis.

The general nonlinear solution methodology will also briefly be overviewed in the Gaussian context with some of the practical limitations. These different formulations are discussed with emphasis on the properties of the covariance matrices and illustrated with some numerical examples and references. General recommendations are included for practical geoscience applications.

## 2. Fourier Transform Approaches

The Wiener-Hopf equations written as

$$\left\langle K_{yy}(s, r), H(r, t) \right\rangle = K_{yX}(s, t) \tag{16}$$

for unknown $H(r, t)$ are trivial for white noise, that is, with $K_{yy}(s, r) \equiv \delta(s, r) \equiv \delta(s - r)$, the Dirac delta function, implying that $H(r, t) \equiv H(r - t) = K_{yX}(r, t) \equiv K_{yX}(r - t)$ in such cases. However, when the white noise has different intensities in different parts of the domain, this simple formulation cannot be used and the subsequent sections will deal with the more general situation.

When the covariance kernel $K_{yy}(s, r)$ can be written as $K_{yy}(s - r)$ meaning that the covariance is only dependent on the spacing between its arguments, then it can be diagonalized using a Fourier transform. The corresponding covariance matrix is Toeplitz; that is, its diagonals are constant in terms of $s - r$ for row index $s$ and column index $r$ with gridded data. Notice that the discrete Fourier transform (DFT) of a finite Toeplitz matrix is not necessarily diagonal unless the matrix is circulant as will be illustrated in Section 6.

Explicitly using a Fourier matrix $F \equiv [f_{rs} \mid f_{rs} = e^{2\pi i rs/N}]$ of order $N$, the DFT of a vector $v$ is simply $\mathcal{F}[v] \equiv v^\wedge \equiv F^* v$, and hence in the case of stationarity,

$$F^* K_{yy}(s - r) F F^* H(r) = N F^* K_{yX}(r) \qquad (17)$$

or

$$K_{yy}^\wedge(\tau) H^\wedge(\tau) = N K_{yX}^\wedge(\tau) \qquad (18)$$

for frequencies $\tau$; that is,

$$H^\wedge(\tau) = N \cdot \frac{K_{yX}^\wedge(\tau)}{K_{yy}^\wedge(\tau)} \qquad (19)$$

obviously for $K_{yy}^\wedge(\tau) \neq 0$ assuming the symmetry and positive definiteness of $K_{yy}$, in which $F \cdot F^* = F^* \cdot F = N$. This means that the problem reduces to estimating frequency components of the $y(s)$ observations to infer the corresponding frequency components of the $X(t)$ process. Furthermore, in general, the transfer kernel or function $H^\wedge(\tau)$ is noncausal, as for each time $t$, it uses all of the information from $-\infty < t < \infty$ in the Fourier transform. Hence in practical cases with only finite or even only past information, complications can be expected (see, e.g., the discussion in [3]).

## 3. Empirical Orthogonal Function Expansions

The covariance kernel $K_{yy}(s, r)$ for the observations is generally assumed to be symmetric and positive definite for physical realizability. Its eigenfunctions are the exponential functions only in the stationary case as $F^* K_{yy}(s, r) F$ is then diagonal and $F^* K_{yy}(s, r) F = \text{diag}(\lambda_\nu)$ or $K_{yy}(s, r) F = N^{-1} F \text{diag}(\lambda_\nu)$ for the corresponding eigenvalues $\lambda_\nu$. The discrete Fourier transform application or orthogonal projection onto the exponential eigenfunctions can be generalized to other systems of orthogonal functions in the sense of Galerkin's orthogonal projection methodology for such functional operator systems. This will be illustrated in Section 6 using Haar wavelet transforms.

Hence for a spectrum of orthogonal eigenfunctions $u_k(r)$ and corresponding eigenvalues $\lambda_k$, usually ordered as $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots > 0$,

$$K_{yy}(s, r) = \sum_{k=1}^\infty \lambda_k u_k(s) u_k^*(r) \qquad (20)$$

can be used for an expansion of the random functions. The eigenfunction expansions of random functions or Karhunen-Loève transformed random variables then have a diagonal covariance kernel or matrix. Notice that, in finite dimensions, the expansion of the covariance matrix is simply the eigenvalue expansion which is a special application of the Singular Value Decomposition (SVD) (see, e.g., the works of Klema and Laub [4], Lewis [5], and Parlett [6]). The eigenfunction expansion of the covariance kernel is often referred to as Mercer's theorem in the mathematical literature. The eigenfunctions derived from the data covariance are also called *empirical orthogonal functions* (EOFs) and have proven most effective in signal data processing and in modeling dynamical systems such as in oceanography and related fields (see, e.g., Kim and Wu [7], North and Cahalan [8]).

The mathematical situation is summarized by the well-known *Karhunen-Loève theorem*: Let $X(t)$ be a zero-mean random function on $T$ with a weight function $w(t) \geq 0$ such that its covariance kernel satisfies

$$\int_T \int_T |K_X(s, r)|^2 w(r) \, ds \, dr < \infty; \qquad (21)$$

then

(i) for the integral equation

$$\int_T K_X(s, r) u(r) w(r) \, dr = \lambda u(s), \qquad (22)$$

there exist eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots > 0$ with corresponding eigenfunctions $u_1(s), u_2(s), u_3(s), \ldots$ such that

$$\int_T K_X(s, r) u_k(r) w(r) \, dr = \lambda_k u_k(s) \qquad (23)$$

for all $k$ such that $\{u_k(t)\}$ is a deterministic orthonormal system on $T$ relative to the weight function $w(t)$ meaning that

$$\int_T u_k(t) u_h^*(t) w(t) \, dt = \delta_{kh}; \qquad (24)$$

(ii) the corresponding generalized Fourier coefficients of $X(t)$ relative to these eigenfunctions

$$\xi_k = \int_T X(t) u_k^*(t) w(t) \, dt \qquad (25)$$

are uncorrelated and $\text{Var}[\xi_k] = \lambda_k$;

(iii) $X(t)$ can then be represented as an eigenfunction expansion as

$$X(t) = \sum_{k=1}^\infty \xi_k u_k(t) \qquad (26)$$

in the mean-square sense.

This last orthogonal expansion of the random function $X(t)$ is known in the literature as a generalized Fourier expansion, a Karhunen-Loève expansion, or a canonical expansion. Also, the Fourier and Karhunen-Loève expansions are usually introduced in the functional analysis context of square-integrable functions such as $L^2(\mathbb{R})$ while the canonical expansion context is simply a decomposition of the covariance matrix or kernel (see, e.g., the works of Dougherty [1] for more discussion).

In practice when only the observational sequence $\{y(s)\}$ is known, the SVD can be used to obtain the eigenvectors of the corresponding covariance function without explicitly evaluating the covariance matrix. As different data matrices are generally rectangular in shape, different strategies are advisable for the SVD computations (see, e.g., the works of Björck [9] and Blais [10]). Furthermore, especially in multiple dimensions, the SVD and Principal Component Analysis (PCA) terminologies are not always consistent in the applied science and engineering literature largely because of related data array conventions and objectives.

## 4. Empirical Biorthogonal Expansions

Rather than solving the eigenvalue equation for the orthonormal eigenfunctions $\{u_k(t)\}$ with corresponding eigenvalues $\{\lambda_k\}$, let us assume that an arbitrary sequence $\{a_k(t)\}$ of square-integrable functions are available. Then a biorthogonal sequence $\{v_k(t)\}$ can be generated to satisfy the conditions:

(i) $x_k(t) = \langle K_X(t,s), v_k(s) \rangle / \langle\langle K_X(t,s), v_k(t)\rangle, v_k(s)\rangle$,

(ii) $\langle x_k(t), v_h(t) \rangle = \delta_{kh}$

recursively in a well-known manner using the following projections:

(i) set $v_1(t) = a_1(t)$ with $x_1(t) = \langle K_X(t,s), v_1(s)\rangle / \langle\langle K_X(t,s), v_1(t)\rangle, v_1(s)\rangle$,

(ii) define $v_2(t) = c_{21}v_1(t) + a_2(t)$ with $c_{21}$ determined from $\langle x_1(t), v_2(t)\rangle = 0$; that is,

$$\langle x_1(t), c_{21}v_1(t) + a_2(t)\rangle = c_{21}\langle x_1(t), v_1(t)\rangle + \langle x_1(t), a_2(t)\rangle$$
$$= c_{21} + \langle x_1(t), a_2(t)\rangle = 0 \tag{27}$$

implying

$$c_{21} = -\langle x_1(t), a_2(t)\rangle \text{ and hence}$$
$$v_2(t) = -\langle x_1(t), a_2(t)\rangle v_1(t) + a_2(t) \tag{28}$$

with

$$x_2(t) = \frac{\langle K_X(t,s), v_2(s)\rangle}{\langle\langle K_X(t,s), v_2(t)\rangle, v_2(s)\rangle}, \tag{29}$$

(iii) assuming $v_1(t), v_2(t), \ldots, v_{k-1}(t)$ and $x_1(t), x_2(t), \ldots, x_{k-1}(t)$ known, define $v_k(t) = c_{k1}v_1(t) + c_{k2}v_2(t) + \cdots +$

$c_{k,k-1}v_{k-1}(t) + a_k(t)$ with $c_{k1}, c_{k2}, \ldots, c_{k,k-1}$ determined from

$$\langle x_1(t), v_k(t)\rangle = \langle x_2(t), v_k(t)\rangle = \cdots$$
$$= \langle x_{k-1}(t), v_k(t)\rangle = 0 \tag{30}$$

with

$$x_k(t) = \frac{\langle K_X(t,s), v_k(s)\rangle}{\langle\langle K_X(t,s), v_k(t)\rangle, v_k(s)\rangle} \tag{31}$$

(see, e.g., [1] for more details and applications).

Notice that when the given sequence $\{a_k(t)\}$ of square-integrable functions are orthogonal, then the generated biorthogonal sequence $\{v_k(t)\}$ is simply the original orthogonal sequence $\{a_k(t)\}$. Furthermore, if the given sequence $\{a_k(t)\}$ happened to be the eigenfunction sequence $\{u_k(t)\}$ from $K_X(t,s)$, then the generated biorthogonal sequence $\{v_k(t)\}$ would simply be the eigensequence $\{u_k(t)\}$ with normalizing coefficients being the corresponding eigenvalues $\lambda_k$ and the situation reducing to the previous orthogonal case with unit weight function $w(t) \equiv 1$.

For a random function $X(t)$ over some domain $T$ with zero-mean value and covariance kernel $K_X(t,s)$, let $\{v_k(t), t \in T\}$ denote a sequence of square-integrable functions such that $\{x_k(t)\}$ and $\{v_k(t)\}$ are biorthogonal function sequences on $T$. Then the following conditions are equivalent:

(i) $X(t)$ is equal in the mean-square sense to its canonical expansion in terms of $\xi_k$ and $x_k(t)$; that is,

$$X(t) = \sum_{k=1}^{\infty} \langle X(t), v_k(t)\rangle x_k(t) = \sum_{k=1}^{\infty} \xi_k x_k(t); \tag{32}$$

(ii) $K_X(t,s)$ has a canonical expansion in terms of $\kappa_k$, $x_k(t)$, and $x_k(s)$

$$K_X(t,s) = \sum_{k=1}^{\infty} \kappa_k x_k(t) x_k^*(s); \tag{33}$$

(iii) $\mathrm{Var}[X(t)]$ has a canonical expansion in terms of $\xi_k$ and $x_k(t)$

$$\mathrm{Var}[X(t)] = \sum_{k=1}^{\infty} \xi_k |x_k(t)|^2. \tag{34}$$

More discussion and applications of these results can be found in, for example, [1].

Again in some applications where an arbitrary orthogonal function sequence is available such as from previous measurements or observations, the biorthogonal basis function approach is generally advantageous over other approaches that do not make use of the prior information.

# 5. Nonstationary and Nonlinear Stochastic Analysis

Although most nonstationary stochastic processes are not amenable to linear analysis, there are exceptional cases where a simple variable transformation renders linear analysis possible. For instance, multiplicative noise contamination becomes additive "noise" under a logarithm transformation, and hence linear analysis can be done in the logarithmic space.

Some stochastic processes become stationary under differentiation in continuous contexts or differencing in discrete contexts. For example, a (covariance stationary) process with an unknown linear trend as first moment becomes stationary under double differentiation or differencing as commonly done in Global Positioning System (GPS) data processing. Such processes are called increment stationary and their second moments remain stationary under linear operations of differentiation or differencing. Obviously only finitely many such operations are considered corresponding to the (finite) degree of the polynomial trend.

Assuming a stochastic process $X(t)$ to have stationary increments of some order $k \geq 0$, then $\nabla^{k+1} X(t) \equiv X^{(k+1)}(t)$ is stationary with some generalized covariance $G(P,Q) \equiv G(P-Q)$. That means that the generalized Fourier transform $\mathscr{F}[G(P,Q)] > 0$ for realizability, but using the well-known Khintchine-Wiener theorem, the spectrum of $X^{(k+1)}(t)$ is also given by the generalized Fourier transform of the autocorrelations of $X^{(k+1)}(t)$, that is, $|\mathscr{F}[X^{(k+1)}(t)]|^2$. Using the mathematical relation $\mathscr{F}[X^{(k+1)}(t)] = (i\omega)^{k+1}\mathscr{F}[X(t)]$ for frequencies $\omega$, then

$$
\begin{aligned}
\mathscr{F}\left[G\left(P,Q\right)\right] &= \left|\mathscr{F}[X^{(k+1)}(t)]\right|^2 \\
&\equiv \mathscr{F}\left[X^{(k+1)}(t)\right] \cdot \mathscr{F}\left[X^{(k+1)}(t)\right]^* \\
&= \omega^{2k+2}|\mathscr{F}\left[X(t)\right]|^2 > 0
\end{aligned}
\tag{35}
$$

which implies the existence and positive definiteness (for realizability) of a generalized covariance $G(P,Q) \equiv G(P-Q)$ for $X(t)$ whenever one exists for $X^{(k+1)}(t)$.

Assuming that $X(t)$ is increment stationary, then its expected value can be modeled using an algebraic polynomial $M_X(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k$ with unknown coefficients $a_0, a_1, a_2, \ldots, a_k$, and the previous Wiener-Hopf equations for unknown $H(r,t)$

$$
\left\langle K_{yy}\left(s,r\right), H\left(r,t\right)\right\rangle = K_{yX}\left(s,t\right)
\tag{36}
$$

have to be generalized to the so-called extended normal equations,

$$
\begin{aligned}
\left\langle K_{yy}\left(s,r\right), H\left(r,t\right)\right\rangle + \alpha M_X\left(t\right) &= K_{yX}\left(s,t\right), \\
\left\langle M_X\left(r\right), H\left(r,s\right)\right\rangle &= M_X\left(s\right),
\end{aligned}
\tag{37}
$$

for some Lagrange multiplier $\alpha$ for constrained optimization.

Explicitly, using the terminology and conventions of regionalized variables in geostatistics, the linear prediction

for $f(x_0) = \sum_{i=1}^{N} \lambda_i f(x_i)$, with an algebraic polynomial trend or drift model $a_0 + a_1 x + \cdots + a_k x^k$ for the function $f(x)$, and denoting $C_{ij} = \mathbf{E}[f(x_i) \quad f^*(x_j)]$, the normal equations become

$$
\begin{bmatrix}
C_{11} & \cdots & C_{1N} & 1 & \cdots & x_1^k \\
\vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
C_{N1} & \cdots & C_{NN} & 1 & \cdots & x_N^k \\
1 & \cdots & 1 & 0 & \cdots & 0 \\
\vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\
x_1^k & \cdots & x_N^k & 0 & \cdots & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\vdots \\
\lambda_N \\
\alpha a_0 \\
\vdots \\
\alpha a_k
\end{bmatrix}
=
\begin{bmatrix}
C_{10} \\
\vdots \\
C_{N0} \\
1 \\
\vdots \\
x_0^k
\end{bmatrix}.
\tag{38}
$$

This is the same extended form of normal equations encountered with empirical Radial Basis Functions (RBFs) instead of (generalized) covariance functions. Notice that when the mean is known and constant, the process can be centered with the Lagrange multiplier $\alpha$ set to zero, and these extended normal equations reduce to the standard Wiener-Hopf equations.

Such an extended covariance matrix is easily invertible under minimal conditions as

$$
\begin{pmatrix} C & B \\ B^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} C^{-1} + C^{-1} B H B^T C^{-1} & -C^{-1} B H \\ -H B^T C^{-1} & H \end{pmatrix} \text{ with}
$$
$$
H = -\left(B^T C^{-1} B\right)^{-1},
\tag{39}
$$

where $H$ is sometimes called Schur's Complement of $C$ in the extended matrix (see, e.g., the work of Kailath et al. [3] for more details). Hence under minimal conditions, these extended normal equations can be solved explicitly for the linear coefficients $\lambda_\nu$ and the trend polynomial coefficients. The inverse of a covariance matrix is usually called an information matrix. Also, generalized covariance functions in the sense of generalized functions are often advantageous for estimation in practical applications (see, e.g., the work of Blais [11] for more discussion with simulations).

Recalling that given observations $\{y(s)\}$, the optimal general estimate (not necessarily linear)

$$
\widehat{X}\left(t\right) = \mathbf{E}\left[X\left(t\right) \mid y\left(s\right)\right]
\tag{40}
$$

with the property that

$$
\text{Var}\left[X\left(t\right) - \mathbf{E}\left[X\left(t\right) \mid y\left(s\right)\right]\right] \leq \mathbf{E}\left[\left|X\left(t\right) - \Psi\left(y\left(s\right)\right)\right|^2\right]
\tag{41}
$$

for any linear or nonlinear estimator $\Psi(y(s))$ can sometimes be evaluated directly using numerical simulations. In fact, using the previous conventions,

$$
\mathbf{E}\left[X\left(t\right)\right] \equiv \mu_X\left(t\right) \equiv \int_{-\infty}^{\infty} x p_X\left(t\right) dx,
\tag{42}
$$

and with the help of Bayes' theorem,

$$
\begin{aligned}
\widehat{X}(t) &= \int_{-\infty}^{\infty} x p_{X|y}(t,s)\, dx \\
&= \int_{-\infty}^{\infty} x \left\{ \frac{p_{y|X}(t,s)\, p_X(t)}{\left\| p_{y|X}(t,s)\, p_X(t) \right\|} \right\} dx
\end{aligned}
\tag{43}
$$

which can be evaluated directly using Monte Carlo simulations in simple cases.

For example with Gaussian variables $Z_\nu$ such that $\mathbf{E}[Z_\nu \mid u] = \sigma_\nu^2 \sum \alpha_{\nu\eta} u_\eta$ and variance $\sigma_\nu^2$, one can write explicitly for the conditional probability density function

$$
p(Z_\nu \mid u) = \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \exp\left\{ -\frac{\left(Z_\nu - \sigma_\nu^2 \sum \alpha_{\nu\eta} u_\eta\right)^2}{2\sigma_\nu^2} \right\},
\tag{44}
$$

and these can be multiplied in cases of several independent Gaussian variables. In general, such explicit formulation is obviously unknown. Using Bayes' Theorem as above and substituting the results into the integral for $\widehat{X}(t)$, the required conditional expectation can be evaluated in simple applications. More discussion and application examples can be found in the works of Pugachev [2], Dougherty [1], and Zhang [12].

## 6. Computational Examples and Discussion

In many practical applications, solving the preceding normal equations and/or their extended form in the most efficient manner is one primary objective. As mentioned above, in stationary cases with equispaced data, the use of FFTs is the common strategy as the diagonalizing of the covariance matrix is well known to be most efficient. However, this is not always possible in many situations. Computational alternatives are briefly discussed in the following paragraphs.

As different software packages use different conventions for discrete Fourier, wavelet Haar, and other transforms, the following is very explicit in the conventions used. First, the Fourier matrix $F$ of order $N$ has elements $f_{rs} = e^{2\pi irs/N}$, and the simple application with $N = 8$ for the covariance function $\mathrm{Cov}[d] = 1/(1 + d^2)$ gives (without any normalization)

$$
F^*\begin{pmatrix} 1 & \dfrac{1}{2} & \dfrac{1}{5} & \dfrac{1}{10} & \dfrac{1}{17} & \dfrac{1}{10} & \dfrac{1}{5} & \dfrac{1}{2} \end{pmatrix}
$$
$$
= \begin{pmatrix} 2.659 & 1.507 & 0.659 & 0.375 & 0.259 & 0.375 & 0.659 & 1.507 \end{pmatrix},
\tag{45}
$$

or using the corresponding circulant Toeplitz matrix (with proper normalization),

$$
F^* \begin{pmatrix}
1 & \frac{1}{2} & \frac{1}{5} & \frac{1}{10} & \frac{1}{17} & \frac{1}{10} & \frac{1}{5} & \frac{1}{2} \\
\frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{5} & \frac{1}{10} & \frac{1}{17} & \frac{1}{10} & \frac{1}{5} \\
\frac{1}{5} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{5} & \frac{1}{10} & \frac{1}{17} & \frac{1}{10} \\
\frac{1}{10} & \frac{1}{5} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{5} & \frac{1}{10} & \frac{1}{17} \\
\frac{1}{17} & \frac{1}{10} & \frac{1}{5} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{5} & \frac{1}{10} \\
\frac{1}{10} & \frac{1}{17} & \frac{1}{10} & \frac{1}{5} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{5} \\
\frac{1}{5} & \frac{1}{10} & \frac{1}{17} & \frac{1}{10} & \frac{1}{5} & \frac{1}{2} & 1 & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{5} & \frac{1}{10} & \frac{1}{17} & \frac{1}{10} & \frac{1}{5} & \frac{1}{2} & 1
\end{pmatrix} F
\tag{46}
$$

$$
= 8 \cdot \mathrm{diag}\begin{pmatrix} 2.659 & 1.507 & \cdots & 0.659 & 1.507 \end{pmatrix}
$$

as expected. Notice that formulating the preceding covariance matrix differently for the given covariance function $\mathrm{Cov}[d] = 1/(1 + d^2)$ would not imply the above diagonal matrix. Also, the scaling of this diagonal matrix would not be needed with normalized Fourier transforms.

Again considering the synthetic covariance function

$$
\mathrm{Cov}[d] = \frac{1}{\alpha^2 + d^2}
\tag{47}
$$

for some real scalar $\alpha$, the plot of $\mathrm{Cov}[d]$ for $d = 0, 1, \ldots, 512$ (using $\alpha = 1$ for simplicity) with corresponding power spectrum or Fourier transform of this covariance function is simply

$$
S(\omega) = \frac{1}{\alpha} e^{-\alpha|\omega|}
\tag{48}
$$

for frequencies $\omega$ which is shown in Figure 1(a). Adding some noise amplification within $-1 \le s \le 1$ in space or time, the covariance function becomes

$$
\mathrm{Cov}[s,d] = \frac{1}{\alpha^2 + d^2} + H(|s|)\, \delta(d)
\tag{49}
$$

for the Heavyside (or step) function

$$
H(|s|) = \begin{cases} 1 & \text{for } |s| \le 1, \\ 0 & \text{otherwise} \end{cases}
\tag{50}
$$

and scalar $\alpha$; its spectrum becomes

$$
S(\sigma, \omega) = \frac{1}{\alpha} e^{-\alpha|\omega|} + \sigma^{-1} \sin \sigma
\tag{51}
$$

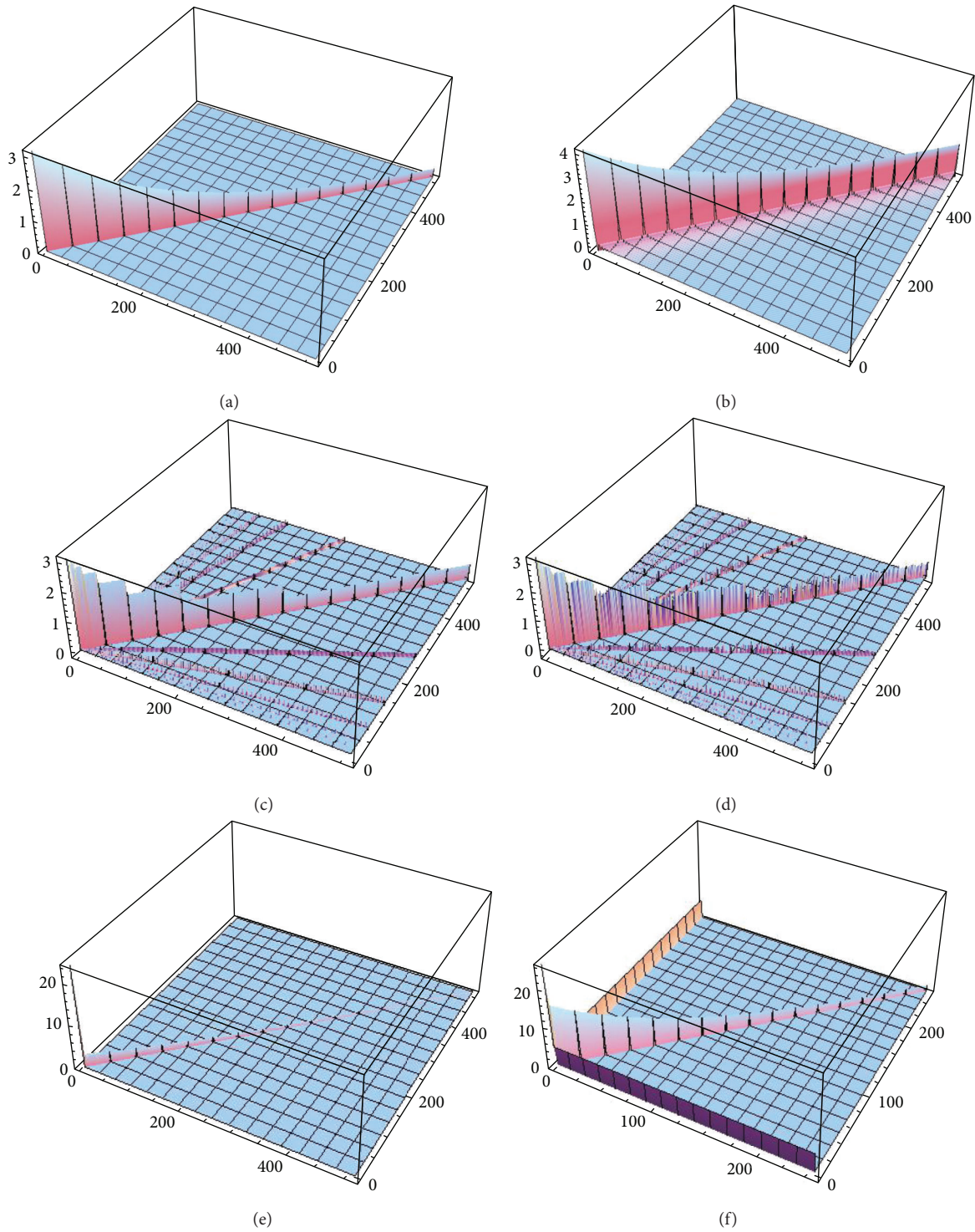which is nondiagonal as shown in Figure 1(b) with ripples away from the diagonal.

Figure 1: (a) Exponential spectrum. (b) Exponential + Sinc spectrum. (c) Haar spectrogram of $[1/(1 + d^2)]$. (d) Haar spectogram of $[1/(1 + d^2) +$ Nonstationary Noise]. (e) Singular spectrum of extended covariance matrix. (f) Power spectrum of extended covariance matrix.

Considering this covariance function Cov$[d]$ for multiresolution analysis with the Haar wavelet basis, using the discrete orthogonal basis $\left(1/\sqrt{2} \;\; 1/\sqrt{2}\right)$ and $\left(1/\sqrt{2} \;\; -1/\sqrt{2}\right)$ with order 512, the corresponding spectrogram is shown in Figure 1(c) which is sparse but not diagonal as the power spectrum. Modifying the level of noise with this covariance function as follows

$$\mathrm{Cov}\,[s, d] = \frac{1}{\alpha^2 + d^2} + \begin{cases} 0.5 & \text{for } s = 1, 2, \ldots, 256, \\ 0.2 & \text{for } s = 257, 258, \ldots, 512 \end{cases}$$

$$(52)$$

does not imply much variation in the corresponding spectrogram (see Figure 1(d)). Similar results were obtained by Keller [13] with Haar wavelets and Zhang [12] using a slightly different orthogonal basis. Although the orthogonal Haar wavelet and similar transforms would seem most appropriate for such covariance matrices, they only succeed in making the covariance matrix sparse but not diagonal. The EOF approach does diagonalize any symmetric positive definite covariance matrix by the very nature of the eigenvector decomposition. Such eigenvector and eigenvalue decompositions are computationally demanding, but it is worth noting that the biorthogonal basis construction is computationally advantageous whenever some arbitrary basis is available, as mentioned above.

Next, assuming an unknown mean for the simulated stochastic process, the covariance matrix Cov[$d$] can easily be modified such that the last column and row are 1's with 0 as the last diagonal element; the corresponding singular spectral matrix is shown in Figure 1(e). Using this modified covariance matrix Cov[$d$], its DFT spectral matrix is shown in Figure 1(f) which shows anomalous effects due to the border modifications of Cov[$d$]. In other words, the DFT and FFT approaches are not appropriate with extended covariance matrices even with equispaced data. This is obviously an important result for applications of extended covariance and RBF normal matrices in geostatistics (see, e.g., the works of Chilès and Delfiner [14]).

## 7. Concluding Remarks

Stochastic time series are very common in geoscience, and better analytical tools and procedures for their modeling and filtering warrant more research and development. Multiresolution analysis and synthesis of sequences of observations and measurements are needed for all kinds of applications in geoscience and other fields.

DFTs and FFTs are normally used in stationary processes with equispaced data. However in practice, nonstationarities are often unavoidable, and even in increment stationary cases with equispaced data, DFTs and FFTs are not appropriate. Galerkin's projection methods especially with EOFs are more general and unfortunately more computationally demanding. However in some applications, only the leading spectral modes are necessary for analysis and inference thus reducing the necessary computational efforts.

For simple prediction applications in stationary cases, the normal equations can be solved using Cholesky's Square-Root algorithm which has well-known numerical properties; while in increment stationary cases, as well as various RBF formulations, some Cholesky and Givens reduction procedure can be used (see, e.g., the work of Blais [10]). For more in-depth numerical analysis, the SVD and KLT are most recommendable in general and visualization tools are very helpful in the structural analysis of covariance and information matrices.

## References

[1] E. R. Dougherty, *Random Processes for Image and Signal Processing*, SPIE/IEEE Series on Imaging Science & Engineering, IEEE Press, New York, NY, USA, 1999.

[2] V. S. Pugachev, *Theory of Random Functions and Its Application to Control Problems*, Revised Translation by O. M. Blunn and Translation Edited by N. L. Johnson, Pergamon Press, Elmsford, NY, USA, 1965.

[3] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, Upper Saddle River, NJ, USA, 2000.

[4] V. C. Klema and A. J. Laub, "The singular value decomposition: its computation and some applications," *IEEE Transactions on Automatic Control*, vol. AC-25, no. 2, pp. 164–176, 1980.

[5] J. G. Lewis, *Algorithms for sparse matrix eigenvalue problems [Ph.D. thesis]*, Stanford University, Palo Alto, Calif, USA, 1977, Report No. 77-595.

[6] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1980.

[7] K. Y. Kim and Q. Wu, "A comparison study of EOF techniques: analysis of nonstationary data with periodic statistics," *Journal of Climate*, vol. 12, no. 1, pp. 185–199, 1999.

[8] G. R. North and R. F. Cahalan, "Predictability in a solvable stochastic climate model," *Journal of the Atmospheric Sciences*, vol. 38, no. 3, pp. 504–513, 1981.

[9] A. Björck, *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics, Philadelphia, Pa, USA, 1996.

[10] J. A. R. Blais, "Least squares for practitioners," *Mathematical Problems in Engineering*, vol. 2010, Article ID 508092, 19 pages, 2010.

[11] J. A. R. Blais, "Generalized covariance functions and their applications in estimation," *Manuscripta Geodaetica*, vol. 9, no. 4, pp. 307–322, 1984.

[12] Z. Zhang, *A study on canonical expansion of random processes with applications in estimation problems [M.S. thesis]*, Department of Geomatics Engineering, University of Calgary, Calgary, Canada, 2011, UCGE Report No. 20327.

[13] W. Keller, "A wavelet approach to non-stationary collocation," in *Geodesy Beyond 2000. The Challenge of the First Decade*, K. P. Schwarz, Ed., vol. 121 of *International Association of Geodesy Symposia*, pp. 208–214, Springer, Berlin, Germany, 2000.

[14] J. P. Chilès and P. Delfiner, *Geostatistics—Modeling Spatial Uncertainty*, Wiley, New York, NY, USA, 2nd edition, 2012.