

Multiresolution models for nonstationary spatial covariance functions

Douglas Nychka¹

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 a multiresolution (wavelet) basis functions and random coefficients. The key is to allow for a limited some 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/biological spatial processes are that nonstationary. Such applications cover a wide range of disciplines ranging from meteorological and ocean measurements, to environmental pollutants, to 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 correlation patterns 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 varying 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. The 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.

A heuristic principle is that although spatial predictions may not be sensitive to the assumed covariance function the implied standard errors by different covariance models can vary widely. For this reason it is important to have accurate models for

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

the covariance structure. Beyond pointwise error estimates, there is a growing interest in 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. [4]), nonlinear deformation of the geographic coordinates ([10]) or variable convolution of a stationary process ([5]). More recently, change of support models have demonstrated the potential for incorporating nonstationary structure ([2]) and a method that is similar in spirit to the approach in this paper is the use of empirical orthogonal functions (e.g. [11]). These methods all have advantages and drawbacks and, while we will focus on a different approach, it should be noted that a collateral benefit of multiresolution model may be simply to provide a compact representation for some of the kinds of models cited above. The chief advantage is that the subsequent use of the covariance function for spatial analysis is efficient using a multiresolution representation. It should be noted that some of the ideas in this paper derive from more theoretical treatments from the work of Donoho and Mallat ([1], [7]) but we have made some extensions that have practical import.

The next section presents a standard spatial model and discusses the main idea of this paper representing the spatial process and sum of fixed basis function with sparsely correlated random coefficients. Section 3 considers spatial processes using multiresolution bases and presents a peculiar wavelet (the W-transform [6]) that we have found useful. The next section gives some results for the approximation properties of the multiresolution for standard families of covariances. Section 5 estimates a nonstationary multiresolution model for daily ozone in the Midwest.

2 Spatial Models

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

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

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 the eigenvalue/eigenfunction decomposition of the form

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

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

$$f(\mathbf{x}) = \sum_{\nu=1}^{\infty} \sqrt{\lambda_{\nu}} \alpha_{\nu} \psi(\mathbf{x})$$

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 and to compute examples it is useful to rephrase this problem for a discrete set of points.

2.1 The Discretized Model

Let 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 .

Accordingly, let

$$\Sigma = COV(z) \tag{1}$$

be the covariance matrix among grid points and to introduce some notation we parallel the continuous representation of z given above. One can always find the eigenvector/eigenvalue decomposition for Σ . $\Sigma = \Psi D \Psi^T$, $\Psi \Psi^T = I$ and D diagonal. Here the columns of Ψ are individual basis functions evaluated on the grid but stacked as a single column vector. One also has the representation $z = \Psi H \mathbf{a}$ where \mathbf{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, Ψ 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.

For most problems m is big and Σ is a gigantic. Even for the small example reported Section 5 of this work, $m = 48^2$ and so Σ will have approximately 2.5M unique elements! Dealing with such large matrices is discussed in the next section.

2.2 Why sparsity?

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 burden.

Consider an observational model

$$\mathbf{y} = K\mathbf{z} + \boldsymbol{\epsilon}. \quad (2)$$

where \mathbf{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 1 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 \mathbf{z} given \mathbf{y} and this distribution has mean vector

$$\hat{\mathbf{z}} = \Sigma K^T (K \Sigma K^T + \sigma^2 I)^{-1} \mathbf{y} \quad (3)$$

and covariance

$$\hat{\Sigma} = \Sigma - \Sigma K^T (K \Sigma K^T + \sigma^2 I)^{-1} K \Sigma \quad (4)$$

$\hat{\mathbf{z}}$ can also be identified as the best linear unbiased (Kriging) estimate of \mathbf{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 [3]), do not demand the storage of Σ but do require that one can multiply Σ 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 Ψ and H . Through the choice of a multiresolution basis there are fast recursive algorithms for multiplying Ψ and Ψ^T . If H is also sparse then the chain is complete and so one can also multiply Σ quickly. Computing the conditional covariance is difficult and we suggest a Monte Carlo strategy that leverages an efficient algorithm for the conditional mean. Let \mathbf{u} be generated as $MN(0, \Sigma)$. Generate pseudo data vector $\mathbf{y}^* = K\mathbf{u} + \boldsymbol{\epsilon}$ and compute $\mathbf{u}^* = \mathbf{u} - \Sigma K^T (K \Sigma K^T + \sigma^2 I)^{-1} \mathbf{y}^*$. Simple matrix algebra will show that $\hat{\mathbf{z}} + \mathbf{u}^*$ will be a sample from the right 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 [8] 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. From a qualitative point of view this is the distinction of a multiresolution basis that facilitates representing nonstationary structure.

In this section we give a qualitative description of the multiresolution bases. One advantage of wavelet methods is the 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. We start with two templates, mother, ψ , and father ϕ wavelets both defined on $[0, 1]$ and a coarsest level of resolution, say J . The W transform wavelets used in this work are plotted in figure 1. The first J basis functions are similar to the father wavelet translated to J equally space locations. These are given in the first row 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. Accordingly, the next J basis functions are the mother wavelets translated in the same manner (second row of Figure 2). The next generation of basis functions has twice the resolution and twice as many members ($2J$ and is similar to a scaling and translation of the mother wavelet. The third row 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.

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

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

These are plotted in Figure 3.

Here S, H, V , and \mathcal{D} stand for smooth, horizontal, vertical and diagonal respectively and the letters H, V and \mathcal{D} correspond to their ability to represent features in these orientations.

The sequence of panels in Figure 4 are in analogy to the one dimensional basis presented above. One starts with a coarsest scale ($J = 4$) and obtains approximates translates of the father \times father function, S , at a 4×4 grid of locations. These functions are depicted as image plots in upper right hand matrix in Figure 4 where each separate image represents one two-dimensional basis function. The next 16 functions are similar

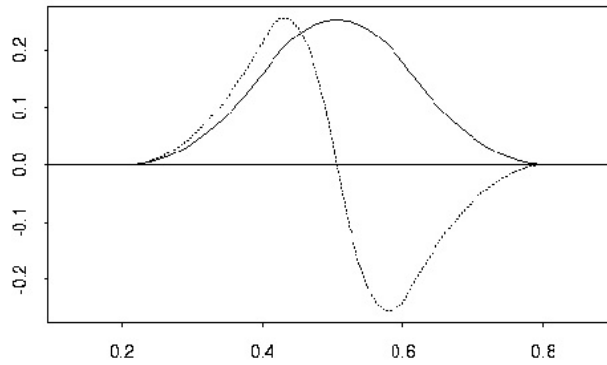


Figure 1: Continuous versions of the Father (dashed) and Mother (solid) W-transform wavelets

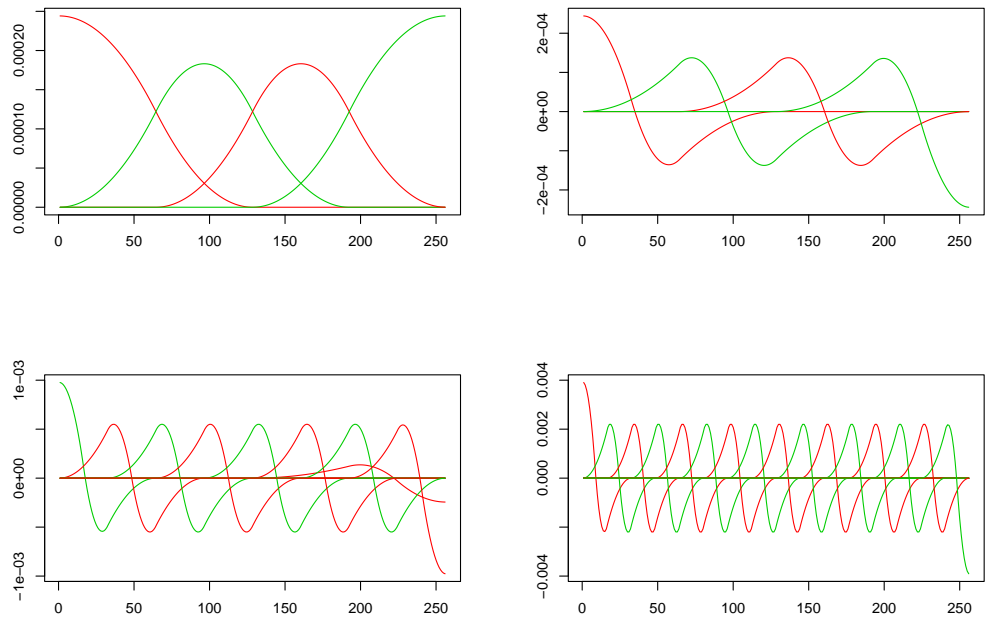


Figure 2: Family of 32 basis functions based on an approximate translation and scaling of the father and mother wavelets from Figure 1. Basis functions at the end points differ shape due to boundary corrections.

to translates of H (Figure 2) and the remaining panels of Figure 4 complete the first generation using V and D for a total of 64 functions. Subsequent generations are based only on the three template functions derived from the mother wavelet and in analogy to the one dimensional case involve scaling by a factor of two and translation. As an example, Figure 5 depicts the 64 functions approximating translations and scalings of H at the next level of resolution.

3.3 Discrete Wavelet transform

The discrete wavelet transform (DWT) is a fast computation method that can compute the coefficients of a basis that approximates the exact multiresolution definition. In the notation of the previous section, the DWT and its inverse can be used to rapidly multiply Ψ and Ψ^{-1} and their transposes by arbitrary vectors. The basis functions comprising the columns of Ψ 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 finite length separable, linear filters into four submatrices of smooth, horizontal, vertical and diagonal terms. The recursive form for the father and mother wavelet functions given below provides the linear filter weights that are used to transform the rows and columns of this image. The resulting block of $n_1/2 \times n_2/2$ smoothed coefficients now becomes the “image” for the next step. The other three blocks are the results for the H , V and D components at this level of resolution. The net result is a set of algorithms that are linear in the image size and are equivalent to multiplication of a vector by Ψ , it’s inverse, or it’s transpose. It should be noted that standard DWT refers to the specific multiplication of a vector by Ψ^{-1} and, for many wavelet choices, Ψ is an orthonormal matrix.

In this work we concentrate on a basis derived from the W transform. Kwong and Tang [6] proposed the father and mother wavelet functions based on simple families of filter weights in the discrete wavelet transform (DWT). The reason for this choice over more common wavelets is that they 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. What is common to most multiresolution bases is a recursive relationship that defines the father and mother wavelets and also subsequently the discrete transform. For example, the recursive equation that defines ϕ is the relation

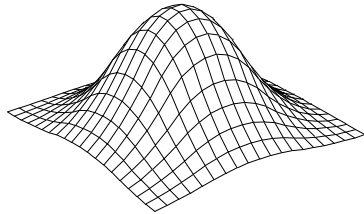
$$\phi(x) = \frac{1}{4}(\phi(2x) + 3\phi(2x - 1) + 3\phi(2x - 2) + \phi(2x - 3))$$

and

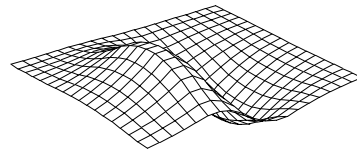
$$\psi(x) = C(\phi(2x) + 3\phi(2x - 1) - 3\phi(2x - 2) - \phi(2x - 3))$$

where C is a constant. The reader is referred to [6] for a development of the multiresolution basis that starts with the DWT. The W-wavelet is implemented in the Fields package [9] for the Splus and R statistical environments.

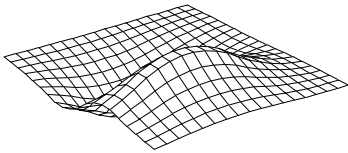
(a)



(b)



(c)



(d)

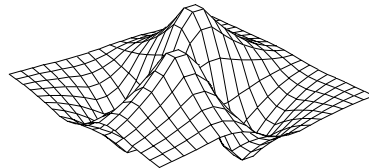


Figure 3: Two dimensions: tensor products of continuous versions of father and mother associated with the W-transform

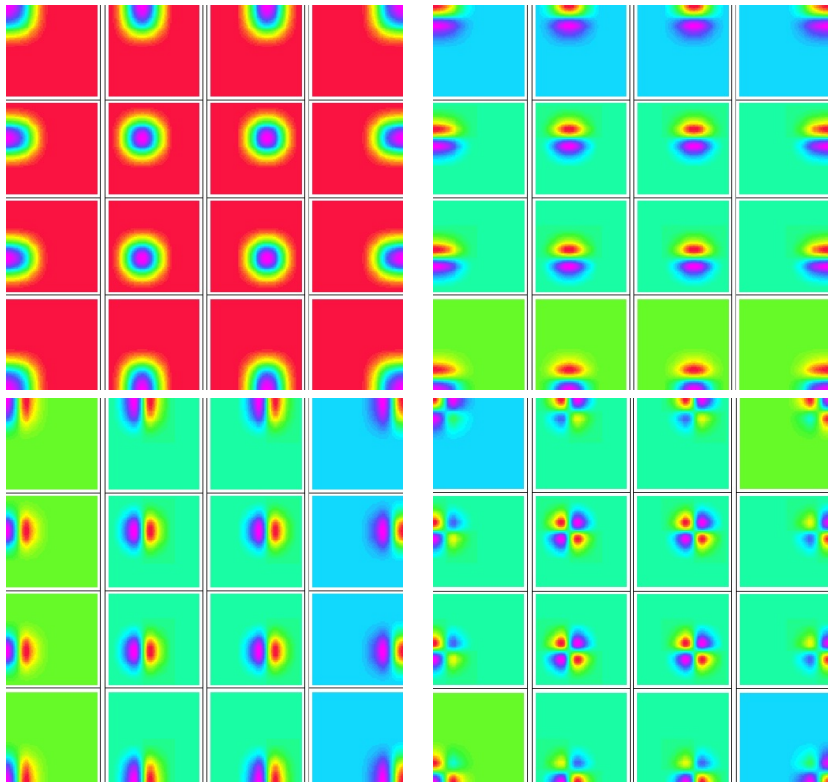


Figure 4: First generation of 64 two dimensional basis functions that approximate translations of the four template functions from Figure

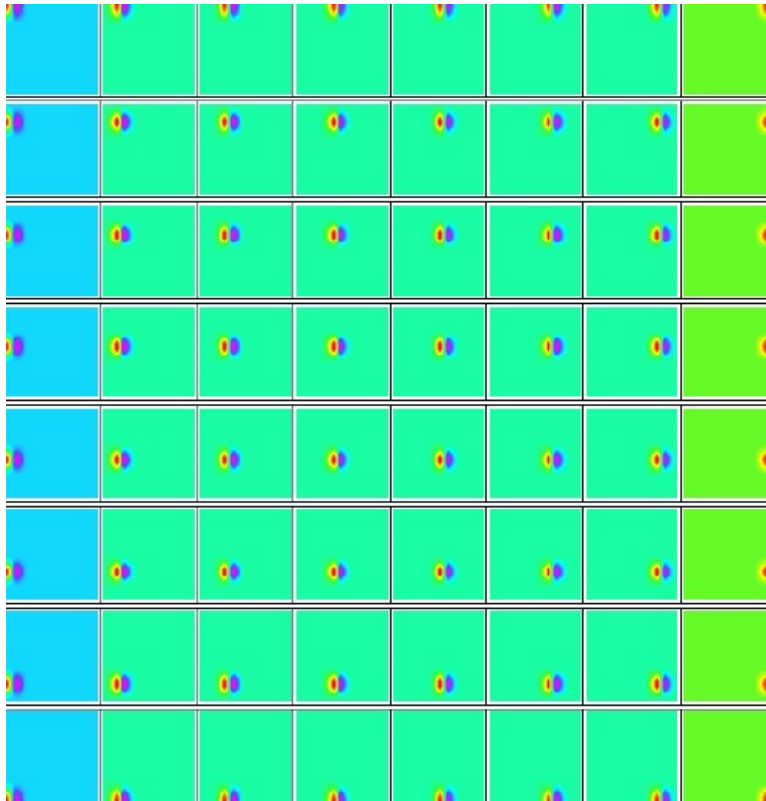


Figure 5: Second generation of 64 two dimensional basis functions based on the mother \times father template. These functions have more resolution in the y coordinate. There are 2 other sets of 64 basis functions in this generation corresponding to horizontal (father \times mother) and diagonal (mother \times mother) resolutions.

Given the DWT and its inverse for a particular wavelet family it is straight forward to calculate the implied basis functions. One simply creates a unit vector, e with a one at the index of interest and evaluates Ψe . In fact this is how the basis functions displayed in Figures 3 through 5 were found.

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, D among coefficients without it losing the approximation properties.

Suppose Σ is the covariance matrix for a spatial process then it is always true that

$$\Sigma = \Psi D \Psi^T = \Psi H^2 \Psi^T \quad (5)$$

for positive definite, D and square root H . Also

$$D = \Psi^{-1} \Sigma (\Psi^T)^{-1} \quad (6)$$

However, this decomposition is only useful if we can find an H that is close to diagonal.

As an example Figure 6 illustrates the structure of these matrices for a one dimensional grid (128 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 Σ , (b) is a scaled version of $\log_{10}(|D|)$ and (c) the leading 32×32 submatrix of D . The image (d) is the corresponding submatrix for H . The main features are clear, the elements of D fall off rapidly with scale and many off diagonal elements are zero. Most of the significant off diagonal elements are associated with the coarsest scale basis function. Moreover, The sparseness is amplified by considering H . This example is motivation for keeping the dominating elements of H but setting to zero small elements.

Another way to interpret the ability of the wavelet basis to nearly diagonalize Σ is to examine a transformed basis that produces an exact diagonalization. Figure 7 plots the normalized columns of the matrix ΨH in a similar format to Figure 2. These new basis functions are only a minor modifications of the original wavelets and thus also illustrate effective sparsity of H .

4.1 Enforcing sparsity in H

In the spirit of threshold estimators used in wavelet regression and the initial ideas from [1] we propose a thresholding to enforce sparsity in H . Based on results that will be given in the next section there are three parts stages to this strategy

- Retain all diagonal elements of H
- For the finest levels of resolution set all off diagonal elements of H to zero.

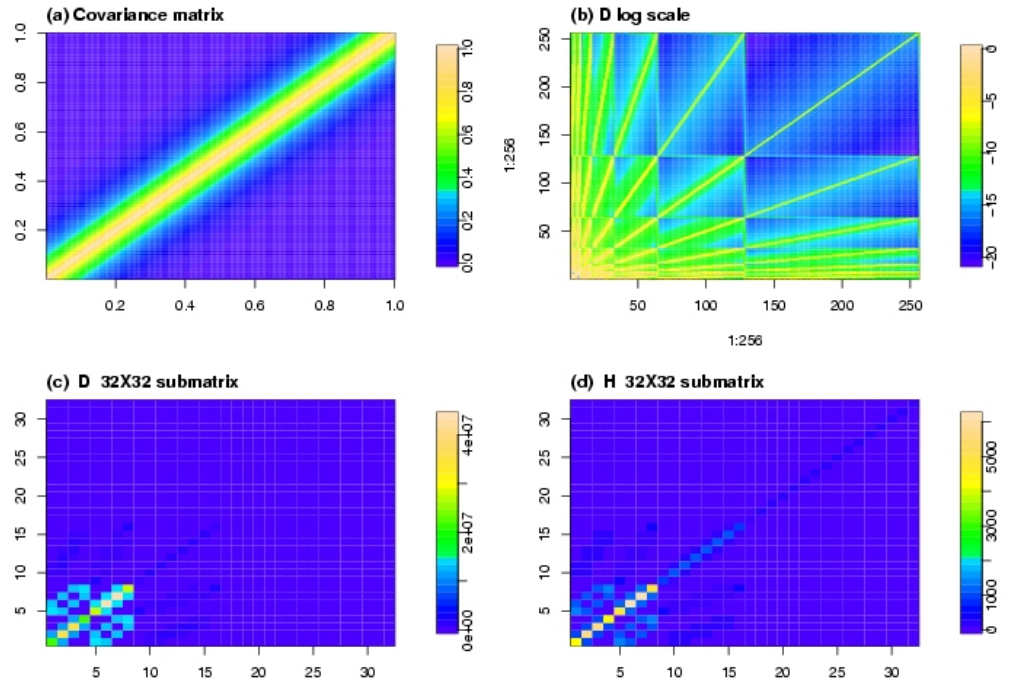


Figure 6: Approximation to a 1-D stationary covariance function. Image (a) is a exponential covariance matrix with range $\theta = 1/8$ over the interval $[0, 1]$ evaluated at 128 points. Image (b) is the matrix D (see equation 6 in a log relative scale, the elements are: $\log_{10}(|D|/\max|D|)$). Basis functions are ordered according the presentation in Figure 2 from coarsest scales to finest scales. Image (c) is an enlargement of the leading 32×32 submatrix of D and (d) is the corresponding H submatrix.

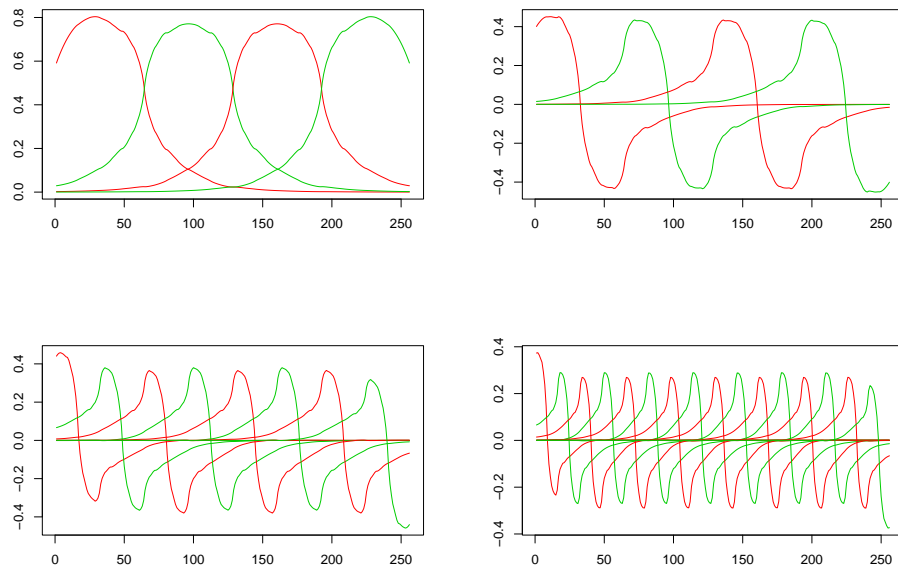


Figure 7: The transformed basis implied by a 1-D stationary covariance function.

- For the remaining elements, given $j \neq i$ and a threshold τ .

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

Figure 8 reports results for approximating for the exponential covariance model used above. Here we have chosen τ to be the 98% and 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 but there are some inaccuracies in the peak height and some ringing for the larger amount of decimation.

As a final illustration we consider a one dimensional nonstationary model obtained by deformation. We apply a logistic-like deformation of the interval $[0, 1]$ plotted in Figure 4.1 to the stationary exponential covariance used previously. This results in the nonstationary covariance matrix 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 reconstruct to assess the approximation (Figure 10). 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 two 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. 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 ν (smoothness) and θ (range) and has the form

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

where

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

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

For numerical results we consider a 16×16 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 4×4 grid. The covariance models were generated according to the smoothness parameters .5, 1.5 and 4 and with range parameters .125, .25, .5 and .75 giving a total of 12 different covariance functions. The wavelet approximation was computed for 98% and 97% decimation and the results are

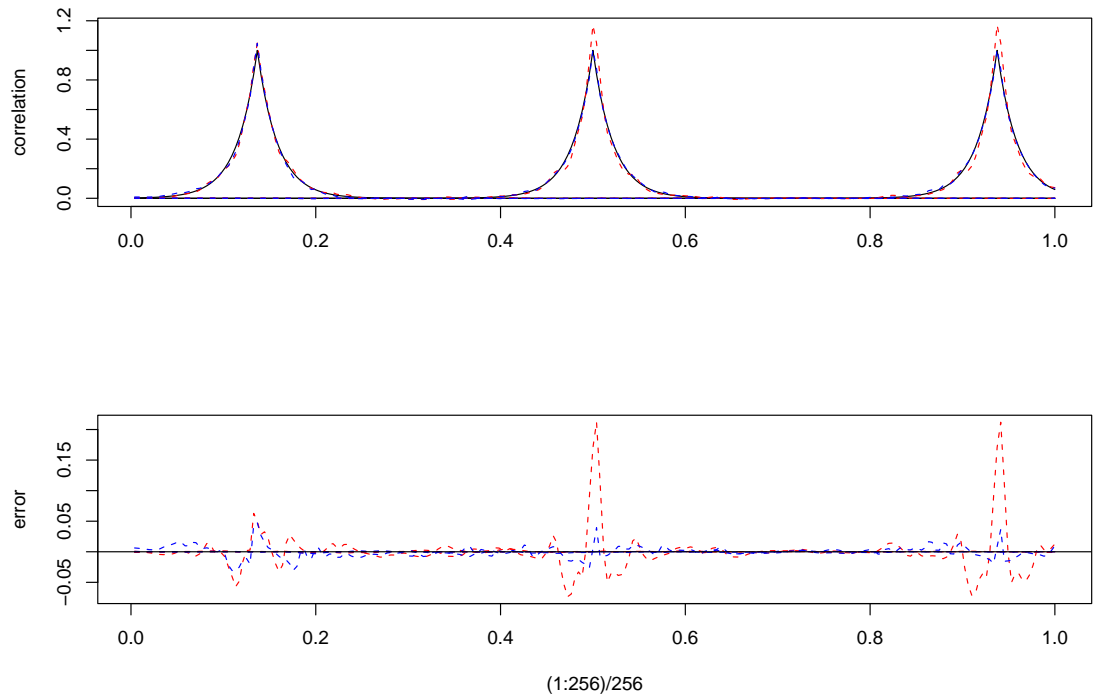


Figure 8: Approximation of a 1-d stationary covariance by wavelets at 97 % and 98 % decimation of H . The exponential covariance matrix with $\theta = 1/8$ is approximated by decimating the elements of H by 97 % (blue) 98 % (red) and then reforming the covariance matrix. Three rows of the true (solid)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

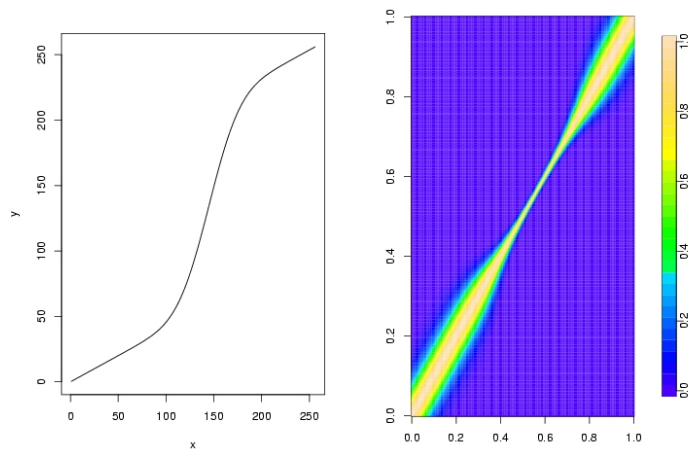


Figure 9: A nonstationary covariance matrix derived from a nonlinear spatial deformation

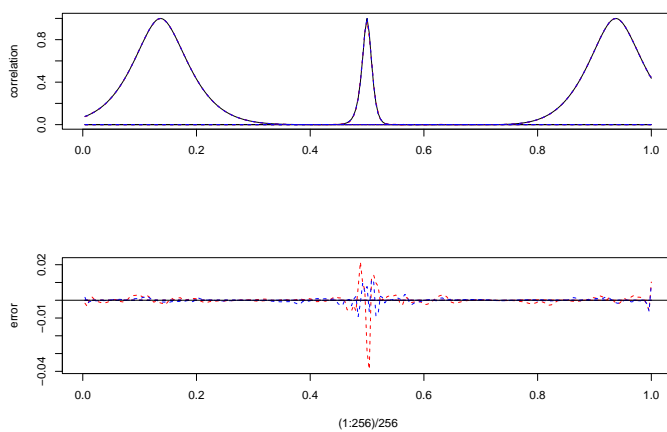


Figure 10: True and approximate covariances based on 97 % and 98 % decimation.

summarized in Figure 11. Here each boxplot is based on 256 values and summarizes the root mean squared errors (RMSE) for each row of the difference between the actual covariance matrix and the wavelet approximation. Large values tend to be located at the edges of the domain. Keeping in mind that the covariance models have marginal variance of 1 these RMSE are small relative to the size of the correlations. We see some degradation in the approximation for small ranges and the exponential model. This behavior may be related to the choice of 4×4 father wavelets setting the coarsest scale and also because the shape of the W-wavelets are closer to Gaussian than exponential at their peak.

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 case with 98% decimation the intersection 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 (approximate) 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 avoid intensive nonlinear optimization and scale to large problems. In general, fitting nonstationary covariance models from a single field, or small number of, fields, observed at irregular points is a difficult problem. We give a possible strategy in the discussion for irregularly spaced observations.

5.1 Sample estimates of H

Assume that independent copies of the field are observed over T time points and let \mathbf{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. Accordingly from the relationship at 1 each column of \mathbf{Z} has covariance Σ . With gridded data and (independent) replications over time, one can get sample estimates of the elements of H . We expect that for most problems $m \gg T$ and so the basic idea is to try to work with matrices of order $m \times T$ instead of $m \times m$. We have the sample covariance as

$$\hat{\Sigma} = (1/T)\mathbf{Z}\mathbf{Z}^T$$

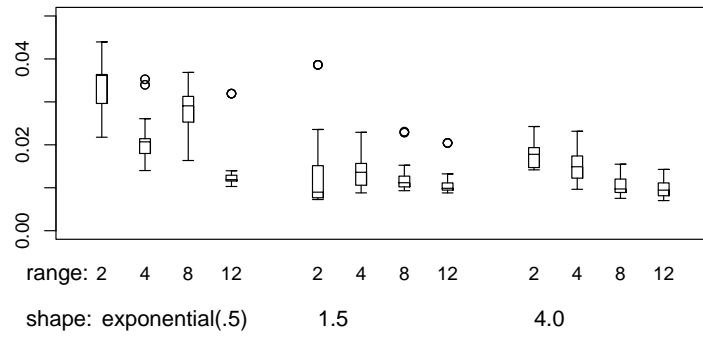
and so

$$\hat{D} = (1/T)(\Psi^{-1}\mathbf{Z})(\Psi^{-1}\mathbf{Z})^T$$

This form motivates estimating H via the singular value decomposition of $(\Psi^{-1}\mathbf{Z})$. Let $V\Lambda U^T = (\Psi^{-1}\mathbf{Z})$ where V and U are orthogonal matrices and Λ a $T \times T$ diagonal matrix of nonnegative singular values. Setting

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

(a) truncation 2%



(b) truncation 3%

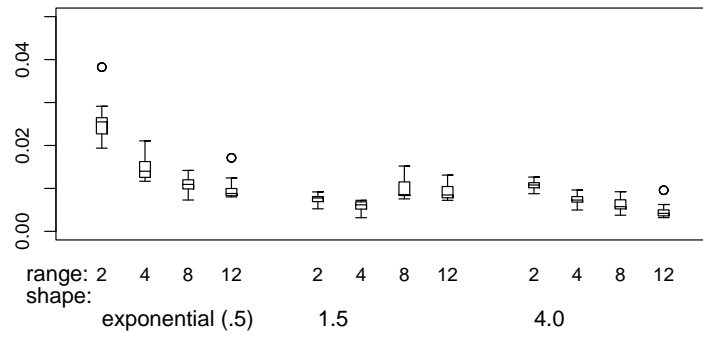


Figure 11: Comparison of the wavelet approximation to the family of Matern covariances. The boxplots are the RMSE for each row of the difference between the true and approximate correlation functions.

it now follows that $\hat{D} = \hat{H}^2$. 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 of and the entires that are nonzero can be computed efficiently based on V and $\Lambda^{1/2}$. Note that here also we are only working with matrices of order $m \times T$ not $m \times m$.

Once the elements of \hat{H} are found one can:

- Further decimate them
- 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 important for large problems.

6 Example with Surface Ozone

We use 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 and simulates ozone formation and transport based on sources of pollution and meteorological conditions. The data used in this study is a 48×48 grid centered on Illinois and Ohio where each grid box is approximately 16×16 miles in size. Output is available for 79 days using meteorology from in June-August 1987, this was a period of high summer ozone. For back ground on this model and its comparison to data see [?].

Ambient average daily ozone has a correlation range of the order of 300 miles and so it is reasonable to start with a 3×3 grid of father wavelets as the coarsest level of resolution. This division is 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 12×12 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 but smoothly varying and is illustrated in Figure 12. Each member of the panel fixes a location and plots the covariance or correlation surface with that point. In terms of matrix notation we plot four separate rows of the estimated matrix Σ having first reformed the row vectors as images. The top left hand location (a) is a rural area of Illinois and has a covariance that is long ranged to the west but with low variability. The location in (d) 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. This might be due to the general motion of weather systems, and the corresponding transport of ozone, from west to east.

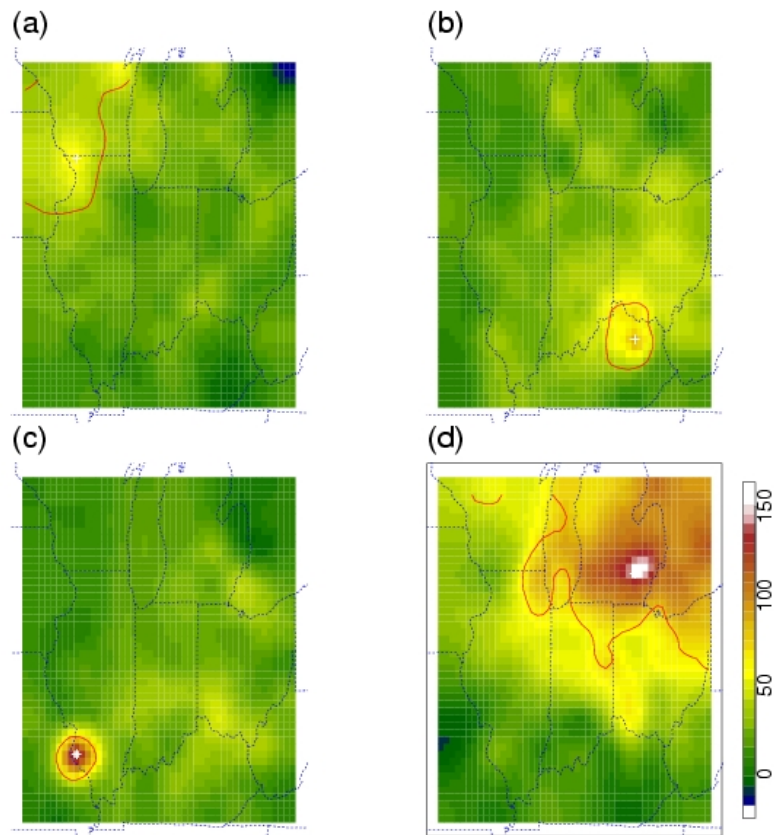


Figure 12: Estimated covariance surface at 4 sample locations. The image plots indicate the covariance between points in the domain and the point location denoted by an +. Contours indicate the locus of points where the correlation is at .5.

7 Discussion

We have shown that wavelets provide flexible methods for introducing nonstationary spatial structure and can also reproduce standard spatial models. Due to 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 with using ROM output is surprising in how a regular basis can produce nonstationary but smooth covariance patterns. A key extension of this methods 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 it one can not take advantage of the DWT for matrix multiplication of the basis functions. Also there are problems with many elements of H simply not being identifiable. A formal approach 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. We think this approach may be productive for moderate size problems but will not be easy to implement for many important large geophysical data sets. Here we suggest an algorithmic 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. It may be useful translate this model into the wavelet approximation format to facilitate computation. Given this starting covariance model at each time point one samples the conditional distribution of the field on a regular grid given the observed data. Given these gridded samples of the field one now fits a model to H based on the ideas from this paper. At this point because there is now “data” on a complete grid, the DWT can be used for matrix multiplications of Ψ . A further elaboration of this strategy is to now use the multiresolution based covariance to extend the data to a grid and reestimate the model for H . 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 enforcing the starting covariance model in regions where direct covariance information is limited.

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

Acknowledgments

This work was supported by National Science Foundation grants DMS-93122686 and DMS-9815344.

References

- [1] Davis, J. M., Nychka, D. and Bailey, B. (2000). "A Comparison of the Regional Oxidant Model with Observed Data." *Atmospheric Environment*, 34, 2413-2423.
- [2] Donoho, D., Mallat, S. and von Sachs, R.(1996) "Estimating covariances of locally stationary processes", In *Proceedings of IEEE Time frequency and Time-Scale Symposium*, Paris, July 1996. IEEE, New York
- [3] 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.
- [4] Golub, G. and Van Loan, C. (1989). *Matrix Computations*. John Hopkins University Press. Baltimore.
- [5] 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.
- [6] 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.
- [7] 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*
- [8] Mallat, S. Papanicolaou, G. and Zhang Z. (1998). "Adaptive covariance estimation of locally stationary processes." *The Annals of Statistics* **26**, 1-47.
- [9] Nason, G.P. and Silverman, B.W. (1994). The discrete wavelet transform. *Journal of Computational and Graphical Statistics*, **3**, 163-191.
- [10] Nychka, D., Meiring, W., Royle, J. A., Fuentes, M. and Gilleland, E. (2000). FIELDS: S Tools for Spatial Data <http://www.cgd.ucar.edu/stats/software>
- [11] Sampson, P. and Guttorp, P. (1992), "Nonparametric estimation of nonstationary spatial covariance structure", *Journal of the American Statistical Association*, **87**, 108-119.
- [12] Wikle, C.K. and N. Cressie (1999) "A dimension reduced approach to space-time Kalman filtering" . *Biometrika* , **86** , 815-829.