# Multiresolution models for nonstationary spatial covariance functions

Douglas Nychka [1]
Geophysical Statistics Project, National Center for Atmospheric Research,
Christopher Wikle,
Department of Statistics, University of Missouri
J. Andrew Royle
U.S. Fish and Wildlife Service Adaptive Management and Assessment Team

Abstract

Many geophysical and environmental problems depend on estimating a spatial process that has nonstationary structure. A nonstationary model is proposed based on the spatial field being a linear combination of multiresolution (wavelet) basis functions and random coefficients. The key is to allow for a limited number of correlations among coefficients and also to use a wavelet basis that is smooth. When approximately 6% nonzero correlations are enforced, this representation gives a good approximation to a family of Matern covariance functions. This sparseness is important not only for model parsimony but also has implications for the efficient analysis of large spatial data sets. The covariance model is successfully applied to ozone model output and results in a nonstationary but smooth estimate.

## 1 Introduction

Many scientific problems involve geophysical or biological spatial processes that are nonstationary. Some examples include meteorological and ocean measurements, environmental pollutants and disease incidence. A statistical problem in all of these areas is to estimate the spatial covariance function without imposing unreasonable restrictions on its form. In this work we propose a multiresolution (wavelet) model that can adapt to heterogeneous spatial correlations and also lends itself to efficient computational algorithms for analyzing large spatial data sets.

As a motivating example in this work we consider the daily average surface (ambient) ozone concentration for a Midwest region of the the US. Based on heterogeneous spatial factors, such as the sources of the precursors to ozone and different meteorological conditions, one expects that the covariance function for the ozone field will vary depending on spatial location. Some issues for ozone pollution are in making spatial predictions at locations where measurements are not made and also attaching a measure of uncertainty to these predictions. Uncertainty measures, such as a prediction standard error, are useful not only for inferences for a given set of data but also to guide thinning, augmenting or designing monitoring networks for the future.

[1]Corresponding Author: National Center for Atmospheric Research, PO Box 3000, Boulder CO 80307-3000, USA

A heuristic principle is that although spatial predictions may not be sensitive to the assumed covariance function the standard errors implied by different covariance models can vary widely. For this reason it is important to have accurate models for the covariance structure. Beyond pointwise error estimates, there is a growing interest in the geophysical sciences to use realizations of a spatial field consistent with observed data that reproduce the stochastic features of the field. In a statistical context it is natural to construct ensembles of possible fields by sampling from the posterior, or conditional distribution, of the spatial field given the observed data. (This is also known as conditional simulation in geostatistics.) The validity of this conditional distribution will depend on how well the spatial process has been modeled and for this reason accounting for nonstationary structure is important.

Past work on estimating nonstationary spatial fields has included moving window kriging using a stationary covariance (e.g. [6]), nonlinear deformation of the geographic coordinates ([12]) or variable convolution of a stationary process ([7]). More recently, change of support models have demonstrated the potential for incorporating nonstationary structure ([4]) and a method that is similar in spirit to the approach in this paper is the use of empirical orthogonal functions (e.g. [14]). These methods all have different advantages and a collateral benefit of a multiresolution model is to provide a compact representation for some of the other kinds of models cited above. It should be noted that some of the ideas in this paper derive from more theoretical treatments from the work of Donoho and Mallat ([3], [9] ) but we have made some extensions that have practical import.

The next section discusses representing a spatial process and as a sum of fixed basis functions with sparsely correlated random coefficients. Section 3 introduces multiresolution bases and presents a particular wavelet (the W-transform [8]) that we have found useful. The next section gives some results for the approximation properties of this multiresolution for standard families of covariances and Section 5 explains how to estimate these models from data. Section 6 gives an example of a nonstationary covariance model for daily ozone in the Midwest.

## 2 Spatial Models

We assume that $z(\boldsymbol{x})$, is the value of a random field, e.g. ozone concentration at location $\boldsymbol{x}$, with covariance function

$$k(\boldsymbol{x}, \boldsymbol{x}') = COV(z(\boldsymbol{x}), z(\boldsymbol{x}'))$$

Throughout this paper we will assume that $z$ is a mean zero, Gaussian process and so the covariance function completely describes its stochastic properties. The covariance kernel has an eigenvalue/eigenfunction decomposition of the form

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sum_{\nu=1}^{\infty} \lambda_\nu \psi_\nu(\boldsymbol{x}) \psi_\nu(\boldsymbol{x}')$$

that holds for all covariance functions both stationary and nonstationary. Moreover, the actual process can be represented as

$$z(\boldsymbol{x}) = \sum_{\nu=1}^{\infty} \sqrt{\lambda_\nu}\alpha_\nu\psi_\nu(\boldsymbol{x})$$

where $\{\alpha_\nu\}$ are independent random variables, distributed $N(0,1)$.

The basic idea of this paper is to use a wavelet basis in place of the eigenfunctions and relax the condition on $\{\alpha_\nu\}$ to allow some correlation among these coefficients. The main contribution is how to estimate these covariances among the coefficients. In order to implement the models for data it is useful to rephrase this problem for a discrete set of points.

## 2.1 The Discretized Model

Let $\boldsymbol{z}$ be the values of the field on a large, rectangular grid (and stacked as a vector). This discretization is partly a computational device and should not influence the statistical analysis; we assume that the grid is chosen fine enough to resolve all relevant spatial features. Throughout this discussion we denote the total number of grid points as $m$ and let $\Sigma = COV(\boldsymbol{z})$ be the $m \times m$ covariance matrix among grid points. One can always find the eigenvector/eigenvalue decomposition for $\Sigma$. $\Sigma = \Psi D \Psi^T$ where $\Psi\Psi^T = I$ and $D$ diagonal. Here the columns of $\Psi$ are individual basis functions evaluated on the grid but stacked as a single column vector. One also has the representation $\boldsymbol{z} = \Psi H \boldsymbol{a}$ where $\boldsymbol{a}$ is a vector of independent $N(0,1)$ random variables and $D = H^2$. To emphasize the basic idea in this work, the eigen decomposition suggests an alternative way of building the covariance by specifying the basis functions and a matrix $H$. However, $\Psi$ need not be orthogonal and $H^2$ need not be diagonal. The primary constraint is that the implied covariance matrix and corresponding spatial estimator be readily computable and approximate a variety of standard models. In the rest of this paper $\Psi$ will denote a matrix whose columns form a discrete basis. But it may not necessarily be based on the eigenvectors of a particular covariance matrix.

For most problems $m$ is big and $\Sigma$ is gigantic. Even for the small example reported in Section 6 of this work, $m = 48^2$ and so $\Sigma$ will have approximately 2.5M unique elements! Dealing with such large matrices is discussed in the next section.

## 2.2 Computing a spatial process estimate

We end this section by a brief motivation for the kind of computational advantages provided by a multiresolution covariance model. In order to do so we review the equations for the conditional multivariate normal. The basic Gaussian model discussed here can be greatly extended, however we focus on the simplest case to isolate the main computational challenge.

Consider an observational model

$$\boldsymbol{y} = K\boldsymbol{z} + \boldsymbol{\epsilon}. \tag{1}$$

where $\boldsymbol{y}$ is a vector of $n$ observations and $\boldsymbol{\epsilon}$ is $MN(0, \sigma^2 I)$. The $n \times m$ matrix $K$ is usually an incidence matrix of ones and zeroes with a single one in each row indicating the position of each observation with respect to the grid.

The simplest spatial inference is to find the conditional normal distribution of $\boldsymbol{z}$ given $\boldsymbol{y}$ and this distribution has conditional mean vector

$$\hat{\boldsymbol{z}} = \Sigma K^T (K\Sigma K^T + \sigma^2 I)^{-1} \boldsymbol{y} \tag{2}$$

and conditional covariance matrix

$$\Sigma - \Sigma K^T (K\Sigma K^T + \sigma^2 I)^{-1} K\Sigma \tag{3}$$

$\hat{\boldsymbol{z}}$ can also be identified as the best linear unbiased (Kriging) estimate of $\boldsymbol{z}$.

When matrices are large the conditional mean vector is most efficiently solved for using approximate, iterative methods. These methods, such as conjugate gradient (see [5]), do not demand the storage of $\Sigma$ but do require that that one can multiply $\Sigma$ and $K$ by arbitrary vectors efficiently. $K$ will typically be a matrix that is sparse and so can be multiplied with a vector easily. Given a decomposition $\Sigma = \Psi H^2 \Psi^T$ efficient multiplication hinges on the structure of $\Psi$ and $H$. Through the choice of a multiresolution basis there are fast recursive algorithms for multiplying $\Psi$ and $\Psi^T$ with a vector. If $H$ is also sparse then the chain is complete and one can also multiply $\Sigma$ quickly.

Computing the conditional covariance is difficult and we suggest a Monte Carlo strategy that leverages an efficient algorithm for the conditional mean. Let $\boldsymbol{u}$ be generated as $MN(0, \Sigma)$. Generate a pseudo data vector $\boldsymbol{y}^* = K\boldsymbol{u} + \boldsymbol{\epsilon}$ and compute $\boldsymbol{u}^* = \boldsymbol{u} - \Sigma K^T (K\Sigma K^T + \sigma^2 I)^{-1} \boldsymbol{y}^*$. Simple matrix algebra will show that $\hat{\boldsymbol{z}} + \boldsymbol{u}^*$ will be a sample from the correct conditional distribution. Performing this several times will give an ensemble of fields and, of course, finding the sample covariance across the ensemble provides a Monte Carlo based estimate of the conditional covariance.

# 3 Multiresolution Bases

We will generate a basis for expanding the covariance using repeated translations and scalings of a few fixed functions. Multiresolution methods, in particular regression on wavelets, have received much recent attention in the statistics literature, especially in their ability to provide estimates of functions that have discontinuities or varying degrees of smoothness over their domain. The reader is referred to the review article by Nason and Silverman [10] for more background and development. The local support of these basis functions lends themselves to nonstationary fields because the stochastic properties can also be controlled locally. Changing the variances and covariances of groups of individual coefficients will only have a local impact on the spatial field and facilitates representing nonstationary structure.

In this section we give a qualitative description of multiresolution bases. One advantage of wavelet models is their efficient computation due to the discrete wavelet transform (DWT). Because of its discrete nature, the DWT only approximates the exact translation and scaling of fixed functions. However, the approximation is accurate for the coarser levels of resolution and the intuition of basis functions having a scale of resolutions and local support is important for understanding their benefit.

## 3.1 A one-dimensional basis

Although the main practical interest is in two dimensional fields, for illustration and some later examples we first give a qualitative description of a continuous multiresolution basis in one-dimension. The key feature that we would emphasize is building a complete basis through the repeated translation and scaling of a few fixed template functions. We start with two templates, mother, $\psi$, and father, $\phi$, wavelets both defined on $[0, 1]$ and a coarsest level of resolution, say J. The W wavelets used in this work are plotted in Figure 1 and a basis of 32 functions is displayed in Figure 2. The first $J$ basis functions are similar to the father wavelet translated to $J$ equally spaced locations. These are given in plot (a) of Figure 2 for $J = 4$. The father wavelet only appears in this first $J$ set and all subsequent basis functions are similar in form to the mother wavelet. The next $J$ basis functions are the mother wavelets translated in the same manner and are in plot (b). The next generation of basis functions has twice the resolution and twice as many members (8) and is similar to a scaling and translation of the mother wavelet. Plot (c) of Figure 2 shows this generation. This cascade continues with the number of members in each subsequent generation and the resolution increasing by a factor of two. Plot (d) completes the basis of size 32.

In this work we concentrate on a basis derived from the W transform. Kwong and Tang [8] proposed the father and mother wavelet functions based on simple families of filter weights with respect to the discrete wavelet transform (DWT). The reason for this choice over more common wavelets is that they appear to approximate the shape of common covariance models and handle boundaries easily. We will refer to these as W wavelets and the quadratic parents of this family are given in (Figure 1). Both are piece-wise quadratic splines but unlike other popular wavelet bases are not orthogonal or compactly supported. The W-wavelet is implemented in the Fields package [11] for the Splus and R statistical environments. The source code is readily available and provides a clear definition of the filter weights for the transforms. The W-wavelet is also similar to the biorthognal, b-spline wavelet (bs3.1) implemented in the S+Wavelets [1] package.

## 3.2 A two dimensional multiresolution basis

A two dimensional basis can be constructed through translations and scaling of four template functions that replace the role of the father and mother wavelets from the one dimensional case. Here we give a qualitative description how

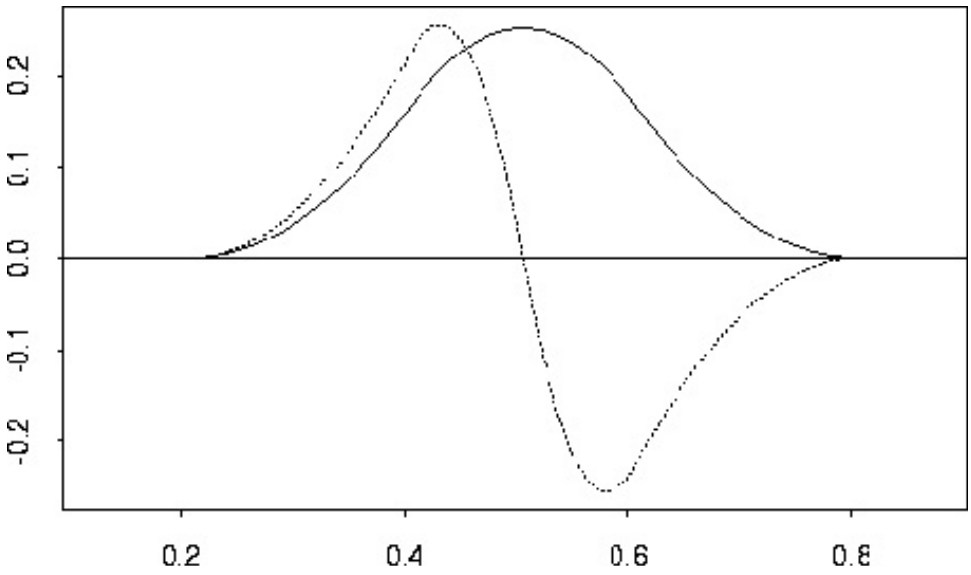


Figure 1: Continuous versions of the Father (solid) and Mother (dashed) W-transform wavelets

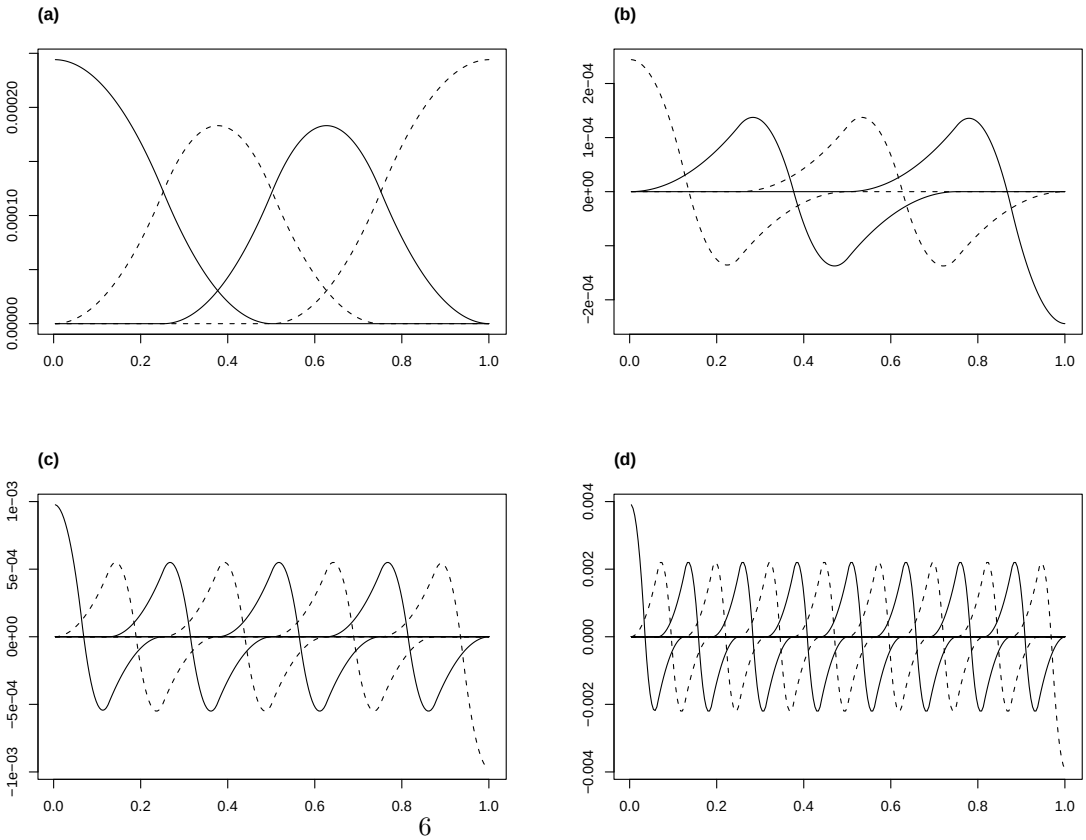




Figure 2: Family of 32 basis functions based on an approximate translation and scaling of the father and mother wavelets from Figure 1. Basis functions at the end points differ slightly in shape due to boundary corrections.

this is done. To start, form the tensor products of the one-dimensional father and mother wavelets functions to obtain four, two-dimensional functions with domain on $[0,1] \times [0,1]$:

$$\begin{aligned}\mathcal{S}(x_1, x_2) &= \phi(x_1)\phi(x_2) \\ \mathcal{H}(x_1, x_2) &= \psi(x_1)\phi(x_2) \\ \mathcal{V}(x_1, x_2) &= \phi(x_1)\psi(x_2) \\ \mathcal{D}(x_1, x_2) &= \psi(x_1)\psi(x_2)\end{aligned}$$

These are plotted in Figure 3. Here $\mathcal{S}, \mathcal{H}, \mathcal{V}$, and $\mathcal{D}$ stand for smooth, horizontal, vertical and diagonal respectively and denote their ability to represent features in these orientations. The later three we refer to as detail functions.

To construct the 2-d basis, one starts with a coarsest scale and translates of the scaled $\mathcal{S}$. As a specific example, if we choose the beginning level $J = 4$ then a scaled $\mathcal{S}$, would be centered at a $4 \times 4$ grid of locations yielding 16 basis functions. The next 16 functions are scaled translates of $\mathcal{H}$ followed by translates of $\mathcal{V}$ and $\mathcal{D}$ for a total of 64 functions in this first generation. Subsequent generations are based only on the three detail functions derived from the mother wavelet and in analogy to the one dimensional case involve reduction by a factor of two and translation on a grid. For example, with the specific choice of $J = 4$, the next generation would involve three sets of 64 basis functions being scaled translates of $\mathcal{H}$, $\mathcal{V}$, and $\mathcal{D}$ on an $8 \times 8$ grid.

### 3.3 Discrete Wavelet transform

The discrete wavelet transform (DWT) is a fast algorithm to compute the coefficients of a basis that approximates the exact multiresolution. In the notation of the previous section, the DWT and its inverse can be used to rapidly multiply $\Psi^{-1}$ and $\Psi$ by arbitrary vectors. The basis functions comprising the columns of $\Psi$ are approximately equal to the exact translations and dilations of the smooth and detail functions with the approximation improving as the level of resolution decreases.

The key idea behind the DWT is recursion. At each step an image of size say, $n_1 \times n_2$ is decomposed through four finite length separable, linear filters into four equal submatrices of smooth, horizontal, vertical and diagonal terms. The three matrices for the $\mathcal{H}$, $\mathcal{V}$ and $\mathcal{D}$ components at this level of resolution are saved. The submatrix of $(n_1/2) \times (n_2/2)$ smoothed coefficients now becomes the "image" for the next step and the filtering is repeated. This process continues until one reaches a smoothed image of a particular size. By varying the filters, the net result is a family of four algorithms that are linear in the image size and are equivalent to multiplication of a vector by $\Psi$, $\Psi^{-1}$ or their transposes. It should be noted that the standard DWT refers to the specific multiplication of a vector by $\Psi^{-1}$ and, for many wavelet choices, $\Psi$ is an orthonormal matrix.

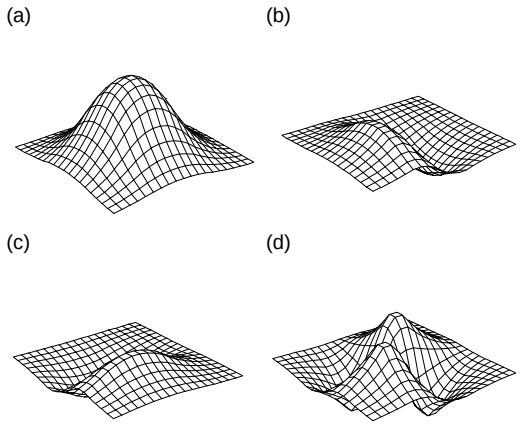


Figure 3: Tensor products of the father and mother wavelets associated with the W-transform. (a) $\mathcal{S}$ , (b) $\mathcal{H}$, (c) $\mathcal{V}$, and (d) $\mathcal{D}$

# 4 Approximating other covariances

Before considering nonstationary models it is of interest to investigate how well the multiresolution representation can approximate standard covariance models. The key issue is enforcing sparsity in the covariance matrix among coefficients without it losing the approximation properties.

Suppose $\Sigma$ is the covariance matrix for a spatial process. For any discrete basis, $\Psi$, it is always true that

$$\Sigma = \Psi D \Psi^T = \Psi H^2 \Psi^T \tag{4}$$

for $D = \Psi^{-1}\Sigma(\Psi^T)^{-1}$ and $H = D^{1/2}$. However, this decomposition is only useful if $D$ or $H$ are close to diagonal (i.e. sparse).

As an example Figure 4 illustrates the structure of these matrices for a one dimensional grid of 256 equally spaced points on $[0, 1]$ and an exponential covariance, $k(x, x') = exp\{-|x - x'|/\theta\}$ with $\theta = 1/8$. The plot (a) is a scaled version of $log_{10}(|D|)$ and (b) the leading $16 \times 16$ submatrix of $D$. The main features are clear, the elements of $D$ fall off rapidly with scale and many off diagonal elements are zero. Moreover most of the significant off diagonal elements are associated with the coarsest scale basis functions. Although not shown in this figure, the sparsity is amplified further by considering $H$. This example is motivation for keeping the dominating elements of $H$ but setting to zero small elements.

## 4.1 Enforcing sparsity in $H$

In the spirit of threshold estimators used in wavelet regression and the initial ideas from [3] we investigate thresholding to enforce sparsity in $H$. We propose a simple strategy:

- Retain all diagonal elements of $H$
- For the finest levels of resolution set all off diagonal elements of $H$ to zero.
- For the remaining elements, given $i \neq j$ and a threshold $\tau$.

$$\widehat{H}_{i,j} = \begin{array}{ll} H_{i,j} & \text{if } |H_{i,j}| > \tau \\ 0 & \text{otherwise} \end{array}$$

Figure 5 reports results for approximating the exponential covariance model used above. The grid is 256 equally spaced points on $[0, 1]$ and we have chosen $\tau$ to be either the 98% or 97% quantiles for $\{|H_{i,j}|\}$. The approximate covariance matrix is reconstructed and three rows of this matrix are plotted along with the true covariance values. Given 97% decimation of $H$, the agreement is striking. However, for more decimation there are some inaccuracies in the peak height and some ringing.

As a final illustration we consider a one dimensional nonstationary model obtained by deformation. Let $T(x) = .4x + .6\Phi((x - .57)/.1)$ where $\Phi$ is the

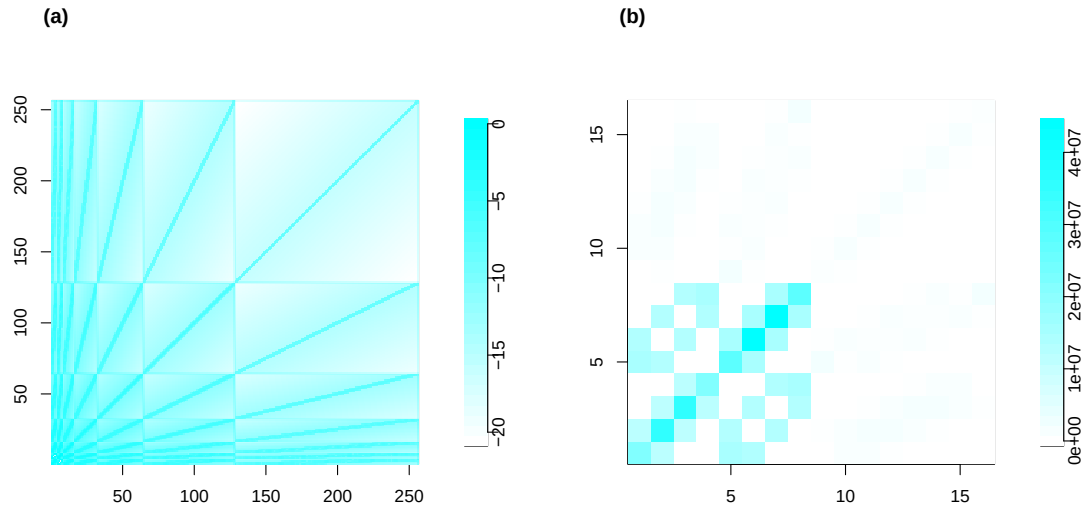


Figure 4: Approximation to a 1-D stationary covariance function. Here the spatial covariance is the exponential with range $\theta = 1/8$. The discrete covariance matrix is this covariance evaluated on 256 equally spaced grid points on [0,1]. Image (a) is the matrix $D$ in a log relative scale, the elements are: $log_{10}(|D|/max|D|)$. Basis functions are ordered according the presentation in Figure 2 from coarsest scales to finest scales. Image (b) is an enlargement of the leading $16 \times 16$ submatrix of $D$. Here the values are not scaled.

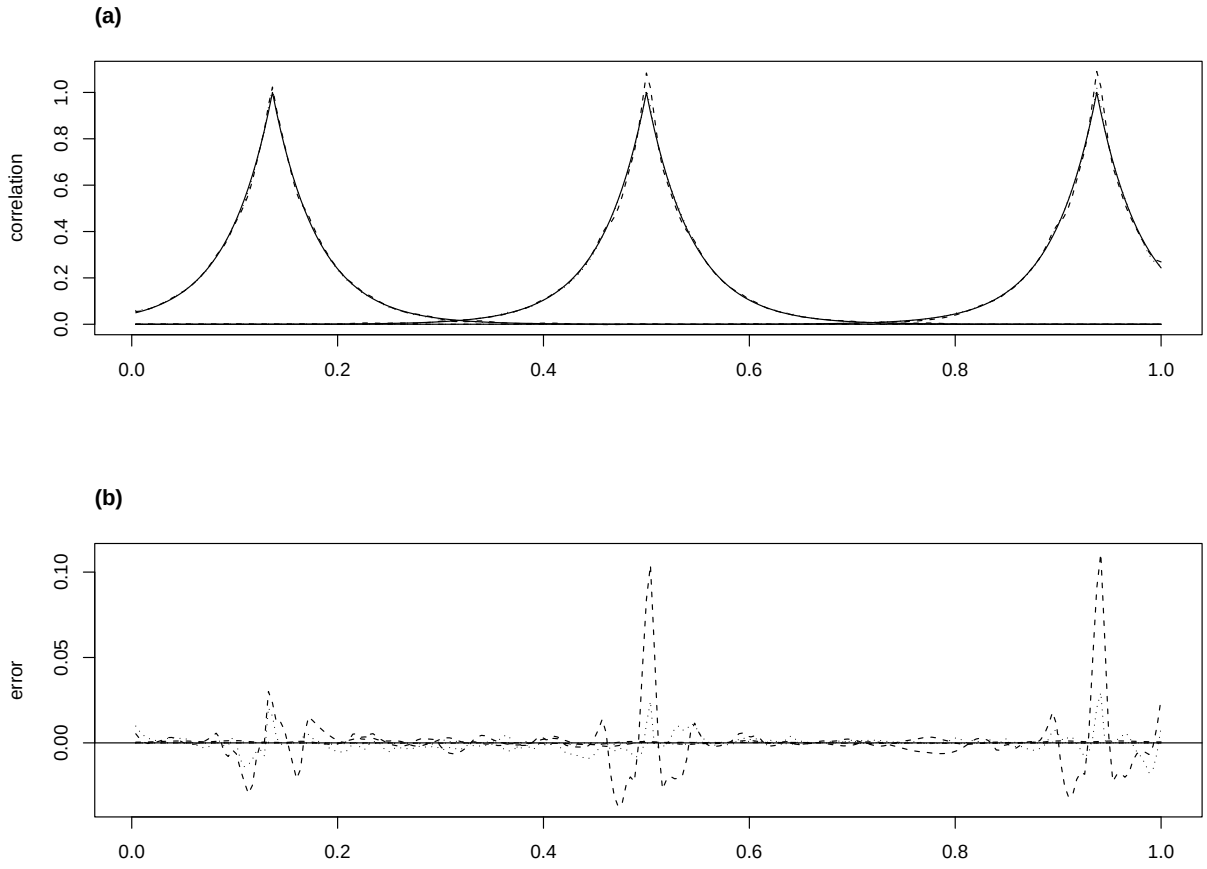


Figure 5: Approximation of a 1-d stationary covariance by wavelets at 97% and 98% decimation of H. The exponential covariance matrix with $\theta = 1/8$ is approximated by decimating the elements of $H$ by 97% or 98% and then reforming the covariance matrix. Three rows of the covariance matrices are plotted in (a) true( solid) 97% (dotted) and 98% (dashed). The differences between the approximate covariances and the true function are plotted in (b).

standard normal distribution function. Define the deformation covariance as $k(x, x') = exp\{-|T(x)-T(x')|/\theta\}$ with $\theta = 1/8$. This results in a nonstationary covariance on $[0, 1]$ where there are short range correlations in the middle of the interval and longer ranges near either end point.

Again we decimate $H$ by 98% and 97% and assess the approximation (Figure 6) to the true covariance matrix. Here the wavelet approximation does well in tracking individual rows of the covariance matrix even with changing shape and range.

## 4.2 Approximating the Matern covariance family

The one-dimensional examples given above help to give some intuition concerning the W-transforms approximation and sparsity properties. However, a more useful comparison is for covariances associated with two dimensional spatial processes. In this section we provide more deliberate comparisons for the Matern family [13]. One key result is a universal mask for sparsity that facilitates estimating these models from data.

The Matern covariance family is indexed by the parameters $\nu$ (smoothness) and $\theta$ (range) and has the form

$$k(\boldsymbol{x}, \boldsymbol{x}') = \Phi_{\nu,\theta}(||\boldsymbol{x} - \boldsymbol{x}'||)$$

where

$$\Phi_{\nu,\theta}(r) = C\frac{(\theta r)^{\nu}}{\theta^{2\nu}}K_{\nu}(\theta r).$$

$K$ is a modified Bessel function of order $\nu$ and $C$ a normalization only depending on $\nu$. A Gaussian process with this covariance will have $\nu$ derivatives that exist in the mean square sense and the parameter $\theta$ controls the correlation range. Two important special cases are the exponential covariance when $\nu = .5$ and the Gaussian, obtained as a limit as $\nu \to \infty$.

For numerical results we consider a $16 \times 16$ grid of locations on $[0, 1] \times [0, 1]$ with the coarsest generation having 16 basis functions for the each of the 4 tensor products from Figure 3 and are centered on a $4 \times 4$ grid. The covariance models were generated according to the smoothness parameters $.5, 1.5$ and 4 and with range parameters $.125, .25, .5, .75$ giving a total of 12 different covariance functions. The marginal variance for these covariances has been normalized to one. The wavelet approximation was computed for 98% and 97% decimation and the results are summarized in Figure 7. For each covariance model the root mean squared error (RMSE) is found for each row of the difference between the actual covariance matrix and the wavelet approximation. Each boxplot summary is based on the $16^2 = 256$ RMSEs calculated in this way. Keeping in mind that the covariance models have marginal variance of one these RMSE are small relative to the size of the correlations. However, we see some degradation in the approximation for small ranges and the exponential model. This behavior may be related to the choice of $4\times4$ father wavelets setting the coarsest scale and also because the shape of the W-wavelets are closer to Gaussian than exponential

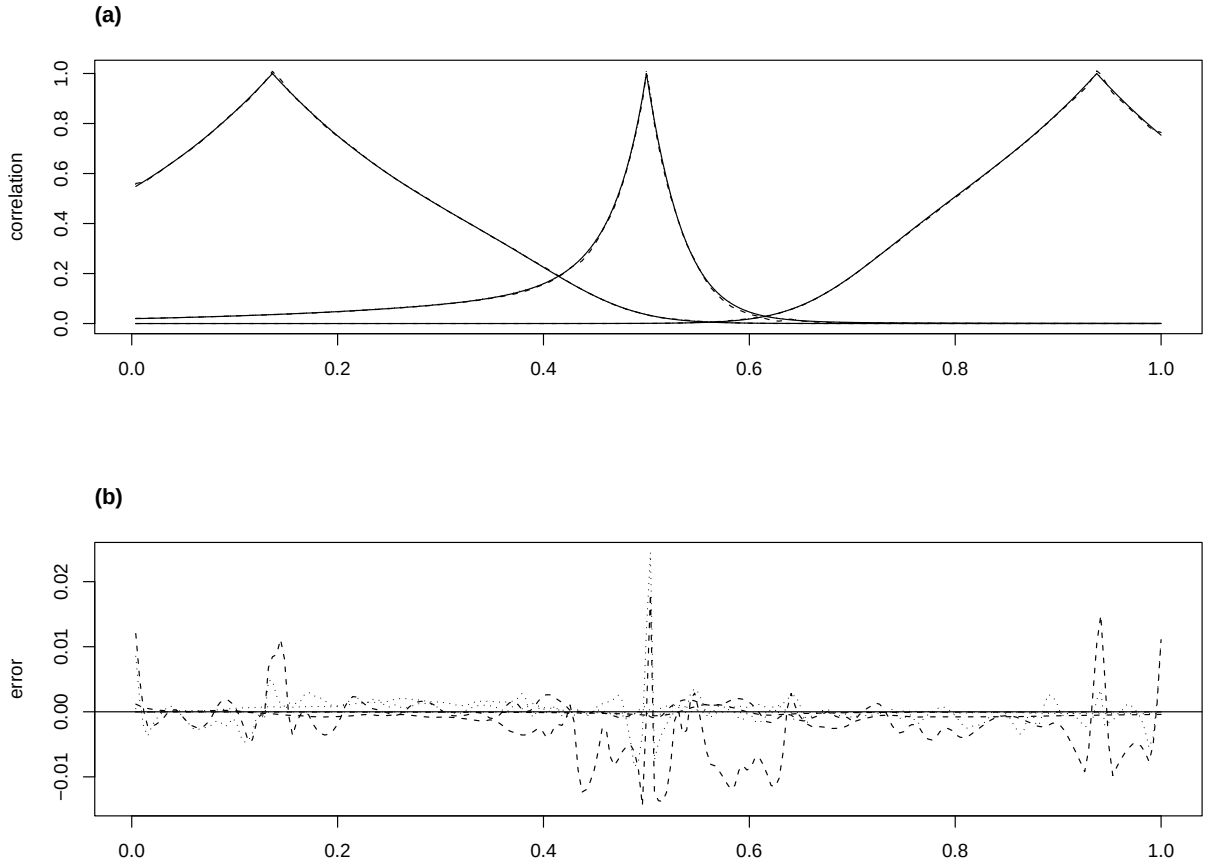


Figure 6: True and approximate covariances based on 97% and 98% decimation for the nonstationary deformation covariance. Three rows of the covariance matrices are plotted in (a) true( solid) 97% (dotted) and 98% ( dashed) are shown in the top. The differences between the approximate covariances and the true function are plotted in (b).

at their peak. Also, large values of the RMSE tend to be associated with edges of the domain.

Another important aspect of this study is investigating how the positions of the nonzero elements in the decimated $H$ matrix vary among the different models. For the case with 98% decimation the union of all nonzero elements across the 12 models amounts to a total of 3.3% nonzero elements and for 97% decimation there are 5.6% nonzero elements. These results suggest that one can restrict the models for $H$ to a limited set of elements and still have a good approximation to a wide range of models.

# 5 Covariance estimates from data

In many geophysical applications the spatial fields are observed over time and one can exploit temporal replication to estimate sample covariances. In this work we focus on this case and also for gridded data with the goal of deriving estimators that scale to large problems.

## 5.1 Sample estimates of $H$

Assume that independent copies of the field are observed over $T$ time points and let $\boldsymbol{Z}$ be an $m \times T$ matrix with each column being the stacked vector of centered observations ( i.e. mean zero) at a single time. By definition each column of $Z$ has covariance $\Sigma$. With gridded data and (independent) replications over time, one can get sample estimates of the elements of $H$. We expect that for most problems $m >> T$ and so the basic idea is to try to work with matrices of order $m \times T$ or $T \times T$ instead of $m \times m$. The sample covariance is

$$\widehat{\Sigma} = (1/T)\boldsymbol{Z}\boldsymbol{Z}^T$$

and so

$$\widehat{D} = (1/T)(\Psi^{-1}\boldsymbol{Z})(\Psi^{-1}\boldsymbol{Z})^T$$

($\widehat{D}$ is the sample covariance of the basis function coefficients.) This form motivates estimating $H$ via the singular value decomposition of $\Psi^{-1}\boldsymbol{Z}$. Let $V\Lambda U^T = (1/\sqrt{T})(\Psi^{-1}\boldsymbol{Z})$ where $V$ and $U$ are orthogonal matrices and $\Lambda$ a $T \times T$ diagonal matrix of nonnegative singular values. Setting

$$\widehat{H} = V\Lambda^{1/2}V^T,$$

it now follows that $\widehat{D} = \widehat{H}^2$. Based on the numerical results from Section 4.2 one is lead to a small number of nonzero elements of $H$ that have good approximation properties across a whole covariance family. This sparseness guarantees that we do not have to consider many off-diagonal elements. The entries that are nonzero can be computed efficiently based on $V$ and $\Lambda^{1/2}$.

Once the (nonzero) elements of $\widehat{H}$ are found one can:

- Further decimate them

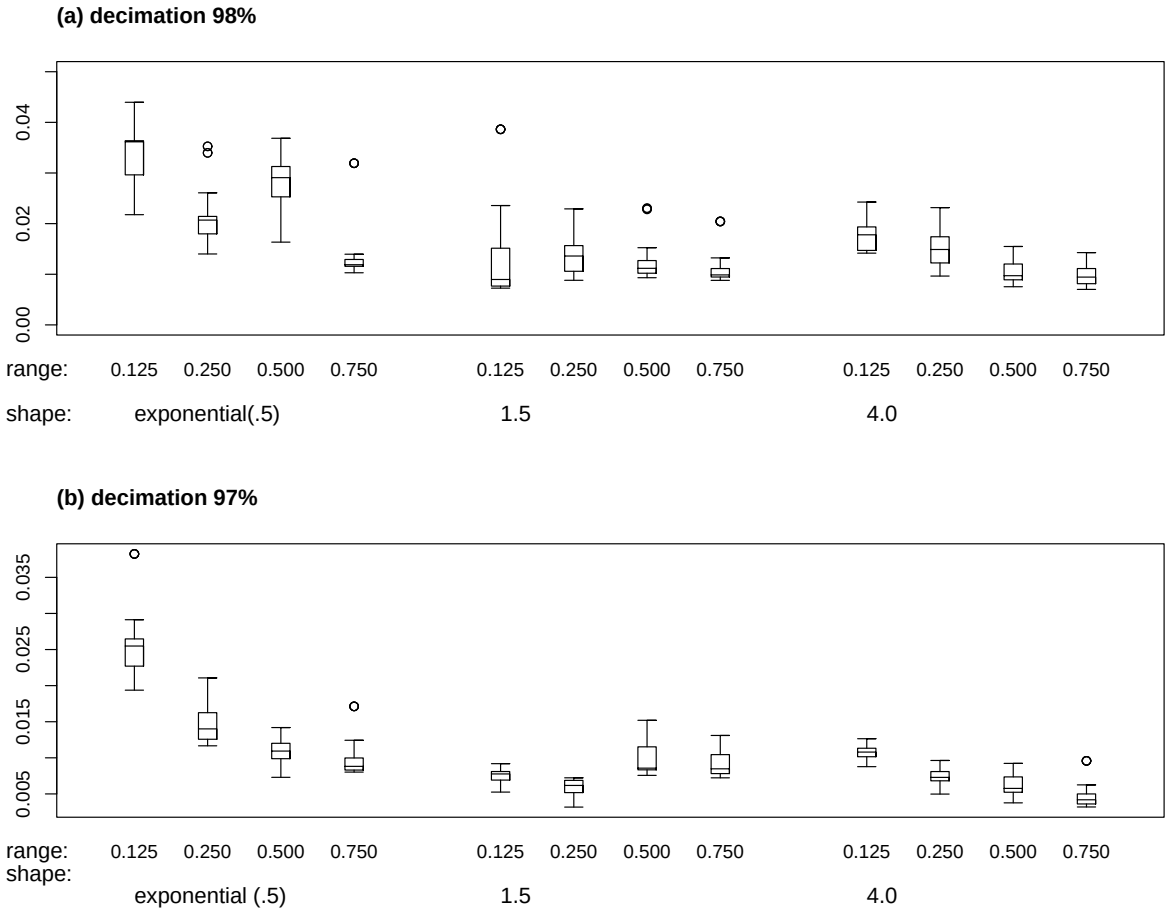


Figure 7: Comparison of the wavelet approximation to the family of Matern covariances. The boxplots are the RMSE for each row of the difference between the true and approximate correlation matrices.

- smooth across "spatially adjacent" entries.
- shrink toward a stationary model

In the example in the next section we simply use decimation, although we believe that smoothing entries may be useful in some cases.

# 6 Example with Surface Ozone

We consider a numerical experiment based on the Regional Oxidant Model (ROM) and analyze daily average ozone produced by the output. ROM is an atmospheric chemistry model that simulates ozone formation and transport based on sources of pollution and meteorological conditions. The data used in this study is a $48 \times 48$ grid centered on Illinois and Ohio where each grid box is approximately $16 \times 16$ miles in size. Output is available for 79 days using meteorology from June-August 1987. For background on this model and its comparison to data see [2].

Ambient average daily ozone has a correlation range of the order of 300 miles and so it is reasonable to start with a $3 \times 3$ grid of father wavelets as the coarsest level of resolution. This division also facilitates the recursive algorithm because then the remaining factor of 48 is 16, a power of 2. In this example we simply decimated the leading $12 \times 12$ block of H by 90% and retained diagonal elements for the remaining levels. A refinement of this could be to smooth the diagonal elements in the finer levels.

The resulting covariance function appears to be nonstationary with some interesting structure. Figure 8 depicts the estimated marginal standard deviations at each ROM grid box. We found that this surface has not been smoothed much relative to the sample variances but note higher variability over urban areas such as Chicago and Detroit. Figure 9 depicts the covariance at four locations. Each member of the panel fixes a location and plots the covariance between all other points and the fixed location. These fixed locations are indicated by an x symbol on the plots. The top left hand location (a) is a rural area of Illinois and has a covariance that is longer ranged to the west but with low variability. The location in (b) is near more urban areas in Michigan. It has higher variability but also a longer correlation range than the locations in the bottom corners of the domain. Another interesting distinction is that the location (c) (near St Louis) has more isotropic contours. Although more extensive analysis of this data set is beyond the scope of this example it should be noted that the correlations tend to have longer range west of the locations such as (a) and (b). This might be due to the general motion of weather systems, and the corresponding transport of ozone, from west to east.

# 7 Discussion

We have shown that wavelets provide flexible methods for introducing nonstationary spatial structure and can reproduce standard spatial models. Due to

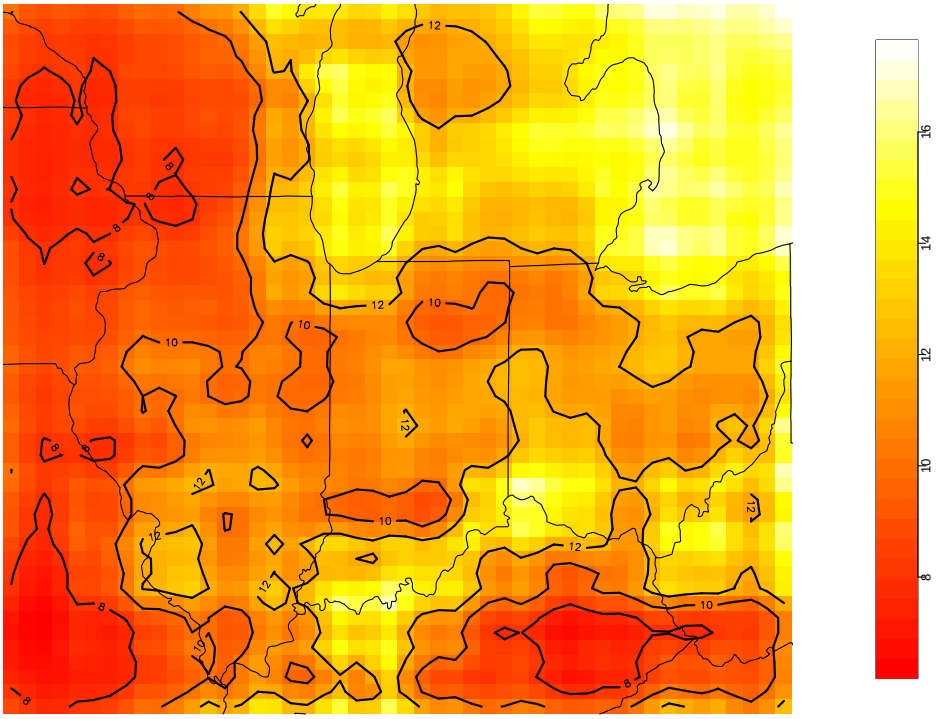


Figure 8: Estimated marginal standard deviations on the ROM grid. Contour levels are at (8,10,12) parts per billon.

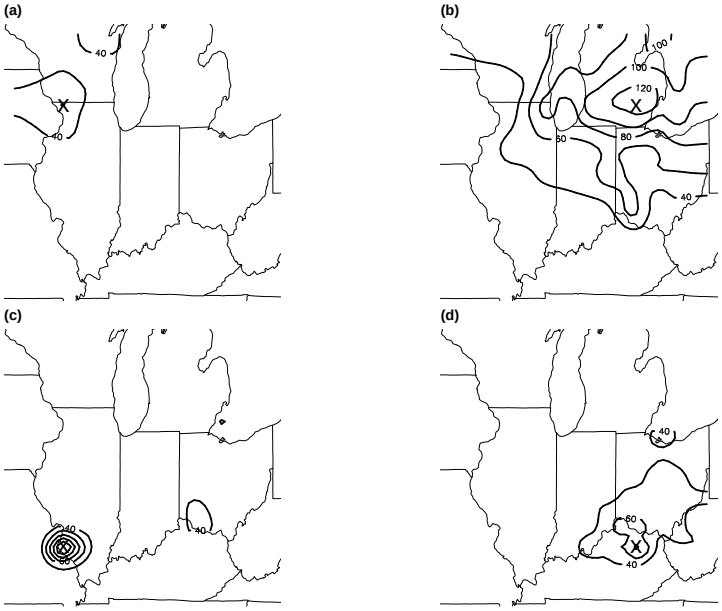


Figure 9: Estimated covariance surface at 4 sample locations for the ROM output. The image plots indicate the estimated covariance between points in the domain and the point location denoted by an x. Contour levels are at (40,60,80,100, 120) .

the efficiency of the discrete wavelet transform and also the enforced sparseness in the covariance matrix among coefficients, these models are also amenable to large spatial problems. The example using ROM output is surprising in how a regular basis can produce nonstationary but smooth covariance patterns. A key extension of this method is to irregular locations and we end this paper by suggesting an approach to this problem when temporal replicates of the field are available.

When data is not observed on a complete grid one can not take advantage of the DWT for matrix multiplication of the basis functions. Also there are more intrinsic problems in that many elements of $H$ are not identifiable. A formal solution is to write out a full Bayesian model for the field including a hierarchical model for $H$ and then attempt to sample from the posterior using Markov chain Monte Carlo techniques. We think this approach may be productive for moderate size problems but may not be easy to implement for many important large geophysical data sets. Here we suggest an iterative, Monte Carlo based approach that has less of a statistical foundation but is more direct.

Given spatial data at irregularly spaced locations and observed at several times, one starts by fitting a simple, possibly stationary model, to the sample covariances. Based on this starting covariance model, one samples the conditional distribution of the field on a regular grid given the observed data at each time point. The result is a time sequence of gridded fields that are consistent with the data and use the initial covariance for extrapolation. Given these gridded samples of the field one now fits a model to $H$ based on the ideas from this paper. Now one uses the multiresolution based covariance to sample the conditional distribution of the field given the data and to reestimate $H$. This last step is now repeated until convergence of the covariance estimates.

The basic idea of extrapolating irregularly spaced data to a regular grid and then using the DWT is not new. However, one crucial difference is that we propose to sample from the conditional distribution rather than use the posterior mean. This will have the effect of favoring the starting covariance model in regions where direct covariance information is limited but adapting to the data in areas where locations are dense.

Clearly there are many open statistical questions posed by the covariance model in this paper and we hope that its merits in flexibility and practicality spur more research.

# Acknowledgments

This work was supported by National Science Foundation grants DMS-93122686 and DMS-9815344. We would like to thank the two reviewers of this paper for their indulgence in reading a manuscript with many errors and their willingness to give constructive comments.

# References


[1] Bruce, A. and Gao, H. (1994) S+Wavelets User's Manual, Version 1.0, StatSci, a division of MathSoft, Inc., Seattle.

[2] Davis, J. M., Nychka, D. and Bailey, B. (2000). "A Comparison of the Regional Oxidant Model with Observed Data." Atmospheric Environment, 34, 2413-2423.

[3] Donoho, D., Mallat, S. and von Sachs, R.(1996) "Estimating covariances of locally stationary processes", In Proceedings of IEEE Time frequence and Time-Scale Symposium, Paris, July 1996. IEEE, New York

[4] Gabrosek, J. , Cressie, N. and Huang, H-C. (1998). "Spatio-Temporal Prediction of Level 3 Data for NASA's Earth Observing System." Proceedings of Third Spatial Accuracy Symposium, Quebec City, May 1998.

[5] Golub, G. and Van Loan, C. (1989). Matrix Computations. John Hopkins University Press. Baltimore.

[6] Haas, T. (1995). "Local Prediction of a spatio-temporal process with an application to wet sulfate deposition." Journal of the American Statistical Association, 90 , 1189-1199.

[7] Higdon,D Swall,J and Kern, J (1998) "Non-Stationary Spatial Modeling" INSTITUTE OF STATISTICS AND DECISION SCIENCES Discussion Paper Series Duke University, 98-17.

[8] Kwong, M.K. and Tang, P.T.P. (1994) "W-Matrices, Nonorthogonal Multiresolution Analysis, and Finite Signals of Arbitrary Length", Technical Report Argonne National Laboratory, MCS-P449-0794

[9] Mallat, S. Papanicolaou, G. and Zhang Z. (1998). "Adaptive covariance estimation of locally stationary processes." The Annals of Statistics 26, 1-47.

[10] Nason, G.P. and Silverman, B.W. (1994). The discrete wavelet transform. Journal of Computational and Graphical Statistics, 3, 163-191.

[11] Nychka, D., Meiring, W., Royle, J. A., Fuentes, M. and Gilleland, E. (2002). Fields: S Tools for Spatial Data http://www.cgd.ucar.edu/stats/software

[12] Sampson, P. and Guttorp, P. (1992), "Nonparametric estimation of nonstationary spatial covariance structure", Journal of the American Statistical Association, 87, 108-119.

[13] Stein, M. (1999) Interpolation of Spatial Data: Some theory for Kriging. Springer Verlag, New York.

[14] Wikle, C.K. and N. Cressie (1999) "A dimension reduced approach to space-time Kalman filtering" . Biometrika , 86 , 815-829.