Using transforms to analyze space-time processes

Peter Guttorp Montserrat Fuentes Paul Sampson

NRCSE

Technical Report Series

NRCSE-TRS No. 080

March 11, 2005

Using transforms to analyze space-time processes[1]

Peter Guttorp, Montserrat Fuentes and Paul Sampson

INTRODUCTION

Transform tools, such as power spectra, wavelets, and empirical orthogonal functions, are useful tools for analyzing temporal, spatial and space-time processes. In this chapter we will develop some theory and illustrate its use in a variety of applications in ecology and air quality.

Harmonic (or frequency) analysis has long been the one of the main tools in time series analysis. Application of Fourier techniques work best when the underlying process is stationary, and we develop and illustrate it here for stationary spatial and space-time processes. However, spatial stationarity is rather a severe assumption for air quality models, and we show how the theory can be generalized beyond that assumption. In some circumstances we can develop formal tests for stationarity.

Wavelets are another set of transform tools that recently have found much use in time series analysis. We illustrate here how wavelets can be used to estimate temporal trends, and to develop different models of nonstationary spatial processes.

In geophysics and meteorology, variants of principal components called empirical orthogonal functions (EOFs) have long been used to describe leading modes of variability in space-time processes. Here we use smoothed EOFs to model the spatio-temporal mean field of a random field, while yet a third type of spatially nonstationary model is used to describe the random part of the field.

Finally, we return to the frequency analysis of space-time processes, and describe a spectral representation and a parametric class of space-time covariances. We also show how to test for separability of a space-time covariance, another frequently made but rather severe assumption in the types of applications we discuss.

1 Spectral analysis of spatial processes

Spectral methods are a powerful tool for studying the spatial structure of random fields and generally offer significant computational benefits. In Section 1 of this chapter we offer a review of the Fourier transform and its properties, we introduce the spectral representation of a stationary spatial process, and we also present Bochner's theorem to obtain a spectral representation for the covariance. We later describe some commonly used classes of spectral densities, and we introduce

---

[1] P. Guttorp is Professor at the Statistics Department, University of Washington, Seattle. M. Fuentes is an Associate Professor at the Statistics Department, North Carolina State University (NCSU), Raleigh, NC 27695-8203, and a visiting scientist at the US Environmental Protection Agency (EPA). Tel.:(919) 515-1921, Fax: (919) 515-1169, E-mail: fuentes@stat.ncsu.edu. P. Sampson is Research Professor at the Statistics Department, University of Washington, Seattle. This research was sponsored by a National Science Foundation grant DMS 0353029, and by the Swedish Association of Graduate Engineers. Professor Guttorp is grateful to the Department of Mathematical Statistics at the University of Lund for their hospitality during this research. We want to thank Peter Craigmile, Ohio State University, for letting us use his analysis of the Malindi data set.

Key words: Fourier transform, nonstationary processes, nonseparable covariances, periodogram, spatial statistics, tapering, wavelets.

the periodogram, a nonparametric estimate of the spectrum, and we study its properties. We also present an approximation to the Gaussian likelihood using spectral methods.

1.1 Fourier Analysis.

In this section we present a review of the Fourier transform and its properties, we also discuss the aliasing phenomenon in the spectral domain. This aliasing effect is a result of the lost of information when we take a discrete set of observations on a continuous process.

1.1.1 Continuous Fourier transform

A Fourier analysis of a spatial process, also called a harmonic analysis, is a decomposition of the process into sinusoidal components (sines and cosines waves). The coefficients of these sinusoidal components are the Fourier transform of the process.

Supposed that $g()$ is a real or complex-valued function,

Define

$$G(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} g(\mathbf{s}) \exp\{i\boldsymbol{\omega}^t \mathbf{s}\} d\mathbf{s} \tag{1}$$

The function G in (1) is said to be the Fourier transform of g. Then, g has the representation

$$g(\mathbf{s}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} G(\boldsymbol{\omega}) \exp\{-i\boldsymbol{\omega}^t \mathbf{s}\} d\boldsymbol{\omega} \tag{2}$$

so that $|G(\boldsymbol{\omega})|$ represents the amplitude associated with the complex exponential with frequency $\boldsymbol{\omega}$. The right-hand side of Equation (2) is called the Fourier integral representation of g. The function g and G are said to be a Fourier transform pair.

The Fourier transform of $g(\mathbf{s})$ is often also defined as

$$G(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} g(\mathbf{s}) \exp\{i2\pi\boldsymbol{\omega}^t \mathbf{s}\} d\mathbf{s} \tag{3}$$

and the inverse Fourier transform of $G(\boldsymbol{\omega})$ as

$$g(\mathbf{s}) = \int_{\mathbb{R}^d} G(\boldsymbol{\omega}) \exp\{-i2\pi\boldsymbol{\omega}^t \mathbf{s}\} d\boldsymbol{\omega} \tag{4}$$

It is often useful to think of functions and their transforms as occupying two domains. These domains are referred to as the upper and the lower domains in older texts, “as if functions circulated at ground level and their transforms in the underworld” (Bracewell, 1999). They are also referred to as the function and transform domains, but in most physics applications they are called the time (or space) and frequency domains respectively. Operations performed in one domain have corresponding operations in the other. For example, the convolution operation in the time (space) domain becomes a multiplication operation in the frequency domain, that is, $f(x) * g(x) \rightarrow F(s)G(s)$. The reverse is also true, as we will see in the next section of this chapter, $F(s) * G(s) \rightarrow f(x)g(x)$. Such theorems allow one to move between domains so that operations can be performed where they are easiest or most advantageous.

1.1.2 Properties of the Fourier Transform

Scaling Property

If $\mathcal{F}\{f(x)\} = F(s)$ is a real, nonzero constant function, where $\mathcal{F}$ denotes the Fourier transform, then

$$\begin{aligned}\mathcal{F}\{f(ax)\} &= \int_{-\infty}^{\infty} f(ax)\exp(i\pi 2sx)dx \\ &= \frac{1}{|a|}\int_{-\infty}^{\infty} f(\beta)\exp(i\pi\beta 2s/a)d\beta \\ &= \frac{1}{|a|}F(s/a).\end{aligned} \tag{5}$$

From this, the scaling property, it is evident that if the width of a function is decreased while its height is kept constant, then its Fourier transform becomes wider and shorter. If its width is increased, its transform becomes narrower and taller.

A similar frequency scaling property is given by

$$\mathcal{F}\left\{\frac{1}{|a|}f(x/a)\right\} = F(as).$$

Shifting Property

If $\mathcal{F}\{f(x)\} = F(s)$ and x_0 is a real constant, then

$$\begin{aligned}\mathcal{F}\{f(x-x_0)\} &= \int_{-\infty}^{\infty} f(x-x_0)\exp(i2\pi sx)dx \\ &= \int_{-\infty}^{\infty} f(\beta)\exp(i2\pi s(\beta+x_0))d\beta \\ &= \exp(i\pi 2x_0 s)\int_{-\infty}^{\infty} f(\beta)exp(i\pi 2s\beta)d\beta \\ &= F(s)\exp(i2\pi x_0 s).\end{aligned} \tag{6}$$

This shifting property states that the Fourier transform of a shifted function is just the transform of the unshifted function multiplied by an exponential factor having a linear phase.

Likewise, the frequency shifting property states that if $F(s)$ is shifted by a constant s_0, its inverse transform is multiplied by $\exp(-i\pi 2xs_0)$

$$\mathcal{F}\{f(x)\exp(-i2\pi xs_0)\} = F(s-s_0).$$

Convolution Theorem

We now derive the previously mentioned convolution theorem. Suppose that $g(x) = f(x) * h(x)$.

Then, given that $\mathcal{F}\{g(x)\} = G(s)$, $\mathcal{F}\{f(x)\} = F(s)$, and $\mathcal{F}\{h(x)\} = H(s)$,

$$\begin{aligned}
G(s) &= \mathcal{F}\{f(x) * h(x)\} \\
&= \mathcal{F}\left\{\int_{-\infty}^{\infty} f(\beta)h(x-\beta)d\beta\right\} \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(\beta)h(x-\beta)d\beta\right]\exp(i\pi 2sx)dx \\
&= \int_{-\infty}^{\infty} f(\beta)\left[\int_{-\infty}^{\infty} h(x-\beta)\exp(i\pi 2sx)dx\right]d\beta \\
&= H(s)\int_{-\infty}^{\infty} f(\beta)\exp(i\pi\beta 2s)d\beta \\
&= F(s)H(s).
\end{aligned} \tag{7}$$

This extremely powerful result demonstrates that the Fourier transform of a convolution is simply given by the product of the individual transforms, that is

$$\mathcal{F}\{f(x) * h(x)\} = F(s)H(s).$$

Using a similar derivation, it can be shown that the Fourier transform of a product is given by the convolution of the individual transforms, that is

$$\mathcal{F}\{f(x)h(x)\} = F(s) * H(s)$$

Parseval's Theorem

Parseval's Theorem states that the power of a signal represented by a function h(t) is the same whether computed in signal space or frequency (transform) space; that is,

$$\int_{-\infty}^{\infty} h^2(t)dt = \int_{-\infty}^{\infty} |H(f)|^2 df$$

(see Bracewell, 2000). The power spectrum, $P(f)$, is given by

$$P(f) = |H(f)|^2,$$

for $-\infty \leq f \leq +\infty$.

1.1.3 Aliasing

If we decompose a continuous process Z into a discrete superposition of harmonic oscillations, it is easy to see that such a decomposition cannot be uniquely restored from observations of Z in $\Delta\mathbb{Z}^2$ (this an infinite lattice with spacing Δ), where Δ is the distance between neighboring observations, and $\mathbb{Z}^2$ the integer lattice. The equal spacing in the space domain of the observations introduces an *aliasing* effect for the frequencies. Indeed,

$$\exp\{i\boldsymbol{\omega}\mathbf{x}\Delta\} = \exp\{i(\boldsymbol{\omega} + \mathbf{y}\mathbf{2}\boldsymbol{\pi}/\Delta)\mathbf{x}\Delta\} = \exp\{i\boldsymbol{\omega}\mathbf{x}\Delta\}\exp\{i2\pi\mathbf{y}\mathbf{x}\}$$

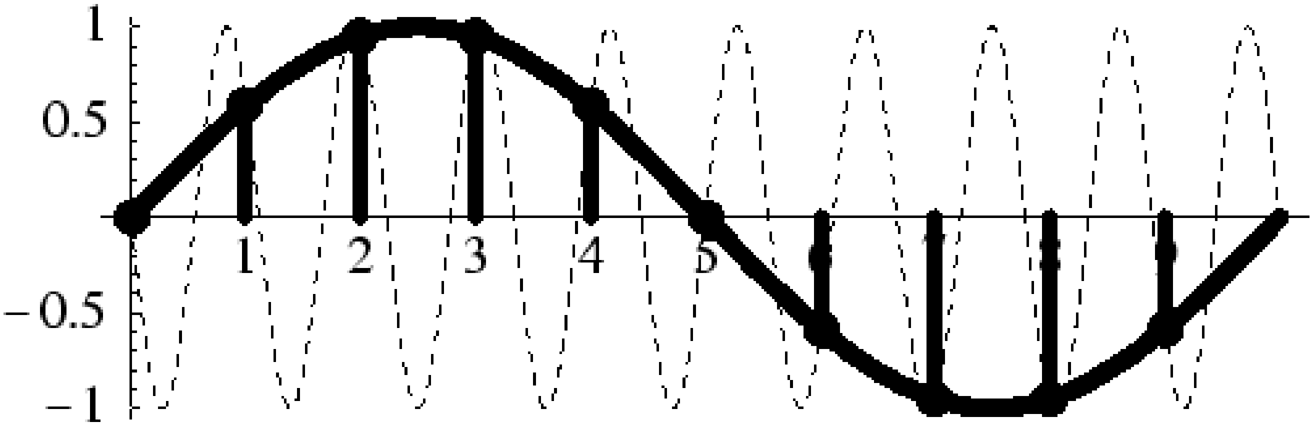


Figure 1: Example of aliasing due to undersampling in space.

for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{Z}^2$. We simply cannot distinguish an oscillation with an angular frequency $\boldsymbol{\omega}$ from all the oscillations with frequencies $\boldsymbol{\omega} + 2\pi\mathbf{y}/\Delta$. The frequencies $\boldsymbol{\omega}$ and $\boldsymbol{\omega}' = \boldsymbol{\omega} + 2\pi\mathbf{y}/\Delta$ are indistinguishable and hence are aliases of each other. The impossibility of distinguishing the harmonic components with frequencies differing by an integer multiple of $2\pi/\Delta$ by observations in the integer lattice with spacing Δ is called *aliasing* effect.

Then, if observation of a continuous process Z is carried out only at uniformly spaced spatial locations Δ units apart, the spectrum of observations of the sample sequence $Z(\Delta\mathbf{x})$, is concentrated within the finite frequency band $-\boldsymbol{\pi}/\Delta \leq \boldsymbol{\omega} < \boldsymbol{\pi}/\Delta$. Every frequency not in that interval band has an alias in the band, termed its principal alias. The whole frequency spectrum is partitioned into bands of length $2\pi/\Delta$ by *fold points* $(2\mathbf{y}+1)\pi/\Delta$, with $\mathbf{y} \in \mathbb{Z}^2$, and the power distribution within each of the bands distinct from the principal band $-\boldsymbol{\pi}/\Delta \leq \boldsymbol{\omega} < \boldsymbol{\pi}/\Delta$, is superimposed on the power distribution within the principal band. Thus, if we wish that the spectral characteristics of the process Z to be determined accurately enough from the observed sample, then the *Nyquist frequency* π/Δ must necessarily be so high that still higher frequencies $\boldsymbol{\omega}$ make only a negligible contribution to the total power of the process. This means that we observe a dense sample of Z (small Δ). The Nyquist frequency is also called the *folding frequency*, since higher frequencies are effectively folded down into the band $-\boldsymbol{\pi}/\Delta \leq \boldsymbol{\omega} < \boldsymbol{\pi}/\Delta$.

It should be noted that aliasing is a relatively simple phenomenon. In general, when one takes a discrete set of observations on a continuous function, information is lost. It is an advantage of the trigonometric functions that this loss of information is manifest in the easily understood form of aliasing. Figure 1 shows an example of aliasing. In the figure, the high-frequency sinusoid is indistinguishable from the lower frequency sinusoid due to aliasing. We say the higher frequency aliases to the lower frequency. Undersampling in the frequency domain gives rise to space-domain aliasing.

1.2 Spectral representation

In this Section we introduce the spectral representation of a stationary spatial process using sine and cosine waves. We also present Bochner's theorem to obtain a spectral representation for the covariance.

1.2.1 The spectral representation theorem

Consider a weakly stationary process Z with mean 0 and covariance C. Before we can apply the ideas of Fourier series and Fourier integrals we must first ask: Can we represent a typical realization as a Fourier series? The answer to this question is clearly “No”, since we have no reason to suppose that a realization of a general stationary process will be periodic in any way.

The next question is: Can we represent a typical realization as a Fourier integral? “No”. Maybe we can not distribute the power over a continuous range of frequencies, but over a set of frequencies with discontinuities. This will lead to a Fourier type integral called Fourier-Stieltjes.

$$Z(\mathbf{s}) = \int e^{is\boldsymbol{\omega}^T} dY(\boldsymbol{\omega}),$$

where Y measures the average contributions from all components with frequencies less than or equal to $\boldsymbol{\omega}$.

The spectral representation theorem

To every stationary $Z(\mathbf{s})$ there can be assigned a process $Y(\boldsymbol{\omega})$ with orthogonal increments, such that we have for each fixed $\mathbf{s}$ the spectral representation:

$$Z(\mathbf{s}) = \int_{\mathbb{R}^2} e^{i\mathbf{s}^T\boldsymbol{\omega}} dY(\boldsymbol{\omega}) \tag{8}$$

$Y(\boldsymbol{\omega})$ is defined up to an additive random variable. The Y process is called the spectral process associated with a stationary process Z. The random spectral process Y has the following properties:

$$E(Y(\boldsymbol{\omega})) = 0$$

(since mean of Z is 0), the process Y has orthogonal increments:

$$E[(Y(\boldsymbol{\omega}_3) - Y(\boldsymbol{\omega}_2))(Y(\boldsymbol{\omega}_1) - Y(\boldsymbol{\omega}_0))] = 0,$$

when $(\boldsymbol{\omega}_3, \boldsymbol{\omega}_2)$ and $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_0)$ are disjoint intervals. If we define F as

$$E|dY(\boldsymbol{\omega})|^2] = dF(\boldsymbol{\omega})$$

F is a positive measure, and

$$E[dY(\boldsymbol{\omega})dY(\boldsymbol{\omega}')] = \delta(\boldsymbol{\omega} - \boldsymbol{\omega}')dF(\boldsymbol{\omega})d\boldsymbol{\omega}'$$

where $\delta(\boldsymbol{\omega})$ is the Dirac δ-function.

The spectral representation theorem may be proved by various methods; using Hilbert space theory, or by means of trigonometric integrals. A good reference is Cramer and Leadbetter (1967).

Bochner’s theorem

We derive the spectral representation of the autocovariance function C :

$$C(\mathbf{s}) = \int_{\mathbb{R}^2} e^{i\mathbf{s}^T\boldsymbol{\omega}} dF(\boldsymbol{\omega})$$

Bochner's theorem says that a function C is nonnegative definite if and only if it can be represented in the form above where F is real, never-decreasing, and bounded. Thus, the spatial structure of Z could be analyzed with a spectral approach or equivalently by estimating the autocovariance function, see Cramer and Leadbetter (1967) for more detail.

If we compare the spectral representation of $C(\mathbf{s})$ and $Z(\mathbf{s})$

$$C(\mathbf{s}) = \int_{\mathbb{R}^2} e^{i\mathbf{s}^T\boldsymbol{\omega}} dF(\boldsymbol{\omega})$$

$$Z(\mathbf{s}) = \int_{\mathbb{R}^2} e^{i\mathbf{s}^T\boldsymbol{\omega}} dY(\boldsymbol{\omega})$$

it will be seen that the elementary harmonic oscillations are respectively $e^{i\mathbf{s}^T\boldsymbol{\omega}}dF(\boldsymbol{\omega})$, $e^{i\mathbf{s}^T\boldsymbol{\omega}}dY(\boldsymbol{\omega})$

If we think of $Y(\boldsymbol{\omega})$ as representing the spatial development of some concrete physical systems, The spectral representation gives the decomposition of the total fluctuation in its elementary harmonic components. The spectral d.f. $F(\boldsymbol{\omega})$ determines the distribution of the total average power in the $Z(\mathbf{s})$ fluctuation over the range of angular frequency $\boldsymbol{\omega}$. The average power assigned to the frequency interval $A = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2]^2$ is $F(A)$, which for the whole infinite $\boldsymbol{\omega}$ range becomes

$$E|Z(\mathbf{s})|^2 = C(\mathbf{0}) = F(\mathbb{R}^2)$$

Thus, F determines the power spectrum of the Z process. We may think of this as a distribution of a spectral mass of total amount $C(\mathbf{0})$ over the $\boldsymbol{\omega}$ axis. F only differs by a multiplicative constant from an ordinary d.f.

If F has a density with respect to Lebesgue measure, this density is the spectral density, $f = F'$, defined as the Fourier transform of the autocovariance function:

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \exp(-i\boldsymbol{\omega}^T\mathbf{x})C(\mathbf{x})d\mathbf{x}.$$

Spectral moments

The spectral moments

$$\lambda_k = \int \boldsymbol{\omega}^k dF(\boldsymbol{\omega})$$

may or may not be finite. As a consequence of Bochner's theorem, the moment λ_{2k} is finite if and only $C(\mathbf{x})$ has all partial derivatives of order $2k$ at the origin.

A process Z is m-times mean square differentiable if and only if $C^{(2m)}(0)$ exists and is finite and, if so, the autocovariance function of $Z^{(m)}$ is $(-1)^m C^{(2m)}$ (eg. Stein (1999), p. 21). Therefore, Z is m-times mean square differentiable if and only if the momment λ_{2m} is finite.

1.3 Some spectral densities

We describe in Section some commonly used classes of spectral densities. We consider in this section a real process, then the spectral density is an even function. We also assume that the covariance is isotropic, so that the spectral density is a function of a single frequency (Matérn, 1960).

1.3.1 Triangular model

For a spatial process with a triangular isotropic covariance:

$$C(\mathbf{x}) = \sigma(a - |\mathbf{x}|)^{+}$$

for σ and a positive, the corresponding spectral density is

$$f(\boldsymbol{\omega}) = \sigma\pi^{-1}\{1 - \cos(\alpha|\boldsymbol{\omega}|)\}/|\boldsymbol{\omega}|^2,$$

The oscillating behavior of the spectral density would be probably quite unrealistic for many physical processes. There is usually no reason to assume the spectrum has much more mass near the frequency $(2n + 1)\pi$ than near $2n\pi$ for n large, which is the case for the spectral density $\{1 - \cos(\{\boldsymbol{\omega})\}/|\boldsymbol{\omega}|^2$. Some kriging predictors under this model have strange properties as a consequence of the oscillations of the spectral density at high frequencies.

1.3.2 Spherical model

One of the most commonly used model for isotropic covariance functions in geological and hydrological applications is the spherical

$$C(\mathbf{x}) = \begin{cases} \sigma\left(1 - \frac{3}{2\rho} + \frac{1}{2\rho^3}|\mathbf{x}|^3\right) & r \le \rho \\ 0 & r > \rho \end{cases} \tag{9}$$

for positive constants σ and ρ. This function is not a valid covariance in higher dimensions than 3. The parameter ρ is called the range and is the distance at which correlations become exactly 0. This function is only once differentiable at $r = \rho$ and this can lead to problems when using likelihood methods for estimating the parameters of this model. In three dimensions, the corresponding isotropic spectral density has oscillations at high frequencies similar to the triangular covariance function in one dimension. Stein and Handcock (1989) show that when using the spherical model in three dimensions, certain prediction problems have rather pathological behavior.

1.3.3 Squared exponential model

The density of a a spatial process with an isotropic squared exponential covariance:

$$C(\mathbf{x}) = \sigma e^{-\alpha|\mathbf{x}|^2}$$

is

$$f(\boldsymbol{\omega}) = \frac{1}{2}\sigma(\pi\alpha)^{-1/2}e^{-\boldsymbol{\omega}^2/(4\alpha)}$$

Note that C and f both are the same type of exponential functions when $\gamma = 2$. The parameter σ is the variance of the process and α^{-1} is a parameter that explains how fast the correlation decays.

1.3.4 Matérn-Whittle class

A class of practical variograms and autocovariance functions for a process Z can be obtained from the Matérn class of spectral densities

$$f(\boldsymbol{\omega}) = \phi(\alpha^2 + |\boldsymbol{\omega}|^2)^{\left(-\nu-\frac{d}{2}\right)} \tag{10}$$

with parameters $\nu > 0$, $\alpha > 0$ and $\phi > 0$ (the value d is the dimension of the spatial process Z). Here, the vector of covariance parameters is $\boldsymbol{\theta} = (\phi, \nu, \alpha)$. The parameter α^{-1} can be interpreted as the autocorrelation range. The parameter ν measures the degree of smoothness of the process Z, the higher the value of ν the smoother Z would be, and ϕ is proportional to the ratio of the variance σ and the range (α^{-1}) to the $2\nu^{\text{th}}$ power, $\phi = \sigma\alpha^{2\nu}$.

The corresponding covariance for the Matérn class is

$$C_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{\pi^{d/2}\phi}{2^{\nu-1}\Gamma(\nu + d/2)\alpha^{2\nu}}(\alpha|\mathbf{x}|)^{\nu}\mathcal{K}_{\nu}(\alpha|\mathbf{x}|), \tag{11}$$

where $\mathcal{K}_{\nu}$ is a modified Bessel function. For instance when $\nu = \frac{1}{2}$, we get the exponential covariance function,

$$C_{\boldsymbol{\theta}}(\mathbf{x}) = \pi\phi\alpha^{-1}\exp(-\alpha|\mathbf{x}|).$$

When ν is of the form $m + \frac{1}{2}$ with m a nonnegative integer, the Matérn covariance function is of the form $e^{-\alpha|\mathbf{x}|}$ times a polynomial in $|\mathbf{x}|$ of degree m, (Abramowitz and Stegun (1965), Stein (1999) p. 31).

Handcock and Wallis (1994) suggested the following parametrization of the Matérn covariance that does not depend on d:

$$C(\mathbf{x}) = \frac{\sigma}{2^{\nu-1}\Gamma(\nu)\alpha^{2\nu}}(2\nu^{1/2}|\mathbf{x}|/\rho)^{\nu}\mathcal{K}_{\nu}(2\nu^{1/2}|\mathbf{x}|/\rho), \tag{12}$$

but the corresponding spectral density does depend on d:

$$f(\boldsymbol{\omega}) = \frac{\sigma g(\nu, \rho)}{(4\nu/\rho^2 + |\boldsymbol{\omega}|^2)^{\nu+d/2}}$$

where

$$g(\nu, \rho) = \frac{\Gamma(\nu + d/2)(4\nu)^{\nu}}{\pi^{d/2}\rho^{2\nu}\Gamma(\nu)}$$

with $\sigma = \operatorname{var}(Z(\mathbf{s}))$, the parameter ρ measures how the correlation decays with distance, and generally this parameter is called the *range*. The parameter α^{-1} has a very similar interpretation to ρ. But both parameters have different asymptotic properties under an infill asymptotic model. If we consider the limit as $\nu \to \infty$ we get the squared exponential covariance

$$K(\mathbf{x}) = \sigma e^{-|\mathbf{x}|^2/\rho^2}.$$

The smoothness of a random field, the parameter ν in the Matérn class, plays a critical role in interpolation problems. This parameter is difficult to estimate accurately from data. A number of the commonly used models for the covariance structure, including spherical, exponential and squared exponential structures assume that the smoothness parameter is known a priori.

As an alternative to the Matérn covariance sometimes the powered exponential model could be used.

$$C(\mathbf{x}) = \sigma e^{-\alpha|\mathbf{x}|^{\gamma}}$$

with $\alpha > 0$ and $\gamma \in (0, 2]$. The parameter γ (when $\gamma < 2$) plays the same role as 2ν in the Matérn, and for $\gamma = 2$ it corresponds to $\nu = \infty$. However, for values of $1 \leq \nu < \infty$ the powered exonential has not elements providing similar local behavior as the Matérn.

Figure 2 shows two Matérn covariances, a squared exponential covariance and a exponential covariance. The squared exponential is more flat the the origin, this indicates that the spatial process is very smooth. On the other hand, the exponential is almost linear at the origin, indicating that the corresponding spatial process is not very smooth, in fact this process is not even once mean square differentiable.

Figure 3 shows other two Matérn covariances with $\nu = 1/2$ (exponential) and with $\nu = 3/2$, which corresponds to a process that is once differentiable. Figure 4 shows the corresponding spectral densities.

1.4 Estimating the spectral density

The periodogram, a nonparametric estimate of the spectral density, is a powerful tool for studying the properties of stationary processes observed on a d-dimensional lattice. Use and properties of spatial periodograms for stationary processes have been investigated by Stein (1995), Guyon (1982,1992), Ripley (1981), Rosenblatt (1985), and Whittle (1954) among others. Pawitan and O'Sullivan (1994) proposed a nonparametric spectral density estimator using a penalized Whittle likelihood for a stationary time series. Guyon (1982) studied the asymptotic properties of various parameter estimation procedures for a general stationary process on a d-dimensional lattice, using spectral methods.

This Section is organized as follows. First, we introduce the periodogram, a nonparametric estimate of the spectral density. Then, by using spectral tools we present an expression for the likelihood function that in practice is very easy to calculate. This version of the likelihood was proposed by Whittle (1954) and avoids computing the determinants an inverses of large matrices.

1.4.1 Periodogram

Consider a spatial stationary process $Z(\cdot)$ with covariance parameter $\boldsymbol{\theta}$ which is assumed here to be known. We observe the process at N equally spaced locations in a regular grid D $(n_1 \times n_2)$, where $N = n_1 n_2$. The distance between neighboring observations is Δ. The periodogram is a nonparametric estimate of the spectral density, which is the Fourier transform of the covariance function. We define $I_N(\boldsymbol{\omega}_0)$ to be the periodogram at a frequency $\boldsymbol{\omega}_0$,

$$I_N(\boldsymbol{\omega}_0) = \Delta^2 (2\pi)^{-2} (n_1 n_2)^{-1} \left| \sum_{s_1=1}^{n_1} \sum_{s_2=1}^{n_2} Z(\Delta \mathbf{s}) \exp\{-i \Delta \mathbf{s}^T \boldsymbol{\omega}\} \right|^2 . \tag{13}$$

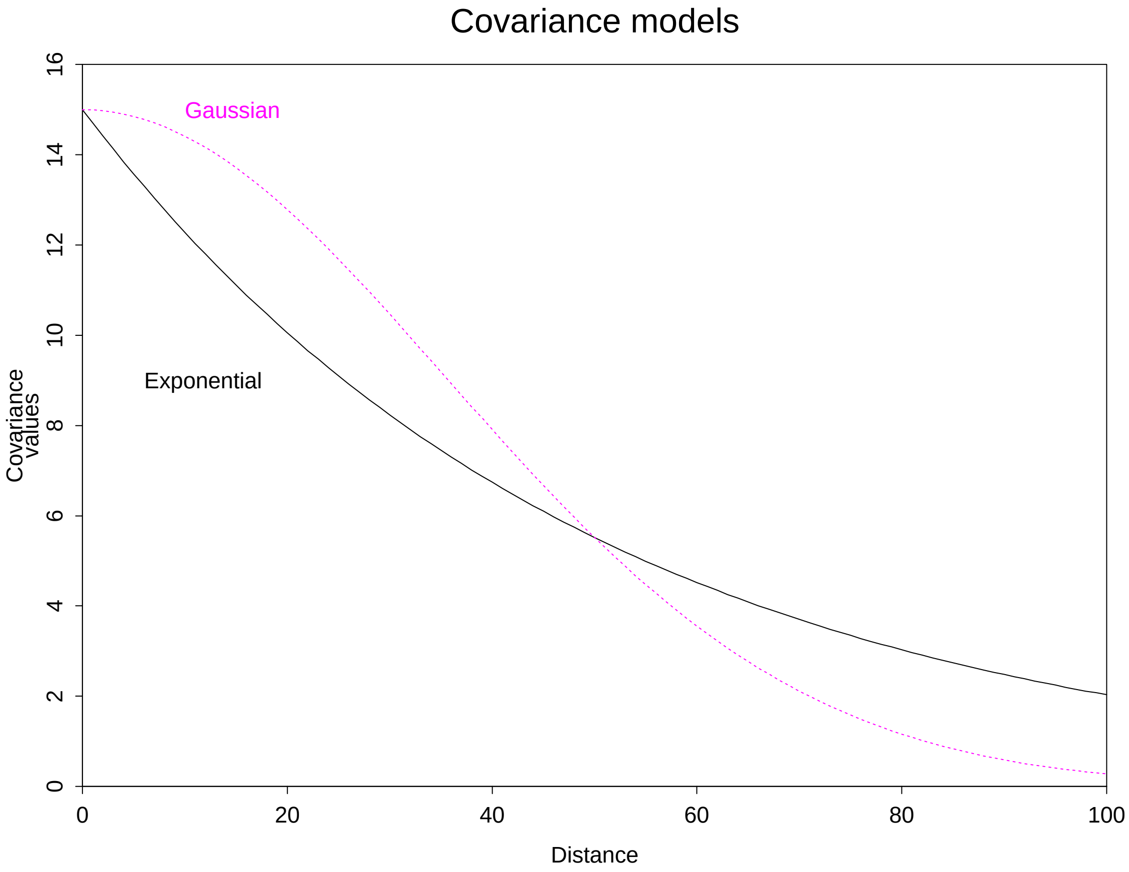


Figure 2: Covariance models: Exponential and squared exponential.

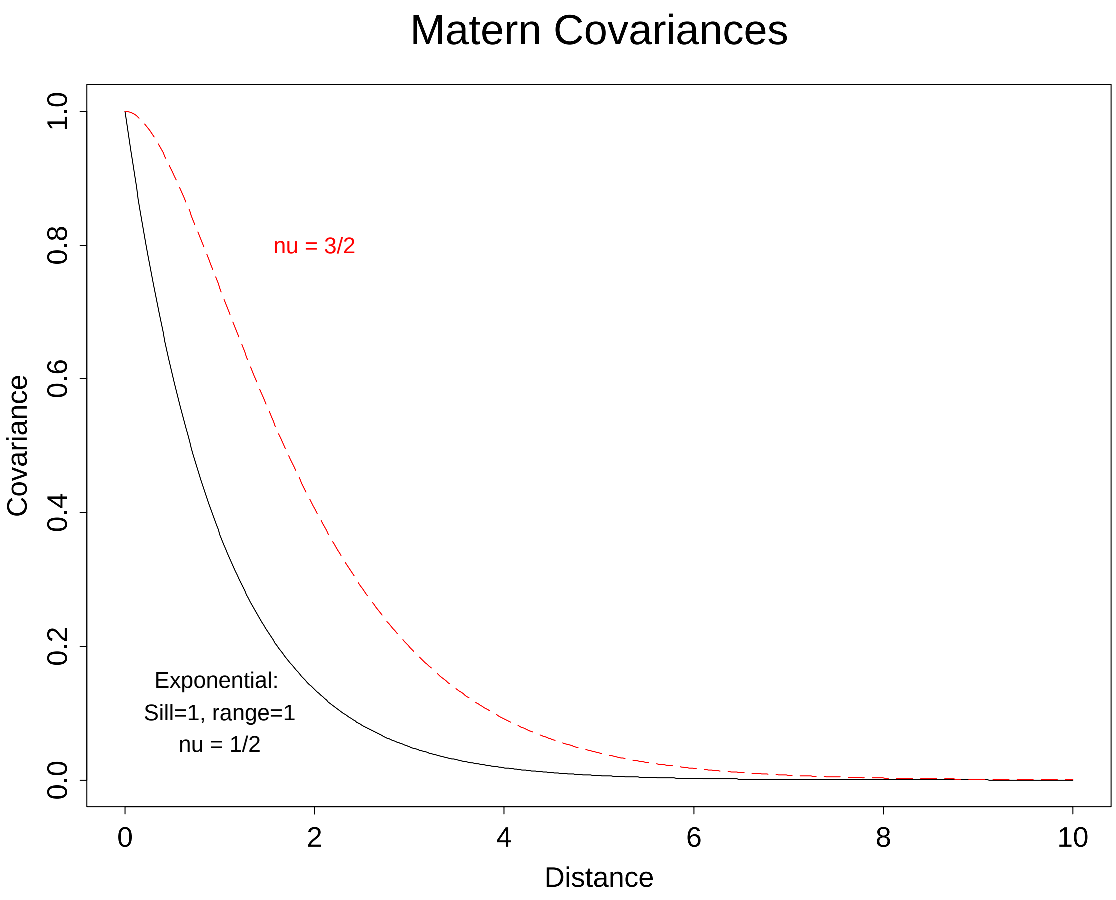


Figure 3: Covariances. Matérn Class for $\nu = 1/2$ (exponential covariance) and $\nu = 3/2$.

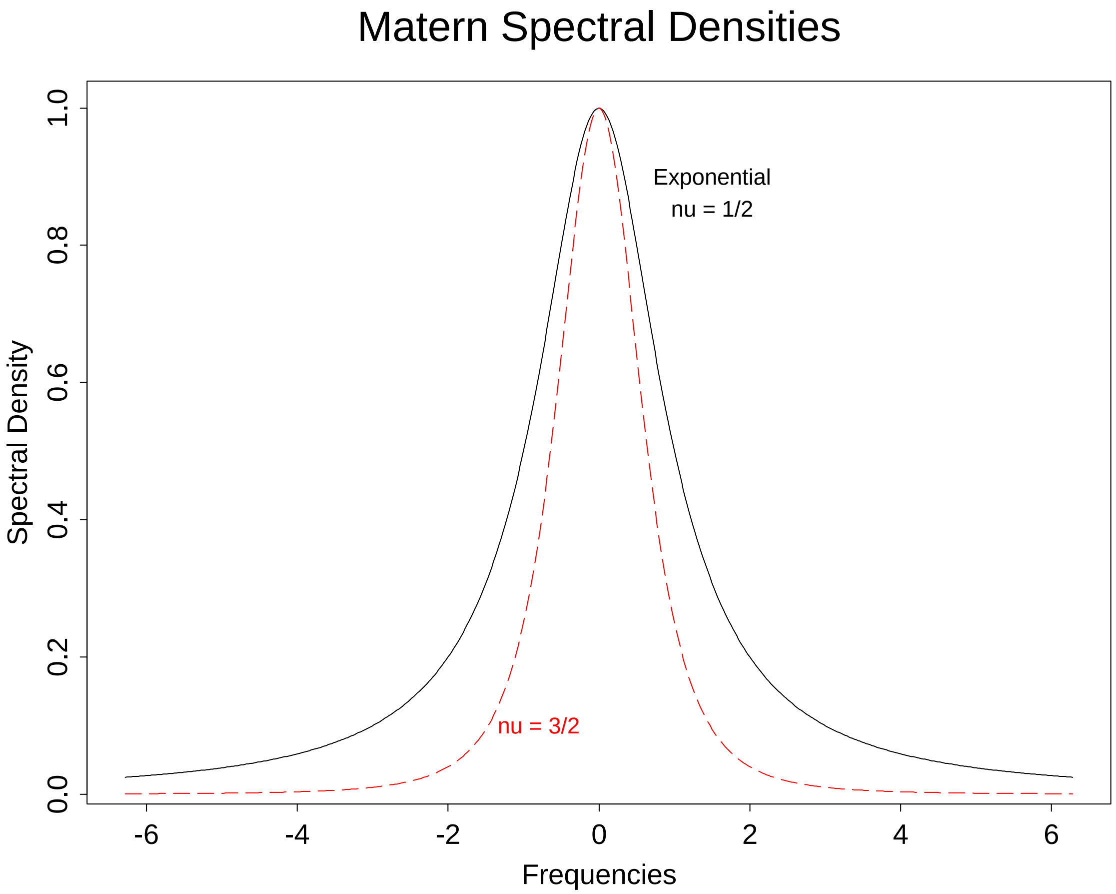


Figure 4: Spectral Densities. Matérn Class for $\nu = 1/2$ (exponential covariance) and $\nu = 3/2$.

If the spectral representation of Z is

$$Z(\mathbf{s}) = \int_{\mathbb{R}^2} e^{i\boldsymbol{\omega}^T\mathbf{x}} dY(\boldsymbol{\omega}),$$

we define $J(\boldsymbol{\omega})$, a discrete version of the spectral process $Y(\boldsymbol{\omega})$, which is the Fourier transform of Z,

$$J(\boldsymbol{\omega}) = \Delta(2\pi)^{-1}(n_1n_2)^{-1/2}\sum_{x_1=1}^{n_1}\sum_{x_2=1}^{n_2} Z(\Delta\mathbf{x})\exp\{-i\Delta\mathbf{x}^T\boldsymbol{\omega}\}.$$

Using the spectral representation of Z and proceeding formally,

$$C(\mathbf{x}) = \int_{\mathbb{R}^2} \exp(i\boldsymbol{\omega}^T\mathbf{x})F(\mathbf{d}\boldsymbol{\omega}) \tag{14}$$

where the function F is called the spectral measure or spectrum for Z. F is a positive finite measure, defined by

$$E\{d|Y(\boldsymbol{\omega})|^2\} = dF(\boldsymbol{\omega}). \tag{15}$$

Thus, we get

$$I_N(\boldsymbol{\omega}) = |J(\boldsymbol{\omega})|^2, \tag{16}$$

this expression for I_N is consistent with the definition of the spectral distribution F in (15), as a function of the spectral processes Y. The periodogram (15) is simply the discrete fourier transform of the sample covariance.

In practice, the periodogram estimate for $\boldsymbol{\omega}$ is computed in the set of Fourier frequencies $(2\pi/Delta)(\mathbf{f}/\mathbf{n})$ where $\mathbf{f}/\mathbf{n} = \left(\frac{f_1}{n_1}, \frac{f_2}{n_2}\right)$, and $\mathbf{f} \in J_N$, for

$$J_N = \{\lfloor -(n_1-1)/2\rfloor, \ldots, n_1 - \lfloor n_1/2\rfloor\} \times \{\lfloor -(n_2-1)/2\rfloor, \ldots, n_2 - \lfloor n_2/2\rfloor\}. \tag{17}$$

where $\lfloor u\rfloor$ denotes the largest integer less or equal than u.

1.4.2 Theoretical properties of the periodogram

The expected value of the periodogram at $\boldsymbol{\omega}_0$ is given by

$$E(I_N(\boldsymbol{\omega}_0)) = (2\pi)^{-2}(n_1n_2)^{-1}\int_{\Pi^2_\Delta} f_\Delta(\boldsymbol{\omega})W(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\mathbf{d}\boldsymbol{\omega},$$

where $\Pi^2_\Delta = (-\pi/\Delta, \pi/\Delta)^2$, and

$$W(\boldsymbol{\omega}) = \prod_{j=1}^{2}\frac{\sin^2\left(\frac{n_j\omega_j}{2}\right)}{\sin^2\left(\frac{\omega_j}{2}\right)}$$

for $\boldsymbol{\omega} = (\omega_1, \omega_2) = (2\pi/Delta)(\mathbf{f}/\mathbf{n})$ and $\mathbf{f} \in J_N\backslash\{0\}$, and $f_\Delta(\boldsymbol{\omega})$ is the spectral density of the process Z on the lattice with spacing Δ. The side lobes (subsidiary peaks) of the function W can lead to substantial bias in $I_N(\boldsymbol{\omega}_0)$ as an estimator of $f_\Delta(\boldsymbol{\omega}_0)$ since they allow the value of f_Δ at frequencies far from $\boldsymbol{\omega}_0$ to contribute to the expected value. Figure 1.4.4 shows a graph of W along the vertical axis. If the side lobes of W were substantially smaller, we could reduce this source of bias for the periodogram considerably. Tapering is a technique that effectively reduces the side lobes associated with the spectral window W. We form the product $h(\mathbf{x})Z_i(\mathbf{x})$ for each value of $\mathbf{x} = (x_1, x_2)$, where $\{h(\mathbf{x})\}$ is a suitable sequence of real-valued constants called a *data taper,* and then we compute the periodogram for the tapered data.

The periodogram values are approximately independent, and this facilitates the use of techniques such as non-linear least squares (NLS) to fit a theoretical spectral model to the periodogram values.

Asymptotic properties of the periodogram

Theorem 1 (Brillinger, 1981):

Consider a Gaussian stationary process Z with spectral density $f(\boldsymbol{\omega})$ on a lattice D. We assume Z is observed at N equally spaced locations in D ($n_1 \times n_2$), where $N = n_1 n_2$, and the spacing between observations is Δ. We define the periodogram function, $I_N(\boldsymbol{\omega})$, as in (34).

Assume $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \lambda$, for a constant $\lambda > 0$.

Then, we get:

(i) The expected value of the periodogram, $I_N(\boldsymbol{\omega})$, is asymptotically $f_\Delta(\boldsymbol{\omega})$.

(ii) The asymptotic variance of $I_N(\boldsymbol{\omega})$ is $f^2_\Delta(\boldsymbol{\omega})$.

(iii) The periodogram values $I_N(\boldsymbol{\omega})$, and $I_N(\boldsymbol{\omega}')$ for $\boldsymbol{\omega} \neq \boldsymbol{\omega}'$, are asymptotically independent.

By Part (i) the periodogram I_N using *increasing domain asymptotics* is asymptotically an unbiased estimate of the spectral density, f_Δ on the lattice. Note, that if f is the continuous process Z, and f_Δ the spectral density on the lattice, then using *increasing-domain* asymptotics, I_N is not asymptotically an unbiased estimate of f but of f_Δ the spectral density of the sampled sequence $Z(\Delta\mathbf{x})$. By Theorem 1 part (ii) the variance of the periodogram at $\boldsymbol{\omega}$ is asymptotically $f^2_\Delta(\boldsymbol{\omega})$. The traditional approach to this inconsistency problem is to smooth the periodogram across frequencies.

By Theorem 1 part (iii), the periodogram values are approximately independent. This property allow us to easily fit in the spectral domain a parametric model to the periodogram values. However, in the space domain, the correlation among the empirical covariance or variogram values thwarts the use of least squares.

Asymptotic distribution of the periodogram

If the process Z is stationary, such that the absolute value of the joint cummulants of order k are integrable (for all k), then the periodogram has asymptotically a distribution that is multiple of a χ^2_2. More specifically, the periodogram $I_N(\boldsymbol{\omega}_j)$ where $\boldsymbol{\omega}_j$ is a Fourier frequency has asymptotically a $f(\boldsymbol{\omega}_j)\chi^2_2/2$ distribution.

1.4.3 Examples of spectral analysis of 2-d processes

A simulated example

In order to get a feel for how bivariate spectra look, we borrow an example from Renshaw and Ford (Renshaw and Ford (1983)). The process is in essence a weighted avarage of cosines with random phase. Figure 5 shows the random field, which is generated by

$$Z(s) = \sum_{i=0}^{16} \sum_{j=-16}^{15} g_{ij} \cos(2\pi \left(\frac{is_1}{m} + \frac{js_2}{n} \right) + U_{ij})$$

where

$$g_{ij} = \exp(- |i + 6 - j \tan(20^\circ)|).$$

The resulting spectrum, Figure 1.4.3 shows a single ridge in the 20° direction, going from $I_{0,-6}$ to $I_{16,0}$.

A forestry example

The main topic of the paper by Renshaw and Ford (Renshaw and Ford (1983)) is the analysis of canopy heights in a managed forest of scots pine at Thetford Forest, U.K. The forest was about 40 years old when the measurements were taken, and had about 1000 tress per hectare. The horizontal extent of individual tree crowns was measure along parallel transects at 1m intervals, and the crown height, location, and perimeter crown height of each tree was determined. Intermediate crown heights were estimated by geometric projection. We focus on a 32 by 32 submatrix of canopy heights. The spectrum is given in Figure 1.4.3.

The spectrum shows a low frequency ridge in the positive quadrat from (1,0) to (8,4), with large elements at (4,1) and (8,2), corresponding to an angle of 76°. Separated from this ridge by a single row is a ridge of high frequency elements with peak value at (12,4), or 72°. This strong directional feature was not obvious when observing the canopy from above. There is also a ridge in the negative quadrant from (4,-4) to (15,1). The high values are clustered around 90° and constitute a basic row effect.

When looking at the entire data set of 36 by 120 meters, there is clear evidence of a changing spectral structure by looking at partially overlapping submatrices. The high frequency components remain relatively stable, but the low frequency features shift between dominating for negative values of $omega_2$ to positive values as one moves along the forest. This is probably related to a thinning that was less severe around a measurement mast used for microclimate measurements in the middle of the study area. Where there has been least thinning, the wave-like aggregation of tree crowns is strongest. Thus, for the entire data set, a model allowing for nonstationary spectral analysis is needed.

1.4.4 Nonlinear WLS estimation in the spectral domain

Consider modeling the spatial structure of Z by fitting a spectral density f to the periodogram values. We could use a weighted non-linear least squares (WNLS) procedure, that gives more weight to higher frequency values because high frequencies are important for interpolation. An approximate expression for the spectral density of the Matérn class for high frequency values is obtained from (33) by letting $|\boldsymbol{\omega}|$ go to ∞:

$$f(\boldsymbol{\omega}) = \phi(|\boldsymbol{\omega}|^2)^{\left(-\nu - \frac{d}{2}\right)} \tag{18}$$

Thus, the degree of smoothness, ν, and ϕ are the critical parameters (and not the range α^{-1}).

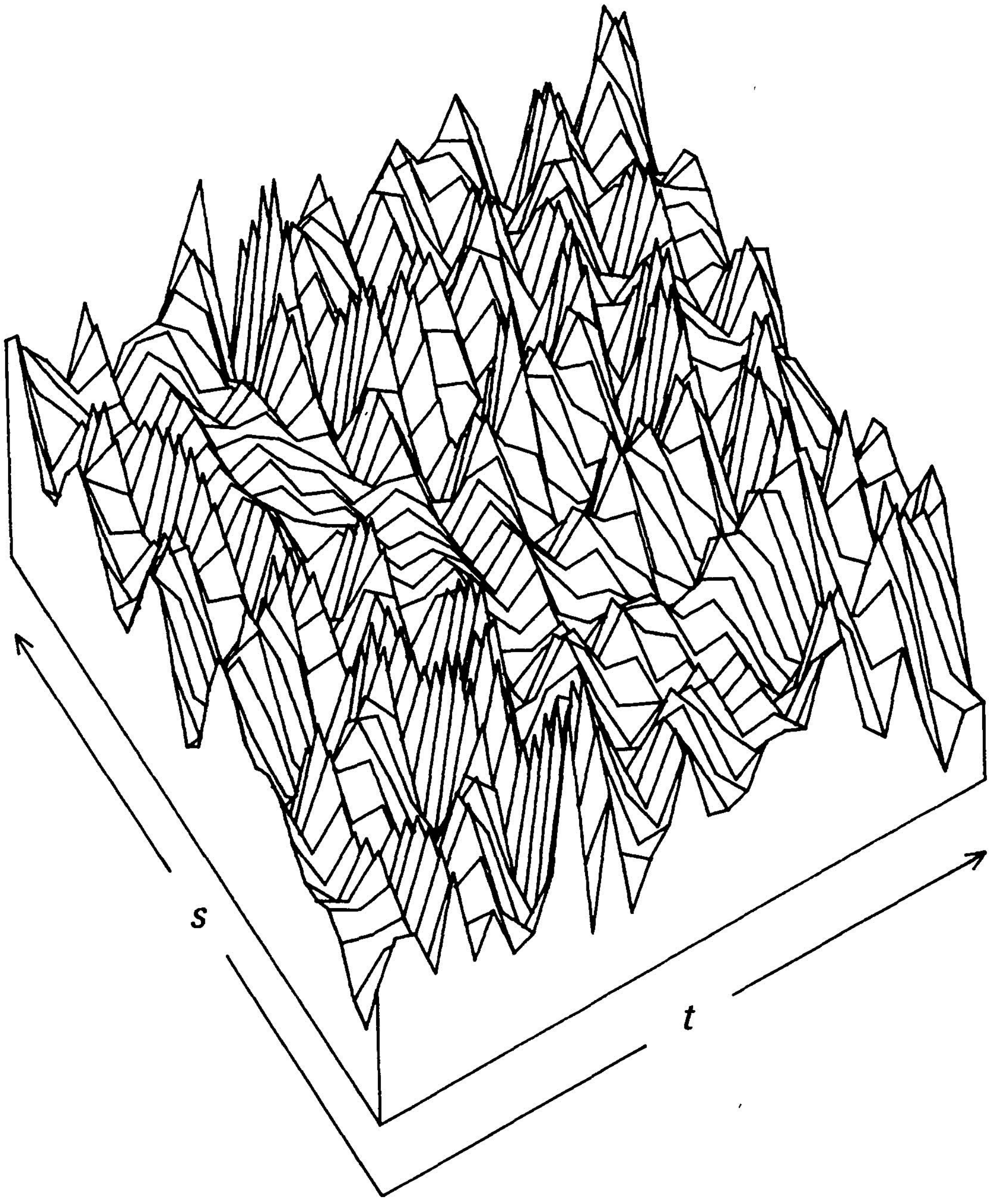


Figure 5: A simulated process with a strong directional component. From Renshaw and Ford (1983).

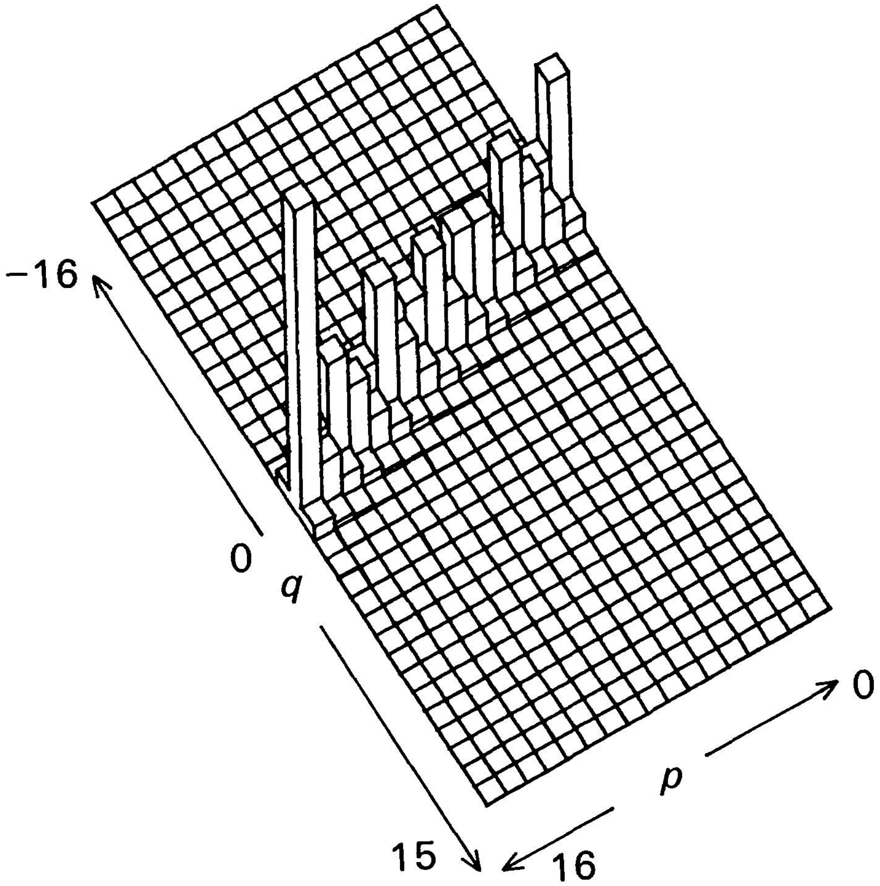


Figure 6: The periodogram for the process in Figure 5. From Renshaw and Ford (1983).

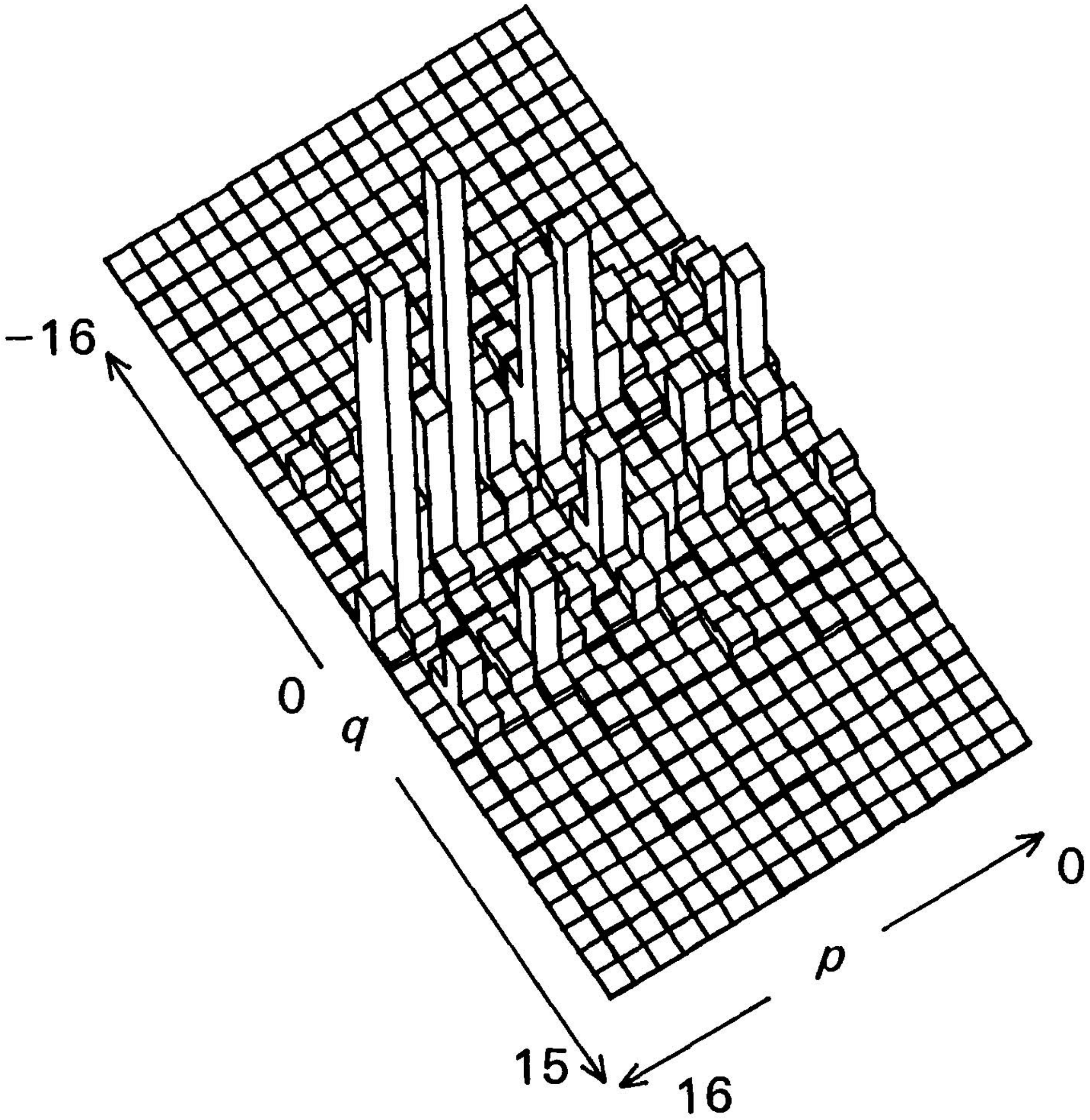


Figure 7: The spectrum of a 32 by 32 submatrix of the Thetford forest canopy heights. The notation (p, q) corresponds to our (ω_1, ω_2). From Renshaw and Ford (1983).

Thus, we propose using the high frequency parametric model for f as in (18) and the weights $f(\boldsymbol{\omega})^{-1}$ to give higher weight to higher frequencies. This is reasonable since for large N the approximate standard deviation of the periodogram I_N is $f(\boldsymbol{\omega})$.

Thus, the proposed weights $f(\boldsymbol{\omega})^{-1}$ also stabilize the variance of the periodogram values. This is similar to the weighted least squares method used in the space domain to fit a variogram model (Cressie, 1985). We recommend to use weighted least squares in the spectral domain rather than in the space domain, because periodogram values are approximately independent while variogram values are not.

An alternative approach would be to fit in the log scale a linear model:

$$\log(f(\boldsymbol{\omega})) = \beta_0 + \beta_1 X \tag{19}$$

where $X = \log(\boldsymbol{\omega})$, $\beta_{0i} = \log(\phi)$ and $\beta_1 = 2\left(-\nu - \frac{d}{2}\right)$.

1.4.5 Likelihood estimation in the spectral domain

For large datasets, calculating the determinants that we have in the likelihood function can be often infeasible. Spectral methods could be used to approximate the likelihood and obtain the maximum likelihood estimates (MLE) of the covariance parameters: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)$.

Spectral methods to approximate the spatial likelihood have been used by Whittle 1954, Guyon 1982, Dahlhaus and Küsch, 1987, and Stein 1995, 1999, among others. These spectral methods are based on Whittle's (1954) approximation to the Gaussian negative log likelihood:

$$\frac{N}{(2\pi)^2} \sum \log f(\boldsymbol{\omega}) + I_N(\boldsymbol{\omega}) f(\boldsymbol{\omega})^{-1} \tag{20}$$

where the sum is evaluated at the Fourier frequencies, I_N is the periodogram and f is the spectral density of the lattice process. The approximated likelihood can be calculated very efficiently by using the fast Fourier transform. This approximation requires only $O(N log_2 N)$ operations. Simulation studies conducted by the author seem to indicate that N needs to be at least 100 to get good estimated MLE parameters using Whittle's approximation.

The asymptotic covariance matrix of the MLE estimates of $\theta_1, \ldots, \theta_r$ is

$$\left\{ \frac{2}{N} \left[\frac{1}{4\pi^2} \int_{[-\pi,\pi]} \int_{[-\pi,\pi]} \frac{\partial \log f(\boldsymbol{\omega}_1)}{\partial \theta_j} \frac{\partial \log f(\boldsymbol{\omega}_2)}{\partial \theta_k} d\boldsymbol{\omega}_1 d\boldsymbol{\omega}_2 \right]^{-1} \right\}_{jk} \tag{21}$$

this is much easier to compute that the inverse of the Fisher information matrix.

Guyon (1982) proved that when the periodogram is used to approximate the spectral density in the Whittle likelihood function, the periodogram bias contributes a non-negligible component to the mean squared error (mse) of the parameter estimates for 2-dimensional processes, and for 3-dimensions this bias dominates the mse. Thus, the MLE parameters of the covariance function based on the Whittle likelihood are only efficient in one dimension, but not in two and higher dimensional problems. Though, they are consistent. Guyon demonstrated that this problem can be solved by using a different version of the periodogram, an "unbiased peridogram", which is the discrete Fourier transform of an unbiased version of the sample covariance. Dahlhaus and Künsch (1987) demonstrated that tapering also solves this problem.

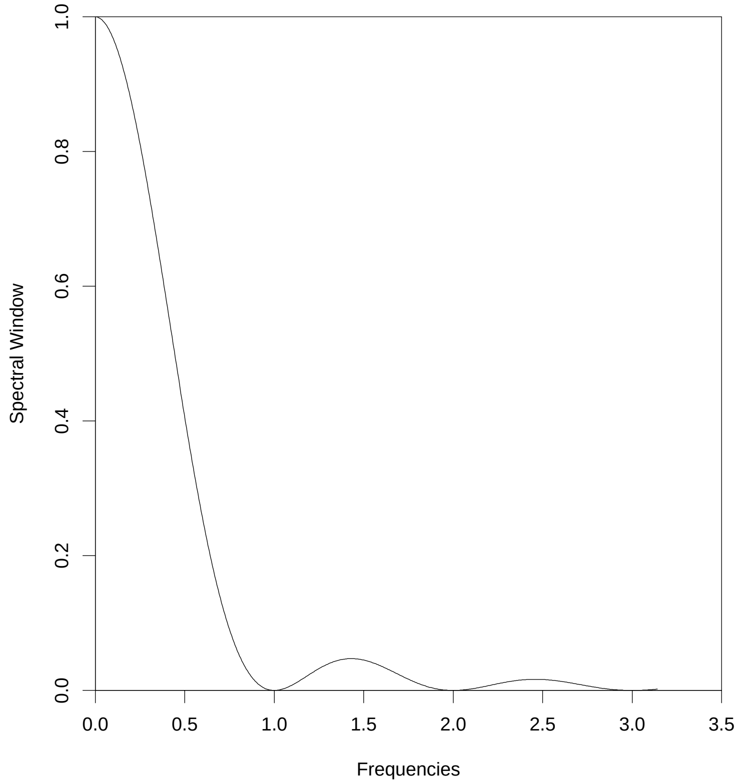


Figure 8: Spectral window along the vertical axis. The horizontal axis shows the frequencies, while the vertical axis shows the spectral window along the vertical axis for the periodogram (without tapering), for $n_2 = 500$.

1.5 Model for nonstationary processes

1.5.1 Convolution of locally stationary processes

In this section, we propose a class of nonstationary processes, based on a convolution of local stationary processes (Fuentes, 2001, 2002, Fuentes and Smith, 2001). This model has the advantage that the model is simultaneously defined everywhere, unlike moving window approaches (Haas, 1998), but it retains the attractive property that, locally in small regions, it behaves like a stationary spatial process.

We model a nonstationary process Z observed on a region D as a convolution of weakly stationary processes (Fuentes, 2001, Fuentes, 2002):

$$Z(\mathbf{x}) = \int_D K(\mathbf{x}-\mathbf{s}) Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}) d\mathbf{s}, \tag{22}$$

where K is a kernel function and $Z_{\theta(\mathbf{s})}(\mathbf{x}),\ \mathbf{x} \in D$, is a family of independent stationary Gaussian processes indexed by $\theta(\mathbf{s})$.

The covariance function $C(\mathbf{x_1}, \mathbf{x_2}; \theta)$ of Z is a convolution of the covariance functions $C_{\theta(\mathbf{s})}(\mathbf{x_1} - \mathbf{x_2})$ of the stationary processes $Z_{\theta(\mathbf{s})}$:

$$C(\mathbf{x_1}, \mathbf{x_2}; \boldsymbol{\theta}) = \int_D K(\mathbf{x_1}-\mathbf{s}) K(\mathbf{x_2}-\mathbf{s}) C_{\theta(\mathbf{s})}(\mathbf{x_1}-\mathbf{x_2}) d\mathbf{s}. \tag{23}$$

The covariance function of $Z_{\theta(\mathbf{s})}$ is stationary with parameter $\theta(\mathbf{s})$, and we assume that $\theta(\mathbf{s})$ is a continuous function on $\mathbf{s}$. The process $Z_{\theta(\mathbf{s})}$ could have a Matérn isotropic covariance function of the form

$$C_{\theta(\mathbf{s})}(\mathbf{x}) = \frac{\pi^{d/2}\phi_s}{2^{\nu_s-1}\Gamma(\nu_s + d/2)\alpha_s^{2\nu_s}} (\alpha_s|\mathbf{x}|)^{\nu_s} \mathcal{K}_{\nu_s}(\alpha_s|\mathbf{x}|), \tag{24}$$

where $\mathcal{K}_{\nu_s}$ is a modified Bessel function, and d is the dimension of $\mathbf{s}$, $\theta(\mathbf{s}) = (\nu_s, \alpha_s, \phi_s)$. The parameter α_s^{-1} can be interpreted as the autocorrelation range, ϕ_s is a scale parameter, and the parameter ν_s measures the degree of smoothness of the process $Z_{\boldsymbol{\theta}(\mathbf{s})}$.

In (23) every entry requires an integration. Since each such integration is actually an expectation with respect to a uniform distribution, we could use Monte Carlo integration or other numeric scheme to approximate the integral (23). We propose to draw an independent set of locations $\mathbf{s}_i$, $i = 1, 2, ..., k$, on D. Hence, we replace $C(\mathbf{x_1}, \mathbf{x_2}; \boldsymbol{\theta})$ with

$$\hat{C}(\mathbf{x_1}, \mathbf{x_2}; \boldsymbol{\theta}) = k^{-1} \sum_{i=1}^{k} K(\mathbf{x}_1 - \mathbf{s}_i) K(\mathbf{x}_2 - \mathbf{s}_i) C_{\boldsymbol{\theta}(\mathbf{s}_i)}(\mathbf{x_1} - \mathbf{x_2}). \tag{25}$$

In this notation, the 'hat' denotes an approximation that can be made arbitrarily accurate and has nothing to do with the data Z. The kernel function $K(\mathbf{x} - \mathbf{s}_i)$ centred at $\mathbf{s}_i$ could be positive for all $\mathbf{x} \in D$, or could have compact support. In the latter case, $K(\mathbf{x} - \mathbf{s}_i)$ would be only positive when $\mathbf{x}$ is in a subregion S_i centred at $\mathbf{s}_i$, and this would simplify the calculations.

The size of the sample, k, is selected using the following iterative algorithm. We first start with a systematic sample of size k, where k is small, and we increase k by adding a new sample point at a time. At each step of the iterative approach we draw a new sample point in between two neighbouring points in the current sample sequence. Thus, in each iteration we decrease by half the distance between two neighbouring draws. We iterate this process until an Akaike information criterion (Akaike, 1974) or Bayesian information criterion (BIC) suggests no significant improvement in the estimation of the nonstationary covariance of Z by increasing k, equivalent to decreasing the distance between draws in the sample sequence.

Throughout the rest of this Section we simplify the notation by writing Z_i to denote $Z_{\boldsymbol{\theta}(\mathbf{s}_i)}$, and $w_i(\mathbf{x})$ to represent $K(\mathbf{x}-\mathbf{s}_i)$, the kernel or weight function centred at $\mathbf{s}_i$.

1.5.2 The spectrum for the convolution model

The nonstationary process Z is modelled here as a mixture of weakly stationary processes Z_i, $i=1,\ldots,k$, with $\mathrm{cov}\{Z_i(\mathbf{x}), Z_j(\mathbf{y})\}=0$ for $i \neq j$:

$$Z(\mathbf{x}) = \sum_{i=1}^{k} Z_i(\mathbf{x}) w_i(\mathbf{x}), \tag{26}$$

and we choose k using the BIC (or AIC) approach discussed in the previous Section.

Each stationary process Z_i has the representation,

$$Z_i(\mathbf{x}) = \int_{\mathbb{R}^2} \exp(i\mathbf{x}^{\mathrm{T}}\boldsymbol{\omega}) dY_i(\boldsymbol{\omega}), \tag{27}$$

where the Y_i are random functions with uncorrelated increments.

Thus, the spectral representation of Z is $Z(\mathbf{x}) = \int_{\mathbb{R}^2} \exp(i\mathbf{x}^{\mathrm{T}}\boldsymbol{\omega}) dY(\boldsymbol{\omega})$, where

$$Y(\boldsymbol{\omega}) = \sum_{i=1}^{k} \mathcal{F}\{w_i\} * Y_i(\boldsymbol{\omega}), \tag{28}$$

$\mathcal{F}\{w_i\}$ is the Fourier transform of w_i, and $*$ denotes the following convolution function:

$$\mathcal{F}\{w_i\} * Y_i(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} \mathcal{F}\{w_i\}(\mathbf{h}) Y_i(\boldsymbol{\omega}-\mathbf{h}) d\mathbf{h}.$$

The covariance function of Z can be defined in terms of the covariance function of the orthogonal stationary processes Z_i :

$$\mathrm{cov}\{Z(\mathbf{x}_1), Z(\mathbf{x}_2)\} = \sum_{i=1}^{k} w_i(\mathbf{x}_1) w_i(\mathbf{x}_2) \mathrm{cov}\{Z_i(\mathbf{x}_1), Z_i(\mathbf{x}_2)\}. \tag{29}$$

This is a valid nonstationary covariance function. Then, the corresponding spectral density is,

$$f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = \sum_{i=1}^{k} f_i * \{\mathcal{F}\{w_i\}(\boldsymbol{\omega}_1) \mathcal{F}\{w_i\}(\boldsymbol{\omega}_2)\}, \tag{30}$$

where

$$f_i * \{\mathcal{F}\{w_i\}(\boldsymbol{\omega}_1) \mathcal{F}\{w_i\}(\boldsymbol{\omega}_2)\} = \int_{\mathbb{R}^2} f_i(\boldsymbol{\omega}) \mathcal{F}\{w_i\}(\boldsymbol{\omega}_1-\boldsymbol{\omega}) \mathcal{F}\{w_i\}(\boldsymbol{\omega}_2-\boldsymbol{\omega}) d\boldsymbol{\omega}.$$

1.5.3 Nonparametric spectral estimation

We present here an asymptotically unbiased nonparametric estimator, $\tilde{I}_N$, of the spectral density f of a nonstationary process Z. We model Z as in (26). Thus, a natural way of defining $\tilde{I}_N$ is as a convolution of the periodograms $I_{i,N}$ of the stationary processes Z_i with domain D:

$$\tilde{I}_N(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = \sum_{i=1}^{k} I_{i,N} * \{\mathcal{F}\{w_i\}(\boldsymbol{\omega}_1)\mathcal{F}\{w_i\}(\boldsymbol{\omega}_2)\}, \tag{31}$$

where $*$ denotes the convolution

$$I_{i,N} * \{\mathcal{F}\{w_i\}(\boldsymbol{\omega}_1)\mathcal{F}\{w_i\}(\boldsymbol{\omega}_2)\} = \sum_{\boldsymbol{\omega} \in J_N} I_{i,N}(\boldsymbol{\omega})\mathcal{F}\{w_i\}(\boldsymbol{\omega}_1 - \boldsymbol{\omega})\mathcal{F}\{w_i\}(\boldsymbol{\omega}_2 - \boldsymbol{\omega}),$$

with J_N the set of the Fourier frequencies (17). The weights w_i have compact support (they are only positive in the corresponding subregion S_i of stationarity) and they help to identify the processes Z_i that are being used. By the definition of f in (30) as a function of the spectral densities f_i, $i = 1, \ldots, k$, and the fact that the periodograms $I_{i,N}$ are asymptotically unbiased estimators of f_i, we obtain that $\tilde{I}$ is asymptotically unbiased. The asymptotic variance of $\tilde{I}_N(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ can be easily obtained because the processes Z_i are orthogonal. Thus, when $n_i \to \infty$, for $i = 1, 2$, $\Delta \to 0$ and $\Delta n_1 \to \infty$, $\Delta n_2 \to \infty$, the variance of $\tilde{I}_N(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ becomes

$$\sum_{i=1}^{k} f_i^2 * \{\mathcal{F}\{w_i\}^2(\boldsymbol{\omega}_1)\mathcal{F}\{w_i\}^2(\boldsymbol{\omega}_2)\}.$$

Furthermore, since $\tilde{I}_N$ is a convolution of independent stationary periodograms, we obtain $\operatorname{cov}\{\tilde{I}_N(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2), \tilde{I}_N(\boldsymbol{\omega}'_1, \boldsymbol{\omega}_2\prime)\} = 0$ asymptotically.

In practice, we compute $I_{i,N}$ as the periodogram of the observed values of Z in the subregion of stationarity S_i,

1.5.4 Parametric spectral estimation

Suppose again that Z takes the form (26), so that we use the expression in (30) for f. The spectral density f is modelled then as a function of the spectral densities f_i, $i = 1, \ldots, k$.

A parametric estimator $\hat{f}$ of the spectral density is easily obtained from parametric estimators of the spectral densities f_i, $i = 1, \ldots, k$:

$$\hat{f}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = \sum_{i=1}^{k} \hat{f}_i * \{\mathcal{F}\{w_i\}(\boldsymbol{\omega}_1)\mathcal{F}\{w_i\}(\boldsymbol{\omega}_2)\}. \tag{32}$$

We study now parametric models for the f_i. A class of practical variograms and autocovariance functions for the stationary processes Z_i can be obtained from the Matérn class of spectral densities

$$f_i(\boldsymbol{\omega}) = \phi_i(\alpha_i^2 + \|\boldsymbol{\omega}\|^2)^{(-\nu_i - \frac{d}{2})} \tag{33}$$

with parameters $\nu_i > 0$, $\alpha_i > 0$ and $\phi_i > 0$, where d is the dimensionality of Z_i. Here, the vector of covariance parameters is $\theta_i = (\phi_i, \nu_i, \alpha_i)$.

1.5.5 An example in air quality

In the air pollution example presented here the main objective is to understand and quantify the weekly spatial structure of air pollutants using the output of a regional scale air quality models, known as Models-3. Models-3 estimates hourly concentrations and fluxes of different pollutants. We study here the nitric acid. The spatial domain D (Figure 9) is a regular 81×87 grid, where the dimensions of each cell on the grid are 36km × 36km. The 81×87 lattice for Models-3 is a two-dimensional grid that takes account of the earth's curvature. Models-3 provides the estimated concentration for the middle point of each cell. In this example we analyse the spatial structure of the hourly averaged nitric acid concentrations for the week starting July 11, 1995. We fit model (22), taking K to be the Epanechnikov kernel $K(u) = \frac{2}{\pi}(1 - ||u||^2/h^2)$ with h an arbitrary bandwidth, and replacing the integral over D by a sum over a grid of cells covering the observation region.

In practice, the choice of h is crucial. The bandwidth should be small to preserve the general 'shape' of the data (Clark, 1977). In a regression setting, reducing the size of the bandwidth reduces the bias but increases the variance. In our spatial setting, since the variance might change with location we do not gain much by increasing h. The shape of the process is represented by the parameter $\boldsymbol{\theta}$, which accounts for the lack of stationarity of Z. Thus, we need to choose h as small as possible to preserve this general shape. However, we also need to ensure that for all $\mathbf{x} \in D$ there is at least one $\mathbf{s}_i$, with $K(\mathbf{x} - \mathbf{s}_i) > 0$, where the $\{\mathbf{s}_i\}$ are the k draws on D to calculate the covariance (25). In this application we choose the smallest value of h that satisfies this condition. When the distance between neighbouring points of the sample sequence $\mathbf{s}_1, \ldots, \mathbf{s}_k$ varies, we could also allow the bandwidth to change with location. If we have k draws from a systematic sample with a distance l between sampling points, then the recommended value for h is $l/\sqrt{2}$. Note that the value of h depends on k.

In this example $k = 9$, which is the optimal value for k based on the AIC criterion. The sample points $\mathbf{s}_1, \ldots, \mathbf{s}_9$ are a systematic sample and they are plotted in Fig. 5. The distance l between the sampling points is 972 km. The value of h in this application is $h = l/\sqrt{2} = 687$ km. We used a likelihood approach to estimate the parameters of the nonstationary covariance matrix, which is a mixture of 9 stationary Matérn models of the form (33). Since the kernel function K has compact support, the covariance matrix of Z is approximately a block matrix, which simplifies the calculations; otherwise the evaluation of the likelihood function requires us to compute the inverse and determinant of a $7,209 \times 7,209$ matrix.

Figure 10 shows the Matérn fitted models (using the likelihood function of Z in the spectral domain and in the subregions S_i's) for the spectral densities, f_i, of the stationary processes Z_i, for $i = 1, \ldots, 9$. The nonstationary spectral density $f(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$, defined in (30), is obtained as a convolution of the densities f_i, $i = 1, \ldots, 9$. Table 1 shows the estimated parameters for the spectral densities f_i and the corresponding standard errors.

Process	Sill	Range	Smoothness
Z_1	2.9 (1.1)	566 (282)	0.74 (0.03)
Z_2	2.1 (0.5)	315 (131)	0.98 (0.10)
Z_3	1.8 (0.1)	231 (18)	1.01 (0.05)
Z_4	1.1 (0.3)	480 (193)	0.63 (0.03)
Z_5	1.22 (0.07)	150 (20)	1.19 (0.22)
Z_6	1.17 (0.01)	476 (17)	0.91 (0.01)
Z_7	1.4 (0.3)	500 (139)	0.46 (0.03)
Z_8	15 (3)	975 (292)	0.67 (0.01)
Z_9	3 (0.2)	252 (45)	0.84 (0.08)

TABLE 1. This table shows the estimated parameters for the spectral densities of the processes Z_i. The values in parenthesis are the standard errors of the estimated parameters. The parameters have been estimated using a likelihood approach.

The smoothing parameter represents the rate of decay of the spectral density at high frequencies; this is an indication of how smooth the corresponding process is. The smoothing parameter is approximately 0.5 corresponding to the exponential model, for the processes Z_1, Z_4, Z_7 and Z_8; these processes explain the spatial structure of the nitric acid concentrations on the eastern part of our domain; see the location of the sampling points in Fig. 6. We observe a relatively faster rate of decay at high frequencies for the processes Z_5, Z_6, Z_9, Z_2 and Z_3, with a smoothing parameter of approximately 1, corresponding to the Whittle model. These processes explain the spatial structure of the nitric acid concentrations on the western part of our domain, mainly over water; the nitric acid seems to be a smoother process over water than over the land surface. The nitric acid is a secondary pollutant, in being the result of photochemical reactions in the atmosphere rather than being emitted directly from sources on the surface. It therefore usually remains in the atmosphere for long periods of time and travels long distances across water.

When the range parameter is large, e.g. for process Z_8, there is a faster decay of the spectral density at short frequencies. We can appreciate this phenomenon by comparing the spectral density of Z_6, large range, to the spectral density of Z_5, small range. In general we observe larger ranges of autocorrelation on the western part of the grid. Furthermore, on the eastern part we should not expect large ranges because of the discontinuity of the nitric acid concentration that results from transition from land to ocean.

The variance of the process, also called the sill parameter, is the integral of the spectral density function, $\int_{\mathbb{R}^2} f(\boldsymbol{\omega})d\boldsymbol{\omega}$. In this example, the sill is relatively large for Z_8. There is higher spatial variability, large sill, mainly on the Great Lakes area, process Z_8, since the area is downwind from sources of pollution, primarily Chicago.

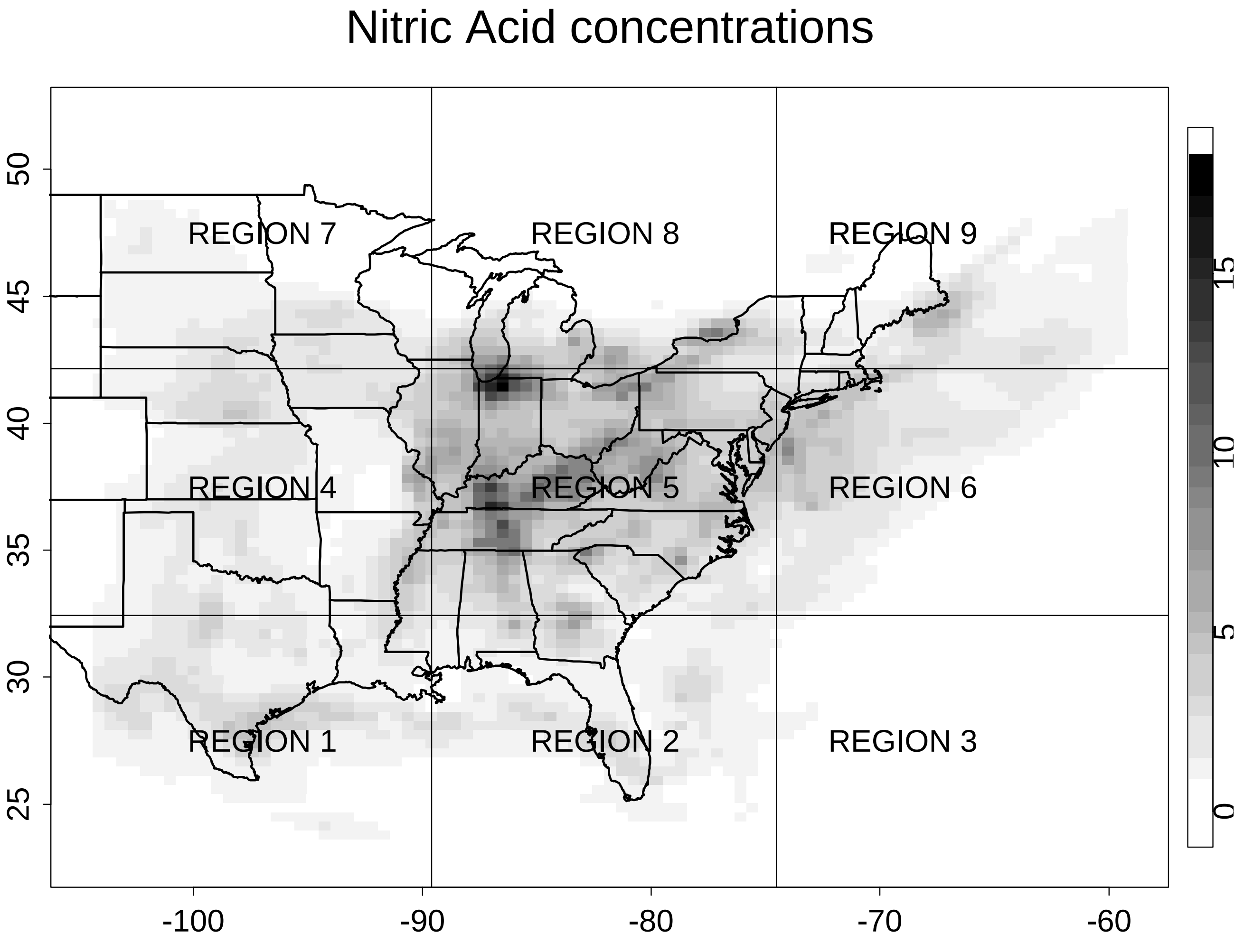


Figure 9: Output of EPA Models-3, showing the estimated concentrations of nitric acid (ppb) for the week starting July 11, 1995. The resolution is 36 km.

1.6 Testing for stationarity

We present now a formal test for stationarity of a spatial process. This test is a generalization of the test for stationarity of time series presented by Priestley and Rao (1969) to spatial processes. We first need to introduce an estimate of the spatial spectral density for a nonstationary process.

1.6.1 Nonparametric estimation of a spatial spectrum

Assume we observe the process Z at N equally spaced locations in a regular grid D $(n_1 \times n_2)$, where $N = n_1 n_2$, and the spacing between observations is Δ. In this Section we propose a nonparametric estimate of the spectral density (in a neighborhood of $\mathbf{x}$), $f_{\mathbf{x}}$. We allow the spectrum to vary from part to part of our domain, by having a spectral density that is a function of location. This estimate is simply a spatial periodogram with a filter function to give more weight to neighboring values of $\mathbf{x}$. We first define $J_{\mathbf{x}}(\boldsymbol{\omega}_0)$,

$$J_{\mathbf{x}}(\boldsymbol{\omega}_0) = \Delta \sum_{u_1=0}^{n_1} \sum_{u_2=0}^{n_2} g(\mathbf{x} - \Delta\mathbf{u}) Z(\Delta\mathbf{u}) \exp\{-i\Delta\mathbf{u}^T\boldsymbol{\omega}_0\}, \tag{34}$$

where $\mathbf{u} = (u_1, u_2)$, and $\{g(\mathbf{s})\}$ is a filter satisfying the following conditions

B.1 $\{g(\mathbf{s})\}$ is square integrable and normalized filter, so that

$$(2\pi)^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |g(\mathbf{s})|^2 d\mathbf{s} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\Gamma(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} = 1$$

Here

$$\Gamma(\boldsymbol{\omega}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(\mathbf{s}) \exp\{-i\mathbf{s}^T\boldsymbol{\omega}\} d\mathbf{s}$$

denotes the Fourier transform of $\{g(\mathbf{s})\}$.

B.2 $\{g(\mathbf{s})\}$ has finite "width" B_g defined by

$$B_g = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\mathbf{s}||g(\mathbf{s})| d\mathbf{s}$$

where B_g is smaller than B_Z.

We refer to $|J_{\mathbf{x}}(\boldsymbol{\omega})|^2$ as the spatial periodogram at a location $\mathbf{x}$ for a frequency $\boldsymbol{\omega}$.

The spectral estimate $|J_{\mathbf{x}}(\boldsymbol{\omega})|^2$ is an approximately unbiased estimate of $f_{\mathbf{x}}(\boldsymbol{\omega})$ (see Theorem 1 in Fuentes (2005a)), but as its variance may be shown to be independent of N it will not be a very useful estimate in practice. We therefore estimate $f_{\mathbf{x}}(\boldsymbol{\omega})$ by "smoothing" the values of $|J_{\mathbf{x}}(\boldsymbol{\omega})|^2$ over neighboring values of $\mathbf{x}$. More precisely, let W_ρ be a weight function or "window", depending on the parameter ρ, which has integral 1 and the squared of W is integrable.

We write

$$w_\rho(\boldsymbol{\lambda}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\{i\mathbf{s}^T\boldsymbol{\lambda}\} W_\rho(\mathbf{s}) d\mathbf{s}$$

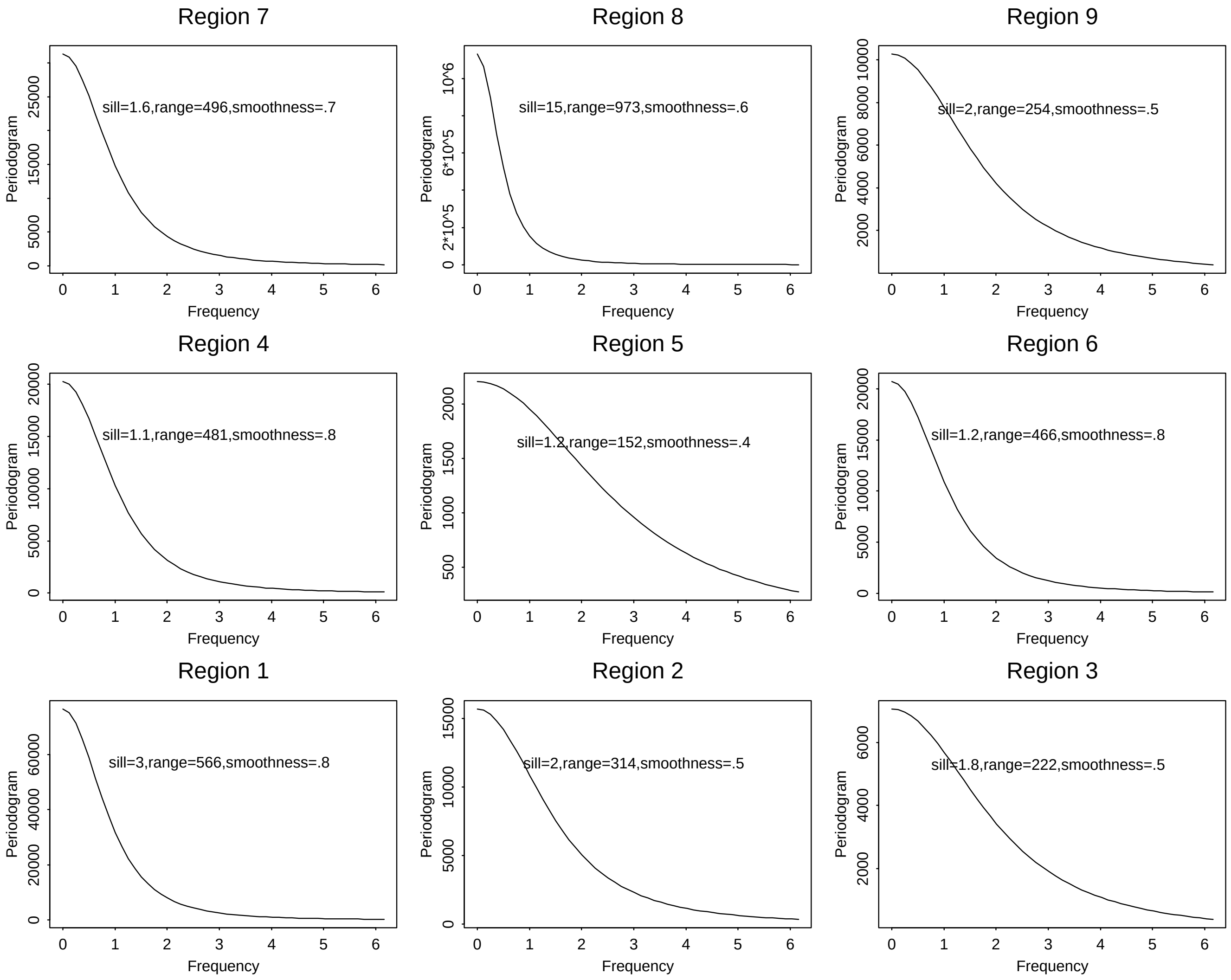


Figure 10: This figure shows the Matérn spectral densities, obtained using Whittle's approximation to the likelihood, for the 9 equally-dimensioned regions shown in Figure 1. Region 1 represents the lower left subarea (TX) of Figure 1.

We assume that there exists a constant C such that

$$\lim_{\rho\to\infty}\left\{\rho^2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|w_\rho(\boldsymbol{\lambda})|^2 d\boldsymbol{\lambda}\right\}=C$$

The parameter ρ determines the bandwidth of $\{W_\rho\}$ and it is chosen, such that it is larger than the width B_g.

Then, we estimate $f_{\mathbf{x}}(\boldsymbol{\omega}_0)$ by

$$\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}W_\rho(\mathbf{x}-\mathbf{s})\,|J_{\mathbf{s}}(\boldsymbol{\omega}_0)|^2\,d\mathbf{s} \tag{35}$$

We evaluate (35) by Monte Carlo integration.

In the application presented in this Section we consider $\{g(\mathbf{s})\}$ for $\mathbf{s}=(s_1,s_2)$ to be a multiplicative filter, i.e. the tensor product of two one-dimensional filters, $g(\mathbf{s})=g_1(s_1)g_1(s_2)$, where g_1 is of the form

$$g_1(s)=\begin{cases}1/\{2\sqrt{h\pi}\} & |s|\le h\\ 0 & |s|>h\end{cases} \tag{36}$$

corresponding to the Barlett window. Then, $\Gamma(\boldsymbol{\omega})=\Gamma_1(\omega_1)\Gamma_1(\omega_2)$ for $\boldsymbol{\omega}=(\omega_1,\omega_2)$. We also choose W_ρ to be of the form $W_\rho(\mathbf{s})=W_{1,\rho}(s_1)W_{1,\rho}(s_2)$, where

$$W_{1,\rho}(s)=\begin{cases}1/\rho & -1/2\rho\le s\le 1/2\rho\\ 0 & \text{otherwise}\end{cases} \tag{37}$$

corresponding to the Daniell window.

The asymptotic properties of $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)$ using a *shrinking asymptotics* model are studied by Fuentes (2005a).

1.6.2 Testing for stationarity

We calculate our estimate of the spatial spectral density $f_{\mathbf{s}_i}(\boldsymbol{\omega})$ at m nodes $\mathbf{s}_1,\ldots,\mathbf{s}_m$ that constitute a systematic sample on D.

We write

$$U(\mathbf{s}_i,\boldsymbol{\omega})=\log\hat{f}_{\mathbf{s}_i}(\boldsymbol{\omega}_j=\log f_{\mathbf{s}_i}(\boldsymbol{\omega})+\epsilon(\mathbf{s}_i,\boldsymbol{\omega}).$$

We obtain that asymptotically $E(\epsilon(\mathbf{s}_i,\boldsymbol{\omega}))=0$ and $\mathrm{var}\{\epsilon(\mathbf{s}_i,\boldsymbol{\omega})\}=\sigma^2$ where

$$\sigma^2=(C/\rho^2)\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|\Gamma(\boldsymbol{\theta})|^4 d\boldsymbol{\theta}\right\}, \tag{38}$$

for $\boldsymbol{\omega}\notin\partial\Pi_\Delta^2$, where $\Pi_\Delta=[-\pi/\Delta,\pi/\Delta]$ and $\partial\Pi_\Delta^2$ denotes the boundary of the region Π_Δ^2. The variance σ^2 is clearly independent of $\mathbf{x}$, and $\boldsymbol{\omega}$.

Now, we evaluate the estimated spatial spectra, $\hat{f}_{\mathbf{s}_i}(\boldsymbol{\omega})$, at the m nodes $\mathbf{s}_1,\ldots,\mathbf{s}_m$ and a set of frequencies $\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\ldots,\boldsymbol{\omega}_n$ that cover the range of locations and frequencies of interest. Assuming the $\mathbf{s}_i$ and $\boldsymbol{\omega}_j$ are spaced "sufficiently wide apart," then the $\epsilon(\mathbf{s}_i,\boldsymbol{\omega}_j)$ will be approximately uncorrelated. This result is based on asymptotic properties of $\hat{f}_{\mathbf{s}_i}(\boldsymbol{\omega}_j)$. The spatial periodogram values

$\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ and $\hat{f}_{\mathbf{y}}(\boldsymbol{\omega}')$ are asymptotically uncorrelated, if either $\|\boldsymbol{\omega} \pm \boldsymbol{\omega}'\| \gg$ bandwidth of $|\Gamma(\boldsymbol{\theta})|^2$ or $\|\mathbf{x} - \mathbf{y}\| \gg$ bandwidth of the function $\{W_\rho(\mathbf{u})\}$. In the application presented in this Section we used the distance between the "half-power" points on the main lobe of $|\Gamma(\boldsymbol{\omega})|^2$ to approximate the bandwidth.

The logarithmic transformation brings the distribution of a smoothed spatial periodogram closer to normality (Jenkins, 1961). Thus, we may treat the $\epsilon(\mathbf{x}_i, \boldsymbol{\omega}_j)$ as independent $N(0, \sigma^2)$. We write

$$U_{ij} = U(\mathbf{s}_i, \boldsymbol{\omega}_j),\ \log f_{\mathbf{s}_i}(\boldsymbol{\omega}_j) = f_{ij},\ \text{ and } \epsilon_{ij} = \epsilon(\mathbf{s}_i, \boldsymbol{\omega}_j).$$

Then we have the model

$$U_{ij} = f_{ij} + \epsilon_{ij}. \tag{39}$$

Equation (39) becomes the usual "two-factor analysis of variance" model, and could be rewritten in the more conventional form:

$$H_1 : U_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ij}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Then, we test for stationarity of Z by using the standard techniques to test the model

$$H_0 : U_{ij} = \mu + \beta_j + \epsilon_{ij}$$

against the more general model H_1. Since we know the value of $\sigma^2 = \text{var}\{\epsilon_{ij}\}$, we can test for the presence of the interaction term, γ_{ij}, with one realization of the process. If the model H_0 is rejected, then there is a significant difference between the parameters α_i, for $i = 1, \ldots, m$, which is evidence of lack of stationarity for Z at the m nodes. Thus, the complex and challenging problem of testing for nonstationarity is reduced to a simple two-factor analysis of variance.

The parameters $\{\alpha_i\}$, $\{\beta_j\}$ represent the main effects of the space and frequency factors, and $\{\gamma_{ij}\}$ represents the interaction between these two factors. A test for the presence of interaction is equivalent to testing if Z is a uniformly modulated process, this means $\log f_{\mathbf{x}}(\boldsymbol{\omega})$ is additive in terms of space and frequency, then $f_{\mathbf{x}}(\boldsymbol{\omega})$ is multiplicative, i.e. $f_{\mathbf{x}}(\boldsymbol{\omega}) = c^2(\mathbf{x}) f(\boldsymbol{\omega})$, so the process Z is of the form: $Z(\mathbf{x}) = c(\mathbf{x}) Z_0(\mathbf{x})$, where Z_0 is stationary with spectral function f and c is a function of space. If the interaction is not significant, we conclude that Z is a uniformly modulated process. If the interaction is significant, we conclude that Z is non-stationary, and non-uniformly modulated. We can study if the non-stationarity of Z is restricted only to certain frequency components, by selecting those frequencies, e.g. $\{\boldsymbol{\omega}_{j_1}, \ldots, \boldsymbol{\omega}_{j_k}\}$ and testing for stationarity at these frequencies.

If Z is an isotropic process, then $f_{\mathbf{x}}(\boldsymbol{\omega})$ depends on its vector argument $\boldsymbol{\omega}$ only through its length $\|\boldsymbol{\omega}\|$. Then, we could test for isotropy by selecting a set of frequencies with the same absolute values, say $\{\boldsymbol{\omega}_{j_1}, \boldsymbol{\omega}_{j_2}\}$ where $\boldsymbol{\omega}_{j_1} \neq \boldsymbol{\omega}_{j_2}$ but $\|\boldsymbol{\omega}_{j_1}\| = \|\boldsymbol{\omega}_{j_2}\|$, and examine whether the "main-effect" effect β is significant.

We could test for "complete randomness", this means constant spectra for the spectral density on the lattice, by testing the "main-effect" β, either at all locations on the lattice when the interaction term is not significant, or at a particular subset of locations. All these comparisons are based on a χ^2 rather than F-tests, because σ^2 is known.

1.6.3 An example in air quality

For the purpose of illustrating the techniques presented in this Section, we take a systematic sample of locations in D, the sample points $\mathbf{x}_1, \ldots, \mathbf{x}_9$ are the centroids of the 9 equally-dimensioned regions $S_1, \ldots, S_9$, in our domain of interest (Figure 9).

$\mathbf{x} \backslash \boldsymbol{\omega}$	$\boldsymbol{\omega}_1$	$\boldsymbol{\omega}_2$	$\boldsymbol{\omega}_3$	$\boldsymbol{\omega}_4$	$\boldsymbol{\omega}_5$	$\boldsymbol{\omega}_6$	$\boldsymbol{\omega}_7$	$\boldsymbol{\omega}_8$	$\boldsymbol{\omega}_9$
$\mathbf{x}_1$	4315.92	31.49	5.35	28.01	9.18	0.20	3.89	5.12	1.02
$\mathbf{x}_2$	3376.27	35.46	3.30	35.81	8.73	4.67	4.58	6.29	4.31
$\mathbf{x}_3$	2670.07	38.01	6.54	40.99	12.38	10.68	7.36	2.86	2.14
$\mathbf{x}_4$	1617.05	13.28	3.52	14.90	2.98	2.08	1.24	4.91	3.20
$\mathbf{x}_5$	1256.20	38.80	5.69	30.58	13.56	3.05	9.66	3.30	2.64
$\mathbf{x}_6$	1765.69	14.20	0.52	13.93	7.29	1.45	1.16	2.36	4.79
$\mathbf{x}_7$	2016.57	12.97	2.55	17.88	0.69	3.27	0.37	0.13	2.21
$\mathbf{x}_8$	13597.65	70.37	10.67	75.72	23.63	8.20	12.55	6.89	6.36
$\mathbf{x}_9$	4618.01	63.28	12.09	56.93	21.94	7.71	10.32	1.23	1.21

Table 1. Values of $\hat{f}_\mathbf{x}(\boldsymbol{\omega})$.

Now, we implement our test for stationarity. We select values of locations $\mathbf{x}$ and frequencies $\boldsymbol{\omega}$ that are sufficiently apart. The estimates, $\hat{f}_\mathbf{x}(\boldsymbol{\omega})$ (Table 1) were obtained using expression (35) in which $W_\rho(\mathbf{u})$ is given by (37) with $\rho = 20$ units (1 unit = 36 km), and $g(\mathbf{u})$ is of the form (36) with $h = 3$. The window $|\Gamma(\boldsymbol{\omega})|^2$ has a bandwidth of approximately $\pi/h = \pi/3$. The distance between the "half-power" points on the main lobe of $|\Gamma(\boldsymbol{\omega})|^2$ was used as an approximation of the bandwidth. The window $\{W_\rho(\mathbf{x})\}$ has a bandwidth of $\rho = 20$. Thus, in order to obtain approximately uncorrelated estimates, the points $\boldsymbol{\omega}_j$ and $\mathbf{x}_i$ should be chosen so that the spacings between the $\boldsymbol{\omega}_j$ are at least $\pi/3$ and the spacings between the $\mathbf{x}_\mathbf{i}$ are at least 20 units, the sample points $\mathbf{x}_1, \ldots, \mathbf{x}_9$ are the centroids of the 9 equally-dimensioned regions $S_1, \ldots, S_9$, covering the domain shown in Figure 1.4.4. The $\boldsymbol{\omega}_j$ were chosen as follows $\boldsymbol{\omega}_j = (\omega_{j_1}, \omega_{j_2}) = (\pi j/20, \pi j_2/20)$ with $j_1 = 1\,(7)\,15$, $j_2 = 1\,(7)\,15$, corresponding to a uniform spacing of $7\pi/20$ (which just exceeds $\pi/3$). The values of $\hat{f}_\mathbf{x}(\boldsymbol{\omega})$ are shown in Table 1, where $\boldsymbol{\omega}_1 = (\pi/20, \pi/20)$, $\boldsymbol{\omega}_2 = (8\pi/20, \pi/20)$, $\boldsymbol{\omega}_3 = (15\pi/20, \pi/20)$, $\boldsymbol{\omega}_4 = (\pi/20, 8\pi/20)$, $\boldsymbol{\omega}_5 = (8\pi/20, 8\pi/20)$, $\boldsymbol{\omega}_6 = (15\pi/20, 8\pi/20)$, $\boldsymbol{\omega}_7 = (\pi/20, 15\pi/20)$, $\boldsymbol{\omega}_8 = (8\pi/20, 15\pi/20)$, and $\boldsymbol{\omega}_9 = (15\pi/20, 15\pi/20)$.

We need to calculate σ^2 (see Equation (38)) to perform the test of stationarity for Z. In this application $\sigma^2 = 16h^2/(9\rho^2) = 0.04$.

Item	Degrees of freedom	Sum of squares	χ^2 = (sum of squares/σ^2)
Between spatial points	8	26.55	663.75
Between frequencies	8	366.84	9171
Interaction + residual	64	30.54	763.5
Total	80	423.93	10598.25

Table 2. Analysis of variance.

The interaction is significant (χ^2 is very large compared to $\chi^2_{64}(0.05) = 83.67$) confirming that we do not have a uniformly modulated model, and both the "between spatial points" and "between frequencies" sums of squares are highly significant (χ^2 is extremely large compared to $\chi^2_8(0.05) = 15.51$), confirming that the process is non-stationary and that the spectra are non-uniform.

If isotropy is a reasonable assumption then columns 3 and 7 in Table 1 should have similar values. We present now an approach to test if there is any significant difference between columns 3 and 7 in Table 1.

The next table (Table 3) presents an analysis of variance to study the significance of the difference between columns 3 and 7 in Table 1. The spatial points are the same as in Table 2, $\mathbf{x_1}, \ldots, \mathbf{x_9}$, and the frequency values used here are $\boldsymbol{\omega}_3$ and $\boldsymbol{\omega}_7$, both having the same absolute value.

Item	Degrees of freedom	Sum of squares	χ^2 = (sum of squares/σ^2)
Between spatial points	8	15.96	399.18
Between frequencies	1	0.12	3.17
Interaction + residual	8	2.87	71.99
Total	17	18.95	473.75

Table 3. Analysis of variance.

The "between frequencies" effect is not significant (χ^2 is smaller than $\chi^2_1(0.05) = 3.84$), suggesting that then there is no evidence of anisotropy. This is not surprising, since for air pollution the lack of anisotropy is usually detected at higher spatial resolutions (here the resolution of the models is 1296 km^2.) However, the "between spatial locations" sums of squares is highly significant (χ^2 is extremely large compared to $\chi^2_8(0.05) = 15.51$), confirming that the process is non-stationary.

We could test for stationarity within the subregions $S_1, \ldots, S_9$, by drawing a larger systematic sample in D with more than one sample point within each subregion. Further testing suggests that $\boldsymbol{\theta}(\mathbf{x})$ does not change significantly within the subregions $S_1, \ldots, S_9$.

2 Wavelet analysis

Wavelet analysis, as opposed to frequency analysis, attempts to decompose the variability of a process both in time and frequency. It is generally an orthogonal expansion. In this section we first describe the continuous and discrete wavelet transforms, following Percival and Walden (Percival *et al.* (2000)). While wavelets can be applied to spatial fields, we focus first for simplicity on a description for a temporal process. Then we describe how one can use the discrete wavelet transform to estimate temporal trends, and derive confidence bands and formal tests. Finally we look at a wavelet transform in the plane, and show how it can be used to create a nonstationary spatial covariance function.

2.1 The continuous wavelet transform

Consider a real-valued time series $x(t)$. For a scale λ and a time t look at the average

$$A(\lambda, t) = \frac{1}{\lambda} \int_{t-\lambda/2}^{t+\lambda/2} x(u)du$$

Essentially, wavelet analysis looks at how much such averages change over time. Let

$$\begin{aligned} D(\lambda,t) &= A(\lambda,t+\tfrac{\lambda}{2}) - A(\lambda,t-\tfrac{\lambda}{2}) \\ &= \tfrac{1}{\lambda}\int\limits_{t}^{t+\lambda} x(u)du - \tfrac{1}{\lambda}\int\limits_{t-\lambda}^{t} x(u)du \end{aligned}$$

The difference D at scale $\lambda = 1$ and time $t = 0$ can be written

$$D(1,0) = \sqrt{2}\int\limits_{-\infty}^{\infty} \psi^{(H)}(u)x(u)du$$

where

$$\psi^{(H)}(u) = \begin{cases} -\frac{1}{\sqrt{2}}, & -1 < u \leq 0 \\ \frac{1}{\sqrt{2}}, & 0 < u \leq 1 \\ 0, & otherwise \end{cases}$$

is the Haar wavelet. We get a family of Haar wavelets by translation

$$\psi^{(H)}_{1,t}(u) = \psi^{(H)}(u-t)$$

and scaling

$$\psi^{(H)}_{\lambda,t}(u) = \frac{1}{\sqrt{\lambda}}\psi^{(H)}(\frac{u-t}{\lambda}).$$

Then the continuous Haar wavelet transform of the process $x(t)$ is

$$\widehat{W}(\lambda,t) = \int\limits_{-\infty}^{\infty} \psi^{(H)}_{\lambda,t}(u)x(u)du \propto D(\lambda,t).$$

Figure 2.1 shows the signal $x(t)$, the wavelet $\psi^{(H)}(t)$ and the integrand in the wavelet transform (i.e., the product of the signal and the wavelet).

A general (mother) wavelet is a function which integrates to zero and square integrates to one. A family of wavelets is generated by translation and scaling; $\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi((x-b)/a)$. There is an additional technical constraint, namely that

$$C_\psi = \int |u|^{-1}|\mathcal{F}(\psi)(u)|^2 du.$$

A few different classes of wavelets are given in a later subsection. The general continuous wavelet transform for a wavelet ψ is

$$\hat{W}(\lambda,t) = \int\limits_{-\infty}^{\infty} \psi_{\lambda,t}(u)x(u)du$$

.

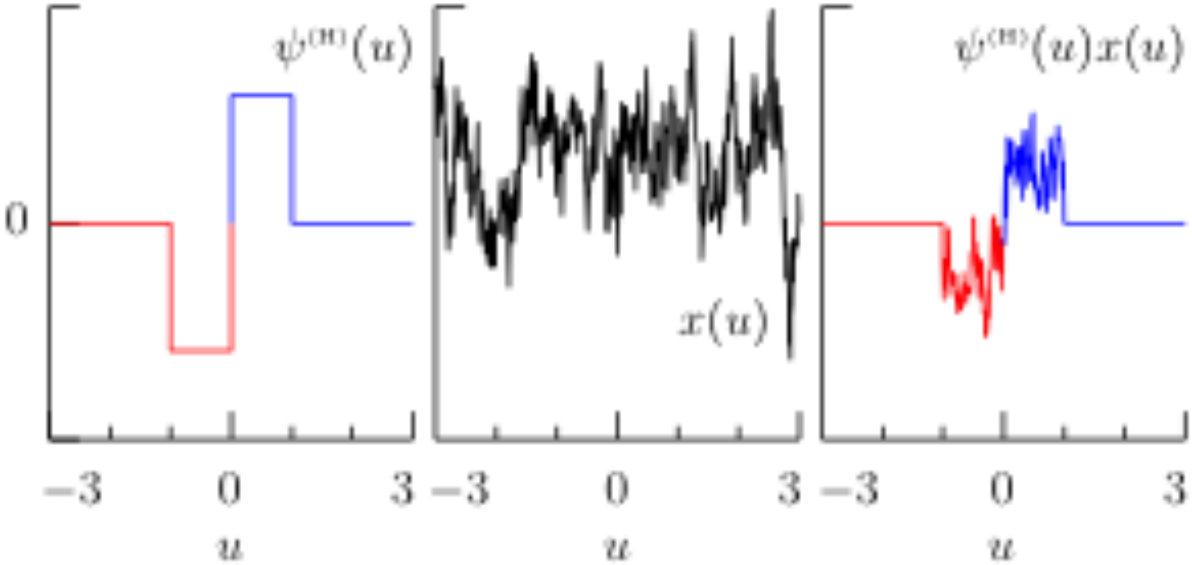


Figure 11: The Haar wavelet (left), a time series (middle) and the product of the two (right). The latter is what is integrated to calculate the wavelet transform at the scale of this wavelet.

The continuous wavelet transform is equivalent to the original process x. We get the process back from the transform by the formula

$$x(t) = \frac{1}{C_\psi} \int_0^\infty \left[\int_{-\infty}^\infty \hat{W}(\lambda, u)\psi_{\lambda,t}(u)du \right] \frac{d\lambda}{\lambda^2}$$

where C_ψ is the finite constant given above.

Like the Fourier transform, the wavelet transform decomposes energy, in the sense (for a mean zero process x) that

$$\int_{-\infty}^\infty x^2(t)dt = \int_0^\infty \int_{-\infty}^\infty \frac{\hat{W}^2(\lambda, t)}{C_\psi \lambda^2} dt d\lambda.$$

Here we think of the lefthand side as the total energy, while the integrand on the righthand side corresponds to the energy at time t and frequency λ.

2.2 The discrete wavelet transform

Usually we do not get to observe our process $x(t)$ continuously. Rather, we observe it at discrete times (for spatial processes, at a grid). Hence we can consider dyadic scales $\lambda = \tau_j = 2^{-j}$ and times restricted to the integers. The discrete wavelet transform then has coefficients $W_{j,k} \propto \hat{W}(\tau_j, k)$. We write $\mathbf{W} = (W_{j,n})$ for the wavelet coefficients, and $\mathcal{W}$ for the wavelet transform, so

$$\mathbf{W} = \mathcal{W}\mathbf{Y}.$$

In many cases we want $(\tau_{j,n})_{n=1}^{\infty}$ be an orthogonal basis for each j, so that $\mathcal{W}^{-1} = \mathcal{W}^T$. This puts some further restrictions on the choice of wavelets, as we will see in the next subsection.

Following Tukey, we can think of $r_n = \sum_{j=1}^{J} W_{j,n}$ as the *rough* of the series, while $s_n = x_n - r_n$ is the *smooth.* A *multiwavelet analysis* shows the *details* or the wavelet coefficients for each scale $j < \log_2(N)$ where N is the number of observation, as well as the smooth. Each of the details describe the variability of the time series at the corresponding scale.

Example

Cole, Dunbar, McClanahan, and Muthiga (Cole *et al* (2000)) studied a 194 year record of the $\delta^1 8$O oxygen isotope measured from a 4m high coral colony growing at a depth of 6m (at low tide) in Malindi, Kenya. The purpose of the study was to measure changes in sea surface temperature over time.

A decrease in the oxygen value corresponds to an increase in the sea surface temperature (SST) (roughly a change of -0.24 pppm concentration corresponds to an SST increase of 1°C). The issue is whether there is a significant decadal trend after we adjust for the variability in the process. Figure 2.2 shows a multiscale analysis of the negative of these data, computed using the pyramid scheme (Mallat, Mallat (1989)) with the LA8 wavelet (see next subsection). We see a tendency towards increase in the smooth of the series towards the end of the period, corresponding to a possible increase in SST.

2.3 Some wavelets

The requirement that a wavelet needs to generate an orthogonal basis affects what functions one can use. There is of course a variety of basis functions. For computational reasons it is desirable that the wavelet has compact support. It turns out that the orthogonality requirement and the compact support together force the wavelet to be asymmetric (if it is not the Haar wavelet). In this subsection we first present a wavelet family without compact support, and then a family with compact support.

2.3.1 Mexican hat

The Mexican hat is proportional to the second derivative of the Gaussian density function, i.e., $\psi(x) \propto (1 - x^2)\exp(-x^2/2)$. Its name comes from the effect of rotating this function around the origin, thus creating a two-dimensional wavelet that has been popular in image analysis. Figure 2.3.1 shows this wavelet.

2.3.2 Daubechies' least asymmetric compactly supported wavelet

The least asymmetric compactly supported wavelet can be written as a low-pass filter supported on $2L$ points. The calculation of the wavelet is done according to a recipe in Daubechies (Daubechies (1992)). Figure 2.3.2 shows the wavelet corresponding to $L = 8$, commonly called the LA(8)-wavelet. It is the default wavelet in software packages such as waveslim in R (R Development Core Team (2004)).

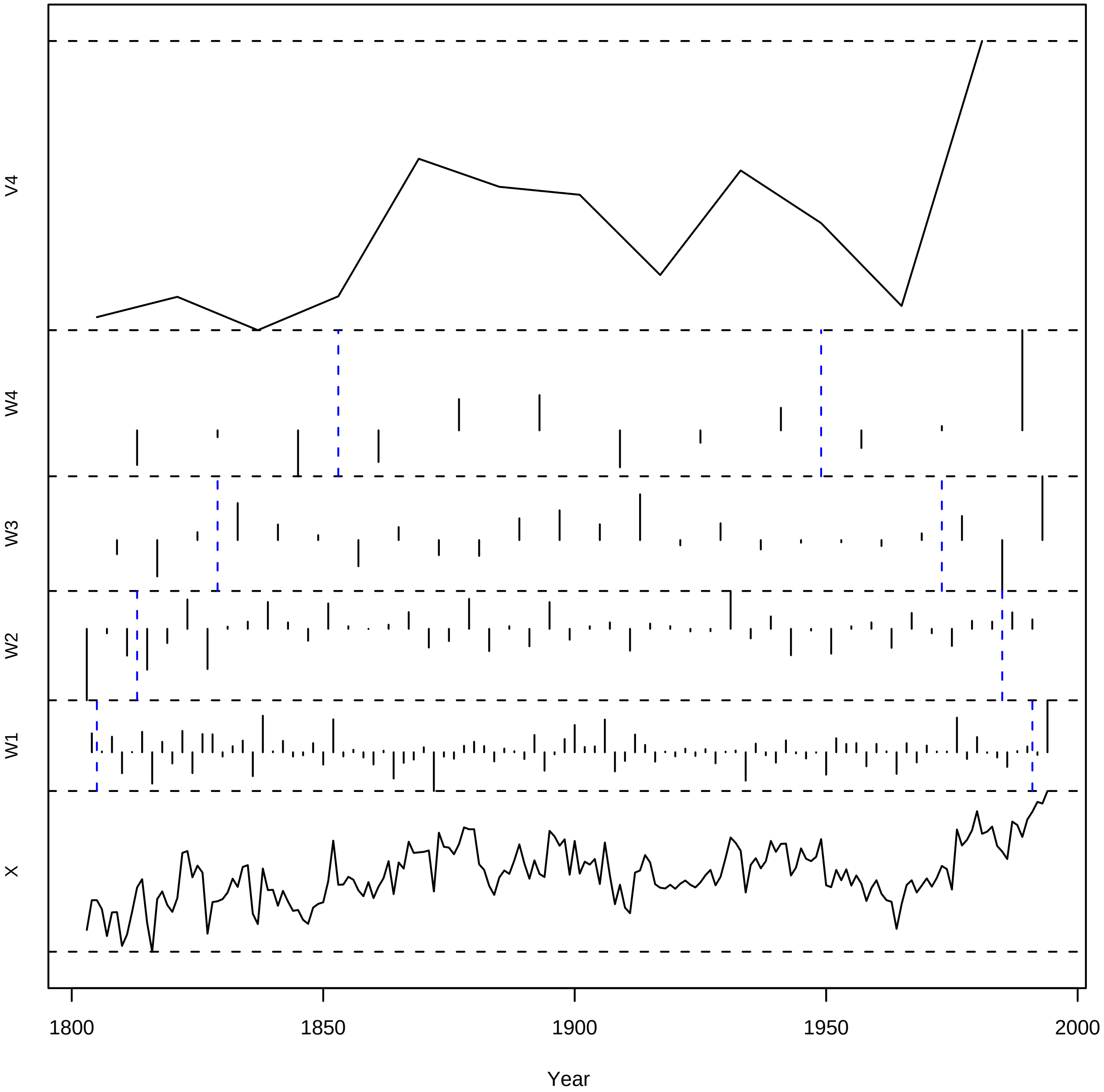
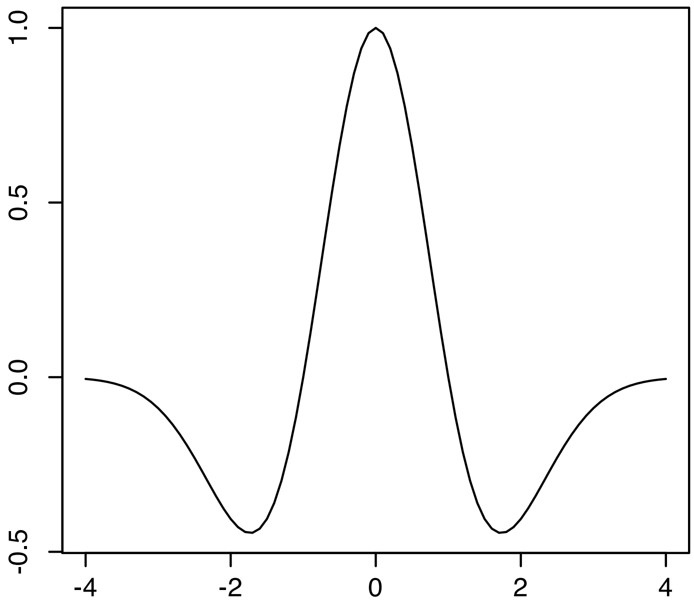
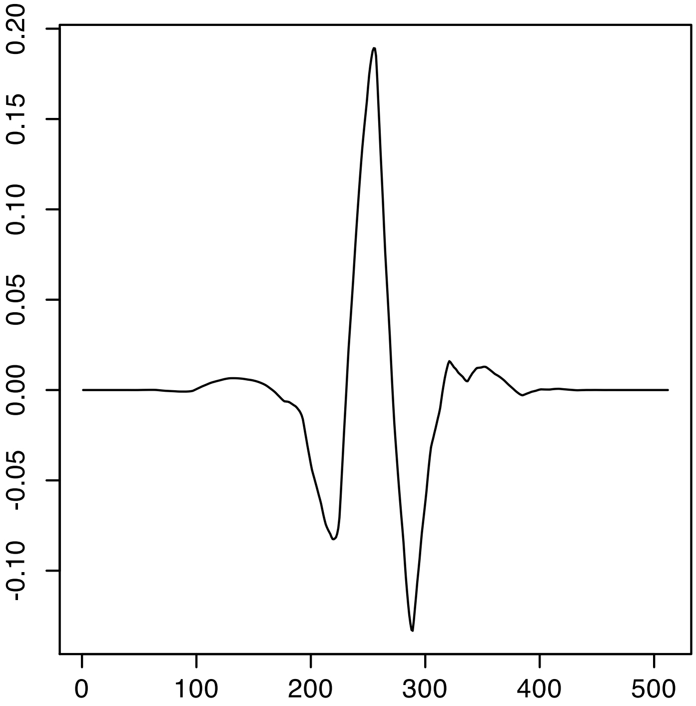


Figure 12: Multiscale analysis of the Malindi time series of oxygen isotope. The smooth is on top, followed successively by the wavelet coefficients at scale 16, 8, 4 and 2 years. The bottom graph shows the time series.


1.0
0.5
0.0
-0.5
-4
-2
0
2
4


0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
0
100
200
300
400
500

2.4 A nonparametric trend estimator and its properties

The scaling part of the wavelet transform is in essence a nonparametric trend estimator, where the trend corresponds to variability at a larger scale than what is captured in the details. In this subsection we outline how Craigmile et al. (Craigmile *et al.* (2004)) have developed tests for polynomial trends and simultaneous confidence bands for a trend estimator which is related to the scaling part of the wavelet transform. The statistical properties are valid even when the underlying time series exhibits long term memory. We consider the usual model $X_t = T_t + Y_t$ where $\mathbf{T} = (\mathbf{T_t})$ is the trend, assumed non-stochastic, and $\mathbf{Y} = (\mathbf{Y_t})$ is the error process.

2.4.1 Fractional difference model for long term memory

Many atmospheric and climatological time series exhibit *long term memory*, i.e., the autocovariance function decays very slowly. This creates difficulties for Fourier analysis, in that the convenient lack of correlation between periodogram values at different Fourier frequencies no longer holds, but turns out to be handled rather well by wavelet analysis.

We will focus here on a simple class of Gaussian long term memory models, namely the *fractional difference* class. It is a mean zero Gaussian time series, with Fourier density

$$f(\omega) = \sigma^2 \left|2\sin(\pi\omega)\right|^{-2\delta}. \tag{40}$$

One can think of this as an ARIMA(0,δ,0)-process, where δ is not necessarily an integer. The parameter can be estimated from the time series (see e.g. Craigmile *et al.* (2004)). General description and theory of long term memory processes can be found in Beran's book (Bera (1994)).

Using the representation 40, and taking logarithms on both sides, we see that for small values of ω we have

$$\log f(\omega) \approx c - 2\delta \log(\pi\omega)$$

so if we plot $\log f$ agains $\log \omega$ we should get a straight line for small values of ω. Figure 2.4.1 shows this plot for the Malindi data.

2.4.2 The trend estimator

Using one of Daubachies' wavelets with compact support, we can divide the wavelet coefficients into those that are affected by boundary values and those that are not. In fact, we can write

$$\mathbf{W} = \mathbf{W_s} + \mathbf{W_b} + \mathbf{W_{nb}}$$

where $\mathbf{W_s}$ has the scaling coefficients and zeros elsewhere, $\mathbf{W_b}$ has the wavelet coefficients affected by the boundary, and zeros elsewhere, while $\mathbf{W_{nb}}$ are the non-boundary wavelet coefficients. Noting that $\mathbf{X} = \mathcal{W}^{\mathcal{T}}\mathbf{W}$ we can write

$$\mathbf{X} = \mathcal{W}^{\mathcal{T}}(\mathbf{W_s} + \mathbf{W_b}) + \mathcal{W}^{\mathcal{T}}\mathbf{W_{nb}} = \tilde{\mathbf{T}} + \tilde{\mathbf{Y}}$$

where $\tilde{\mathbf{T}}$ is an estimate of trend and $\tilde{\mathbf{Y}}$ is a tapered estimate of error. If the true trend is polynomial of degree K, and we use a wavelet filter of length $L >= 2(K+1)$, then nonboundary wavelet coefficients will not contain the trend. We can then use $\tilde{\mathbf{Y}}$ to estimate parameters of the error process, and these will be uncorrelated with the trend estimate.

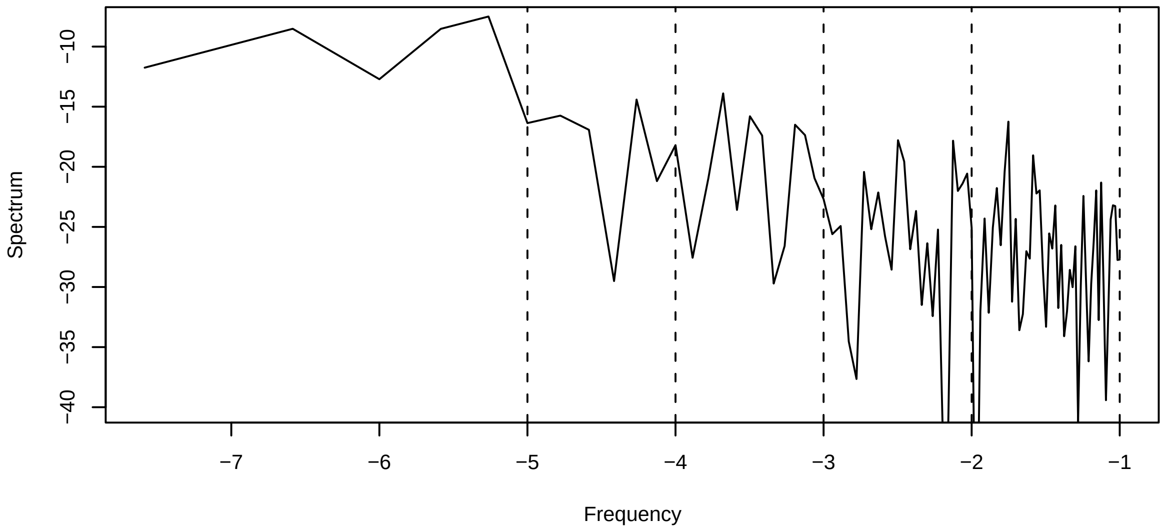


Figure 15: Spectral density function for the Malindi oxygen isotope series on a log-log scale.

2.4.3 Testing for polynomial trend

In order to test the null hypothesis of no trend ($\mathbf{T} = \mathbf{0}$ we use, for a FD error process with parameter δ, the test statistic

$$P(\delta) = \frac{\left\|\tilde{\mathbf{T}}\right\|^2}{\left\|\tilde{\mathbf{Y}}\right\|^2} = \frac{\|\mathbf{AW}\|^2}{\|(\mathbf{I}-\mathbf{A})\mathbf{W}\|^2} = \frac{\|\mathbf{X}\|^2}{\|(\mathbf{I}-\mathbf{A})\mathbf{W}\|^2} - 1$$

where $\mathbf{A}$ contains the indicators of boundary and scaling coefficients. For a given value of δ we can simulate the distribution of $P(\delta)$. In order to take into account the variability in the parameter estimates, we repeatedly simulate FD processes from random samples of the limit law of the parameter estimates, and then compute the test statistic for each simulated path. In the case of the Malindi data, the estimated parameters are $\hat{\sigma} = 0.0667$ and $\hat{\delta} = 0.359$, with a confidence band for δ of (0.143,0.597), indicating that the series does indeed have long term memory. We cannot rule out nonstationarity. Performing the simulation as outlined above, the resulting P-value is below 0.000, indicating strong evidence of a non-constant mean value (in order to look at non-zero mean, one can look at anomalies, i.e., first take out the overall mean of the series).

2.4.4 Computing simultaneous confidence band for the trend estimator

The rejection of a constant mean value only gives part of the story. Ideally we would like to derive a simultaneous confidence band for the trend. Let $\mathbf{v} = (\mathrm{var}\tilde{\mathbf{T}}_{\mathbf{0}}, ..., \mathrm{var}\tilde{\mathbf{T}}_{\mathbf{N-1}})$ and write $\mathbf{U} = \tilde{\mathbf{T}} - \mathbf{T}$. We want to find λ so that

$$1 - \alpha = P(\tilde{\mathbf{T}} - \lambda\mathbf{v} \leq \mathbf{T} \leq \tilde{\mathbf{T}} + \lambda\mathbf{v}) = \mathbf{1} - \mathbf{2P}(\mathbf{U} > \lambda\mathbf{v}).$$

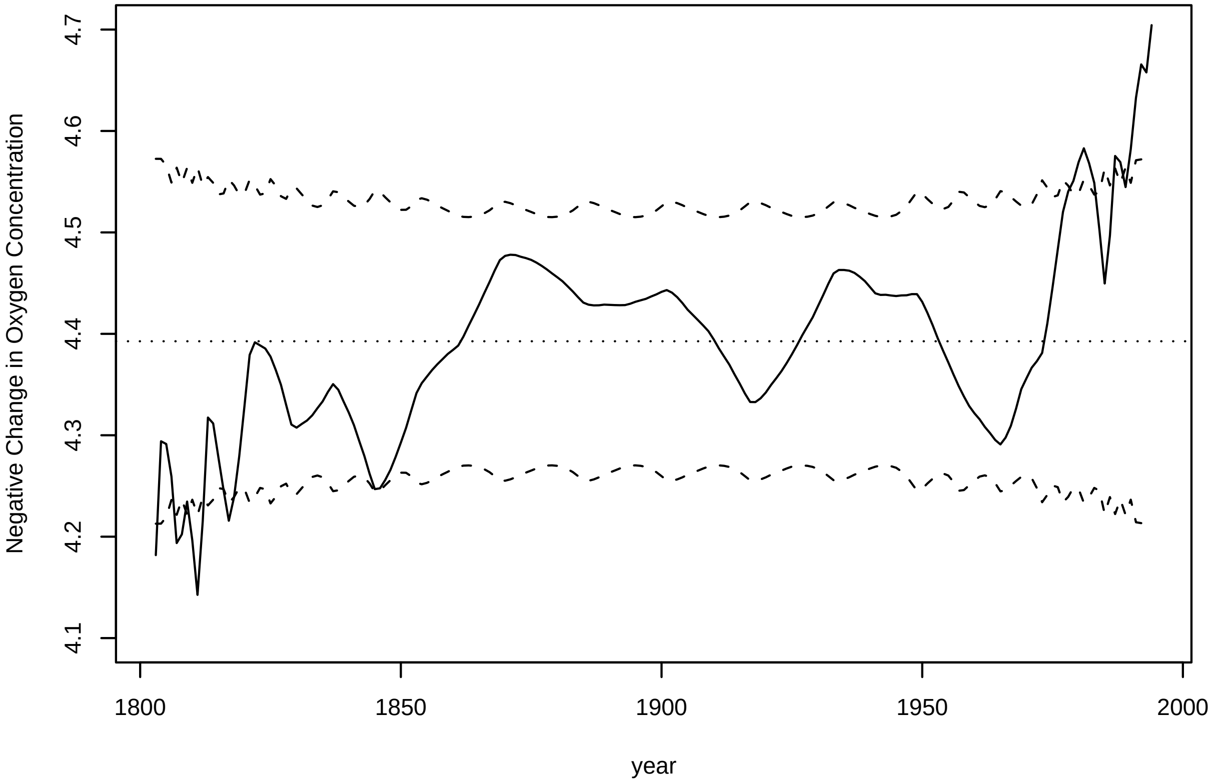


Figure 16: 95% confidence limits for the trend of the Malindi oxygen isotope series.

Given the process parameters, $\mathbf{U}$ has a known multivariate normal distribution, and conditionally we can do a bootstrap. In order to get a confidence band that takes into account the estimation of process parameters, we do again a parametric bootstrap from the limiting distribution. The result for the Malindi data is seen in Figure 2.4.4, where we see that the temperatures are high in the later years, but also somewhat low in the earlier years of the study.

2.4.5 Some spatial wavelets

In order to create spatial wavelets we can start with two wavelet funcions, the "mother wavelet" ψ and the "father wavelet" ϕ. We then create the following functions

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

where S stands for smooth, H for horizontal, V for vertical and D for diagonal. Figure 2.4.5 shows the four functions. The horizontal, vertical and diagonal are called *detail functions*.

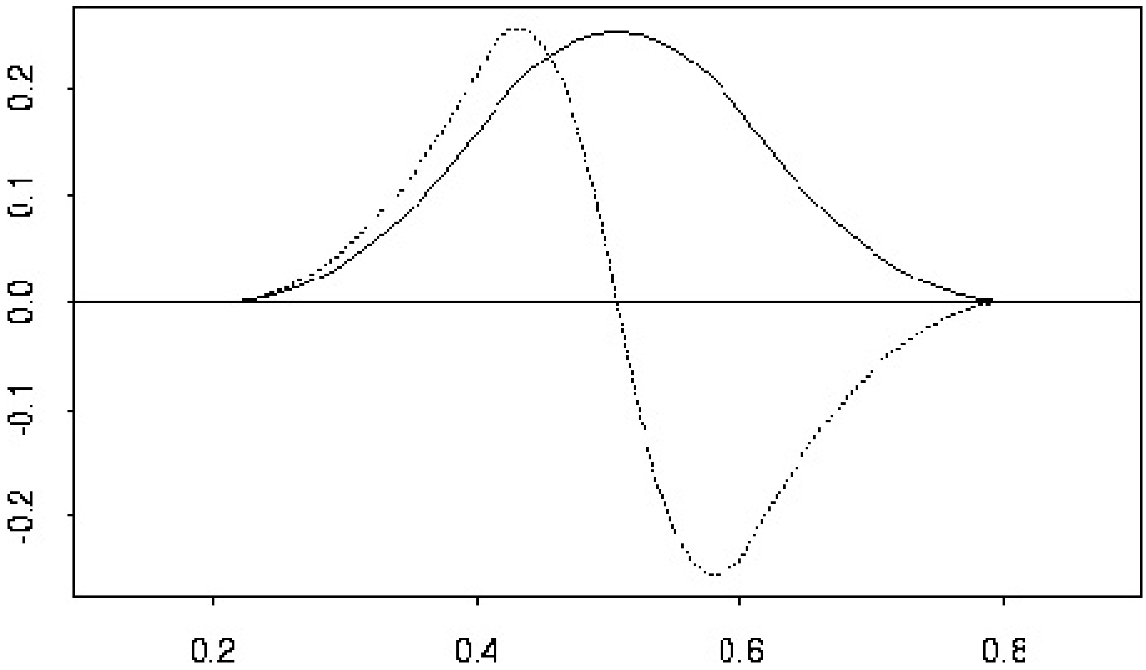


Figure 17: The four wavelet functions for spatial analysis: (a) the smooth S; (b) the horizontal detail H; (c) the vertical detail V; and (d) the diagonal detail D. From Nychka *et al*, (2002).

The key idea behind the wavelet transform 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 H, V and 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.

Technically we create a set of basis functions by starting with a coarsest level of resolution, say J. A basis of 32 functions is displayed in Figure 2.4.5. 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 2.4.5 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 2.4.5 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.

2.5 A nonstationary covariance structure

The Karhunen-Loève expansion of a Gaussian process $Z(s)$ is given by

$$Z(s) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} A_i \psi(s)$$

where

$$\begin{aligned} Cov(Z(s_1), Z(s_2)) &= C(s_1, s_2) \\ &= \sum_{i=1}^{\infty} \lambda_i \psi_i(s_1) \psi_i(s_2) \end{aligned}$$

and the A_i are iid N(0,1) random variables. Instead of using the eigenfunctions of the covariance, Nychka *et al.* (2002) suggested using the wavelet basis described in the previous subsection. This, of course, would lead to dependent coefficients A_i. Specifically, we can write the covariance

$$\mathbf{\Sigma} = \mathbf{\Psi} \mathbf{D} \mathbf{\Psi}^{\mathbf{T}}; \mathbf{D} = \mathbf{\Psi}^{-\mathbf{1}} \mathbf{\Sigma} (\mathbf{\Psi}^{\mathbf{T}})^{-\mathbf{1}}. \tag{41}$$

Letting the vector $\mathbf{\Psi}$ consist of the wavelet functions evaluated on a grid and stacked into a column vector, the matrix $\mathbf{D}$, which in the eigenfunction expansion would be diagonal, is no longer so. However, one may be able to approximate the covariance by an expansion of type 41 with a nearly diagonal $\mathbf{D}$. Something nearly diagonal will be needed, since even for a small grid $\mathbf{\Sigma}$ is huge.

Assume now that a random field Z is observed at m points on a regular grid, and that we have K independent replications. Typically m would be much larger than K. Write the data $\mathbf{Z}$ as an $m \times K$ matrix, and remove the spatial mean for each time point. Then we can estimate $\mathbf{\Sigma}$ by the sample covariance, and consequently we have $\hat{\mathbf{D}} = (1/K)(\mathbf{\Psi}^{-1}\mathbf{Z})(\mathbf{\Psi}^{-\mathbf{1}}\mathbf{Z})^{\mathbf{T}}$. It is convenient to work with the square root $\mathbf{H}$ of $\mathbf{D}$, i.e. a matrix such that $\mathbf{H}^{\mathbf{2}} = \mathbf{D}$. Using the singular value decomposition of $(\mathbf{\Psi}^{-1}\mathbf{Z})$ we can then write $\hat{\mathbf{H}} = \mathbf{V}\Lambda^{\mathbf{1/2}}\mathbf{V}^{\mathbf{T}}$, where Λ contains the singular values and $\mathbf{V}$ is the left singular vector. Numerical studies indicate that if $\hat{\mathbf{H}}$ has a small number of nonzero values one

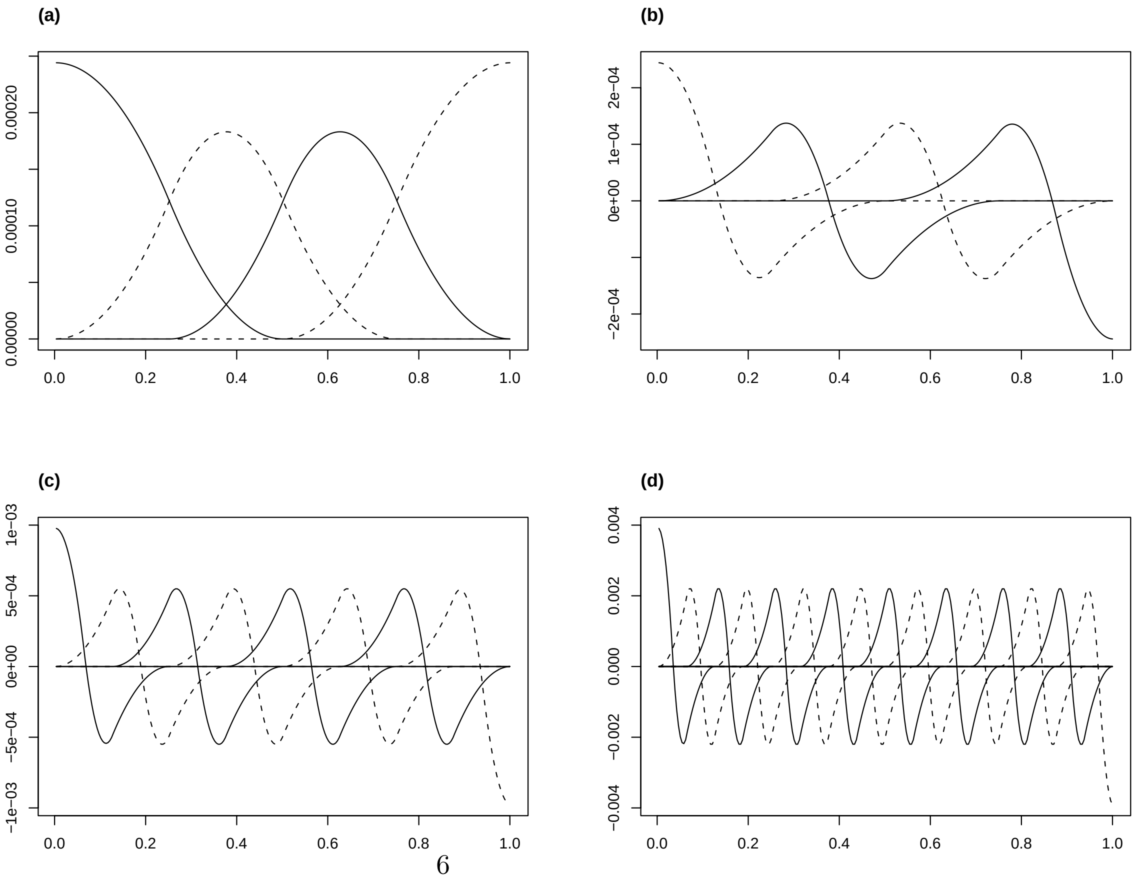


Figure 18: Family of 32 basis functions based on an approximate translation and scaling of father (a) and mother (b)-(d) wavelets. See text for details. From Nychka *et al*, (2002).

can achieve a good approximation across a family of covariance values (Nychka *et al*, (2002)). Once the nonzero elements are determined, one can further decimate them by smoothing over spatially adjacent values.

Example

To illustrate this process, Nychka *et al.* (2002) apply it to the output of a model for tropospheric ozone, the regional oxidant model (ROM), studied on a 48 by 48 grid over Illinois and Ohio in the summer of 1987. Each grid square is 16 km by 16 km, and since the correlation of ozone decays at about 300 km, we start with a 3 by 3 grid father wavelet at the coarsest resolution. In this application the leading 12 by 12 block of $\hat{\mathbf{H}}$ was decimated by 90%, and the diagonal elements were retained for the remaining levels. The resulting covariance structure is shown in Figure 2.5, and is highly nonstationary. This shows isocovariance curves for four different locations (each marked by an x). In particular, we notice that the sites (a) and (b) have covariances that are longer ranged to the west, presumably related to the dominant meteorological structure of weather systems moving from west to east.

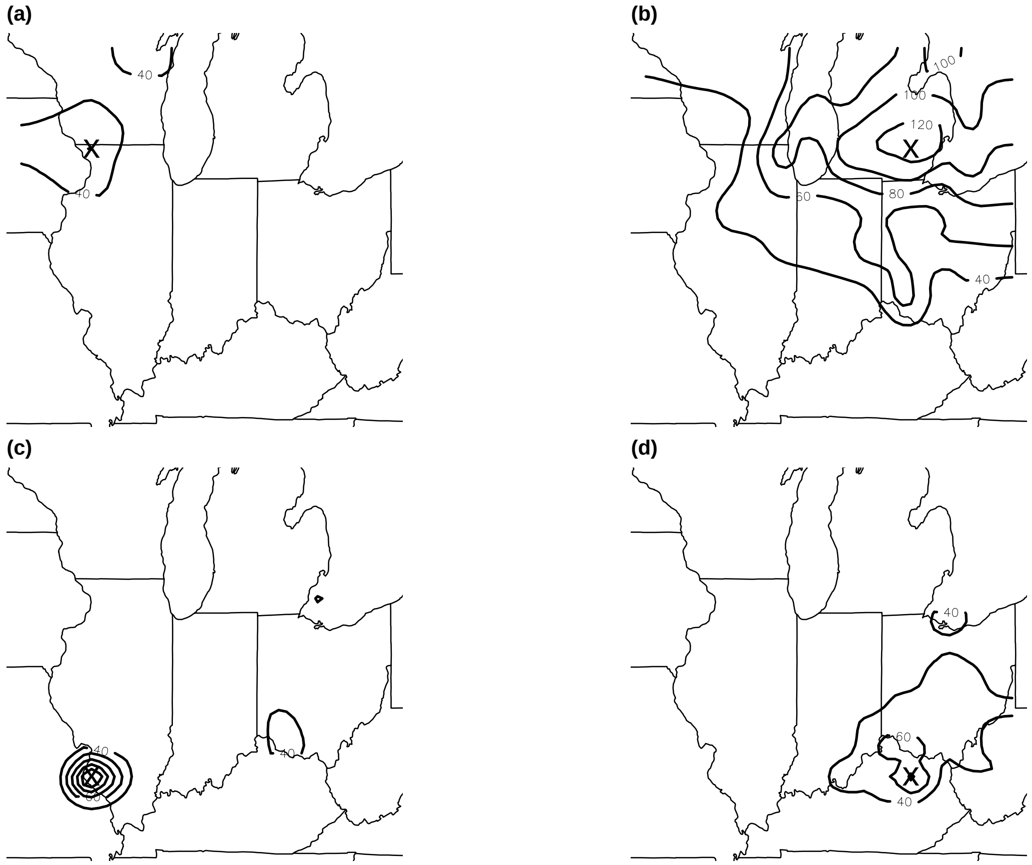


Figure 19: Estimated covariance surface at 4 sample locations for the ROM output. The image plots indicate the estimated covariance between points in the domain and the point location denoted by an x. Contour levels are at (40, 60, 80, 100, 120). From Nychka *et al*, (2002).

3 Analyzing space-time processes

3.1 Empirical orthogonal function analysis

In geophysics and meteorology, variants of principal components called empirical orthogonal functions (EOFs) have long been used to describe leading modes of variability in space-time processes. Here we use smoothed EOFs to model the spatio-temporal mean of a random field viewed as spatially varying systematic temporal trends. In contrast to the following section (3.2), which also concerns space-time modeling, this section does not require the data to derive from a regular grid for spectral modeling; on the other hand, it is not concerned with temporal correlation structure or, more generally, the modeling of nonseparable spatio-temporal correlation structure.

It is common in the space-time modeling literature to decompose observations into the sum of a systematic trend component and residuals

$$Z(x,t) = \mu(x,t) + \varepsilon(x,t). \tag{42}$$

We saw such a decomposition for the purely temporal case in section 2.4 where the trend was denoted T_t and modeled nonparametrically using wavelets. By contrast, the temporal trend for hourly averaged ozone fluxes in the application of section 3.2.5 is a simple trigonometric model, which was fitted and removed at the regular grid of locations for the spatio-temporal air quality model predictions. The trend in that application was believed to be constant in space, although that assumption is not generally necessary. The trigonometric model is one of the most common examples of the decomposition of temporal trend in terms of a series of orthogonal basis functions. Where details of the trend structure vary spatially, we write such a decomposition more generally, with spatially varying coefficients, as

$$\mu(x,t) = \beta_{x0} + \sum_{j=1}^{J} \beta_{xj} f_j(t), \tag{43}$$

the $\{f_j(t)\}$ being a set or orthogonal temporal basis functions. Our focus is on applications with smooth seasonal (or, as in sect 3.2.5, diurnal) trends. Many air quality parameters display a dominant seasonal trend structure that is not conveniently represented by sums of trigonometric basis functions. In these cases we seek a parsimonious set of nonparametric basis functions, $\{f_j(t)\}$ where the first basis function, $f_1(t)$ typically represents the dominant or average trend over the spatial region of interest and subsequent basis functions, computed to be orthogonal to the first, along with the spatially indexed parameters β_{xj}, permit the shape and amplitude of the spatial structure to vary. We illustrate these empirical trend characteristics and a simple approach to computation of useful basis functions using maximum 8-hour average daily ozone observations at 94 monitoring sites in southern California for the 8-year period 1987-1994.

Carrying out spatial analysis, including predictions of air quality concentrations $Z(x_0,t)$ at unmonitored locations x_0 requires spatial prediction of both the trend and the temporal deviations about the trend. The former can be achieved by more-or-less conventional geostatistical kriging of the multivariate spatial data set defined by the spatial dataset of vectors of trend coefficients. The latter is addressed here using the Sampson-Guttorp spatial deformation model for nonstationary spatial covariance structure, first introduced in Sampson and Guttorp (1992), but implemented here using the Bayesian framework of Damian et al. (2001, 2003).

3.1.1 Computation of temporal trend basis functions from incomplete data using an SVD

Assuming a complete data matrix of T observations on each of N monitoring sites, and letting $\mathbf{Z}$, $\mathbf{M}$, and $\mathbf{E}$ represent the $T \times N$ matrices corresponding to the terms of (42), the decomposition can be written

$$\mathbf{Z} = \mathbf{M} + \mathbf{E} \tag{44}$$

where

$$\mathbf{M} = \mathbf{F}\,\mathbf{B} \tag{45}$$

$\mathbf{F} = [f_0(t)\ f_1(t)\ \cdots\ f_J(t)]$ being a $T \times J$ matrix with columns the $T-$vectors of values of the basis functions $f_j(t),\ j = 1, 2, \ldots, J$, with $f_0(t) \propto \mathbf{1}$. Without loss of generality we can scale these basis functions to norm one, $f_j(t)' f_j(t) = 1$. The matrix $\mathbf{B}$ is the $J \times N$ matrix of trend coefficients for the N sites,

$$\mathbf{B} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0N} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1N} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_{J1} & \beta_{J2} & \cdots & \beta_{JN} \end{bmatrix} \tag{46}$$

As $\mathbf{M}$ is of rank J, we obtain the most parsimonious set of basis functions for a least squares approximation of the data matrix $\mathbf{Z}$ by taking $\mathbf{F}$ to be the matrix of the first J left singular vectors in the singular value decomposition (SVD) $\mathbf{Z} = \mathbf{U}\,\mathbf{D}\,\mathbf{V}'$. To the extent that a large number of sites share similar temporal (seasonal) patterns, the left singular vectors will represent these patterns. There are, however, two problems in proposing these as the empirical orthogonal function basis for this trend modeling problem. First: although the first singular vector does usually represent the dominant shared seasonal pattern, based on observational data with substantial variation about the trend, it turns out to be a noisy representation of this pattern. Second: air quality data matrices always contain substantial numbers of missing observations, so the usual SVD routines cannot be applied directly.

The SVD can be computed using an iterative algorithm in which the left singular vectors are computed by regressions in which smoothness is imposed directly. However, it has proven adequate to simply compute smoothed versions of the left singular vectors and then compute the coefficient matrix $\mathbf{B}$ by ordinary least squares regressions of the columns of $\mathbf{Z}$ on the set of J smoothed columns of $\mathbf{F}$. To deal with missing data, we use a simple, "EM-like" iterative algorithm for the SVD. This algorithm can be explained as follows:

1. Specify a dimension (rank), J, for the EOF model
2. Scale the observations at each monitoring site (columns of $\mathbf{Z}$) to norm (variance) one; call this matrix $\tilde{\mathbf{Z}}$
3. Fill in the missing observations in the data matrix $\tilde{\mathbf{Z}}$ using elements of an initial rank-one approximation provided by a regression through the origin of each column of $\tilde{\mathbf{Z}}$ on the vector u_1 computed as the average over sites of the columns of $\tilde{\mathbf{Z}}$.

4. Compute the rank-J SVD-approximation of the now complete data matrix $\tilde{\mathbf{Z}}$

5. Replace the missing values in $\tilde{\mathbf{Z}}$ by the elements of the rank-J SVD approximation.

6. Return to step 3 and iterate to convergence.

Following this calculation we use an ordinary spline smoother to smooth the plots of the left singular vectors against time $t = 1, 2, \ldots, T$. As the imputation of missing values in this algorithm depends explicitly on the assumed rank of the model, one must fit models of varying dimension and then choose a preferred model based on conventional measures like the percent of summed squared variation in $\tilde{\mathbf{Z}}$ explained, along with graphical analysis of the fitted trend models.

3.1.2 Spatial deformation modeling of the nonstationary spatial covariance structure of the detrended space-time residuals

Sampson and Guttorp (1994) introduced an approach to nonstationary spatial covariance modeling in which the geographic coordinates are deformed to create a geography (the disperion plane, or D-plane) in which the covariance structure is approximately isotropic. This approach is usually applied to detrended residuals. We assume for simplicity that the temporal structure of the residuals $\hat{\varepsilon}(x,t) = Z(x,t) - \hat{\mu}(x,t)$ is white noise.

We decompose the residuals $\hat{\varepsilon}(x,t) = \nu(x)\, H(x,t) + E(x,t)$, where $H(x,t)$ is a mean zero, variance one spatial process with covariance structure

$$\mathrm{Cov}(H(x,t), H(y,t)) = \rho_\theta\left(|f(x) - f(y)|\right),$$

and $E(x,t)$is a white noise process, uncorrelated with $H(x,t)$. The function $f$is the deformation of the geographic plane, and is fitted using a pair of thin-plate splines (Bookstein, 1989).

Technically we use a Gaussian-based Bayesian approach using MCMC, detailed in Damian et al. (2001,2003). This has the advantage that we can draw samples from the deformations, and get a good feeling for the uncertainty in the fit.

3.1.3 Example: Application to 8-hour maximum average daily ozone concentrations from southern California

The analysis of 8-hour maximum average daily ozone concentrations from southern California for the period 1987-94 was one of seven similar analyses of data from regions spanning most of the continental United States. Ozone seasonal trends are similar nationwide and we hoped to be able to use a single set of temporal trend basis functions for all regions. The computation of trend components was based on data from 513 monitoring sites monitoring nearly throughout the year across the country. A large fraction of the monitors in the U.S., especially in the north and northeast, are in operation only over the primary ozone season from April through October. Because of the lengthy time series available, and because we were not explicitly concerned with the spatially varying temporal autocorrelation in the ozone data, we avoided this issue by basing inference on the data obtained by subsampling every third daily observation, which results in series that have little autocorrelation (about their trend).

Temporal trend components. Analysis of the ozone concentration data was judged most appropriate on a square root scale. Figure 1 slows the four temporal basis functions computed from the first four left singular vectors of the 2912 $\times$ 513 data matrix (considering only those sites sampling essentially year-round). The first function clearly represents the dominant seasonal ozone cycle with highest concentrations during the sunny summer months. The shape and amplitude of this seasonal feature varies from year to year, another notable distinction with many seasonal trend models which do not adapt to fluctuations in trend. The second trend component is necessary for many sites which display a pair of ozone peaks, one in spring and another in later in the summer. The third and fourth components serve mainly to adjust the exact shape and locations of the seasonal peaks defined by the first two components. Approximately 67% of the variance in the entire (scaled) 2912 $\times$ 513 data matrix is explained by the first four unsmoothed components.

Figure 2 illustrates trends for two monitoring sites in Los Angeles county. These were computed as linear combinations of the four trend components illustrated in Figure 1. We note that there is distinct variation in the shape of the seasonal trend even over this relatively small spatial region with the first site, 060370002 (which was inoperative in early 1987 and from 1992 through 1994), showing the dominant seasonal pattern and the second site, 060371902, showing two reasonably distinct seasonal maxima. Figure 3 illustrates the prediction of the trends at these and two other sites in Los Angeles county, all of which were left out of the spatial analysis for purposes of cross-validation. Site 060371902 was selected as an example of the greatest error in the prediction of the trend. The error is primarily in the intercept rather than the shape of the seasonal trend.

Spatial-deformation and illustration of validation predictions of actual observations. The coastline, complex topology, and typical weather patterns combine to effect a complex nonstationary spatial correlation structure in the spatio-temporal residuals from the fitted temporal trends. Figure 4 depicts the posterior mean estimate of the spatial deformation computed to permit fitting of stationary isotropic correlation models in the deformed coordinate system. The predominant feature of compression along the coastline running NW-SE indicates that spatial correlation is strongest parallel to the coast and weaker orthogonal to the coast. Figure 5 shows empirical spatial correlations vs distance in the geographic and new deformed coordinate system. Spatial prediction (kriging) of the trend coefficients combined with prediction of the spatio-temporal residuals produce the cross-validation predictions of the time series for the two LA county sites, as illustrated in Figure 6.

3.2 Spectral analysis

3.2.1 A spectral representation

A stationary spatial-temporal process $\{Z(\mathbf{x},t) : \mathbf{x} \in D \subset \mathbb{R}^d, t \in T \subset \mathbb{R}\}$ has a spectral representation in terms of sine and cosine waves of different frequencies $(\boldsymbol{\omega}, \tau)$, where $\boldsymbol{\omega}$ is d-dimensional spatial frequency and τ is temporal frequency. If $Z(\mathbf{x},t)$ is a stationary random field with spatial-temporal covariance $C(\mathbf{x},t)$, then we can represent the process in the form of the following Fourier-Stieltjes integral:

$$Z(\mathbf{x},t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\boldsymbol{\omega}^T \mathbf{x} + i\tau t) dY(\boldsymbol{\omega}, \tau). \tag{47}$$

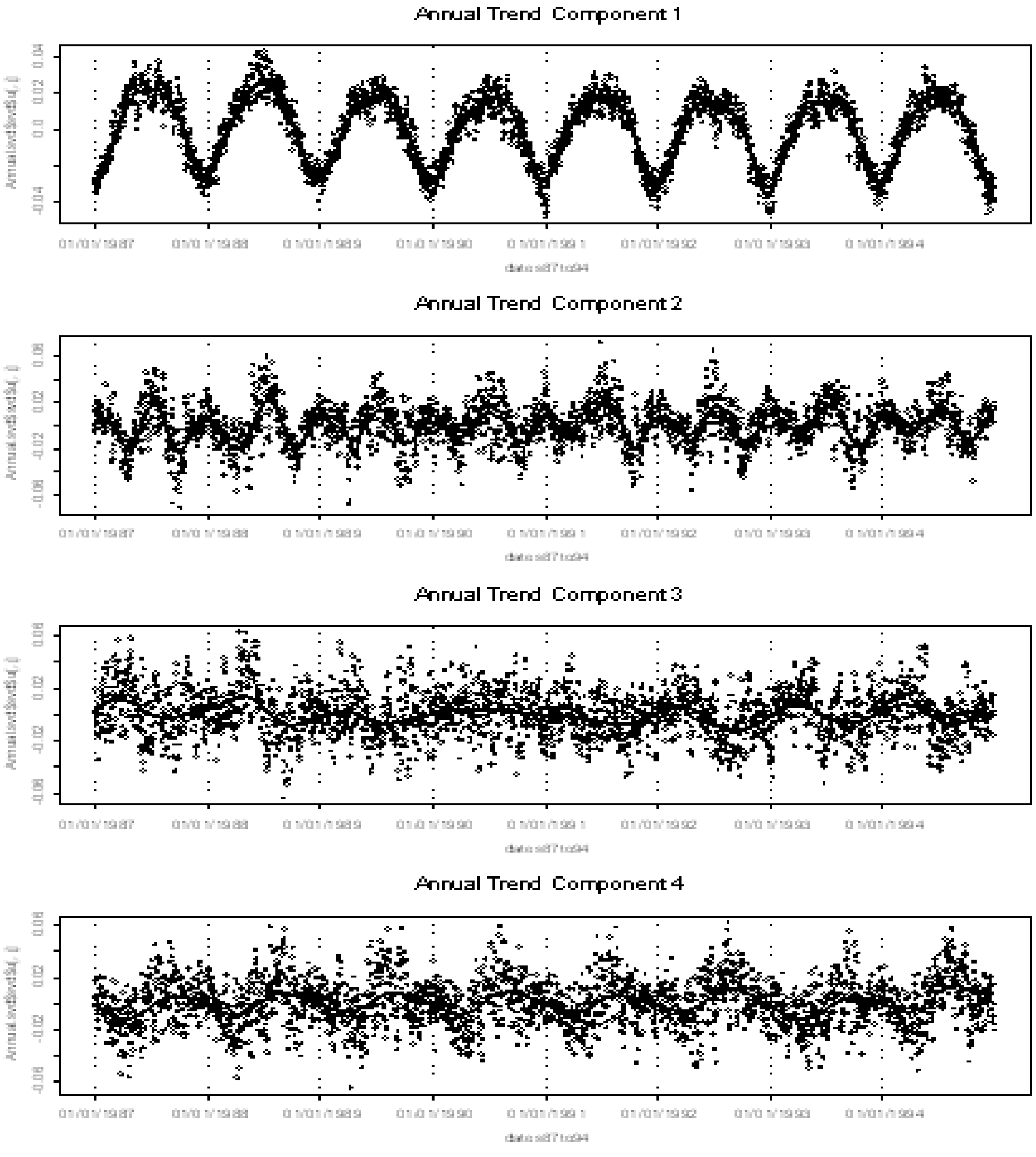


Figure 20: First four singular vectors (dots) and smooth trend components derived from the 2912 x 513 matrix of square root transformed ozone concentrations.

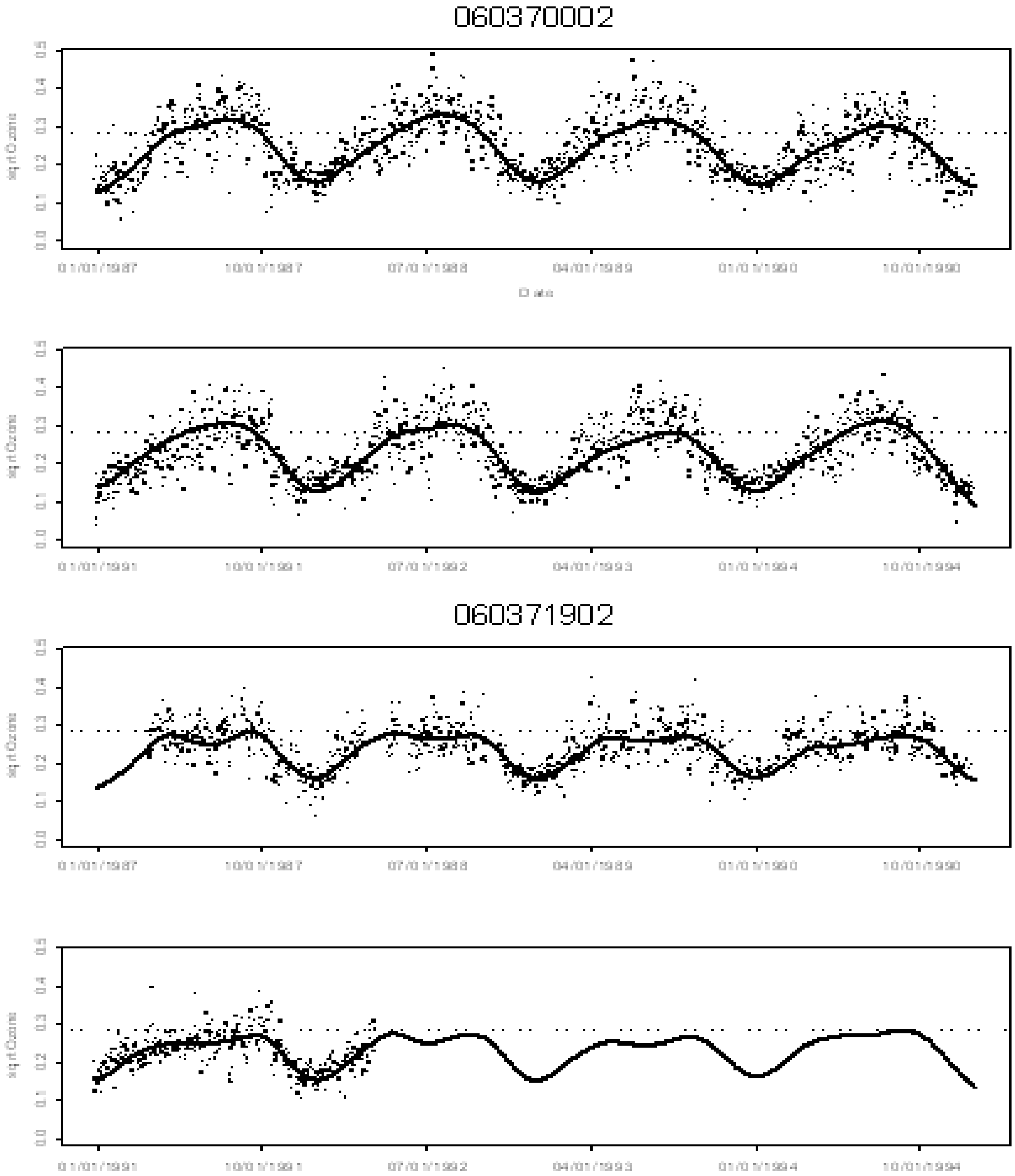


Figure 21: Fitted temporal trends for two monitoring sites in Los Angeles county.

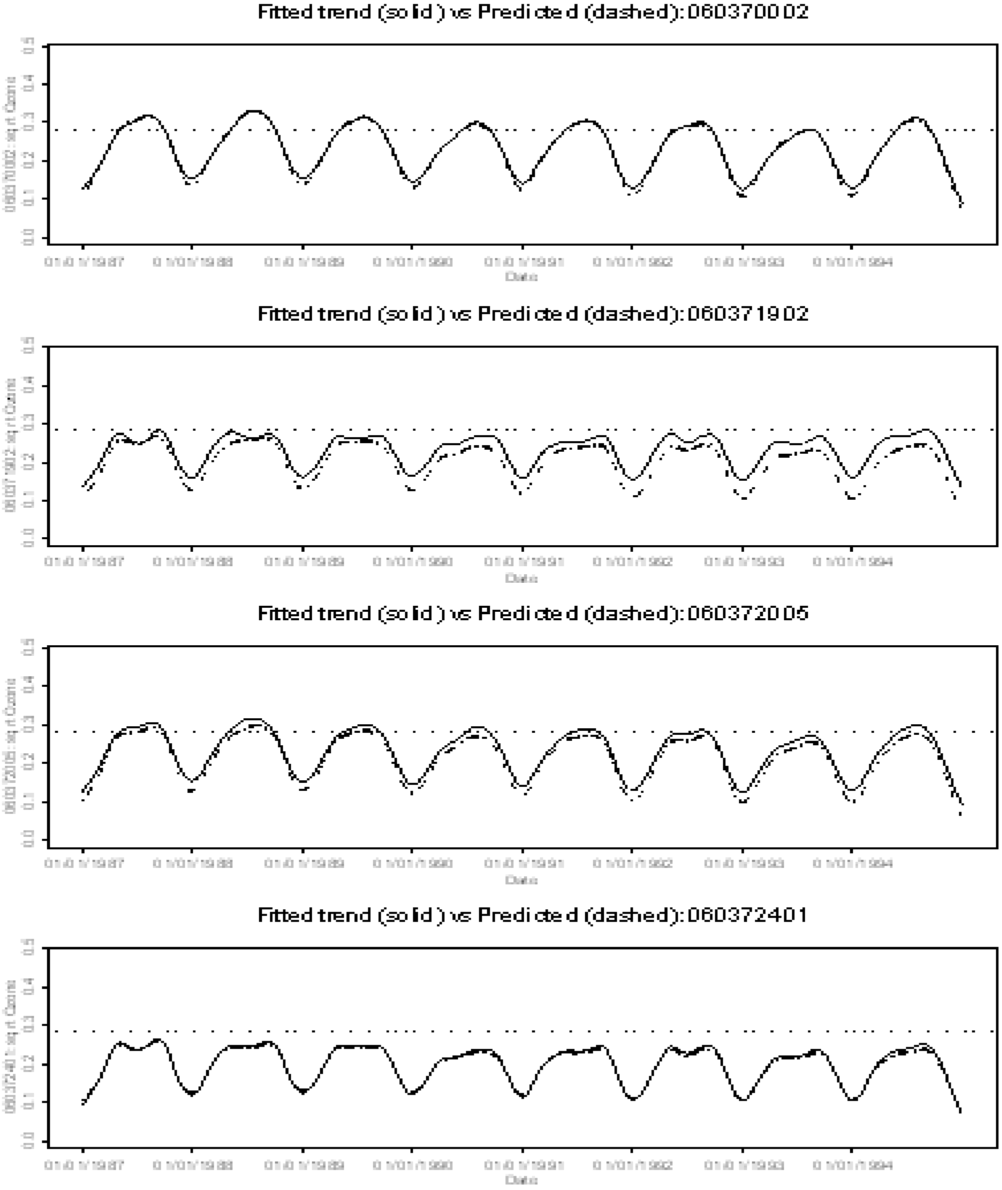
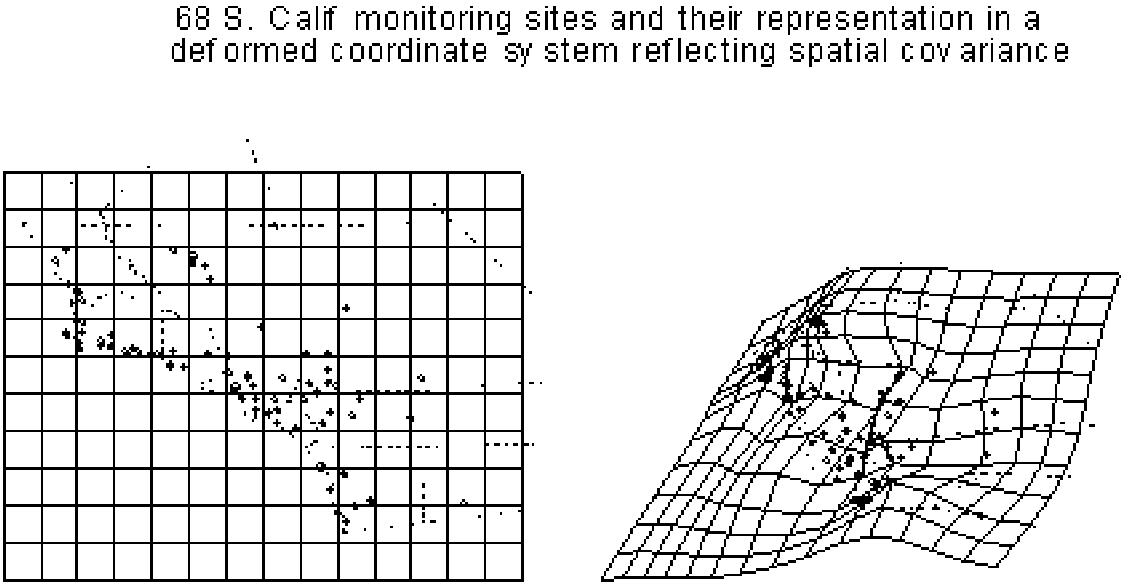


Figure 22: Predicted trend curves for four validation monitoring sites in Los Angeles county.



68 S. Calif monitoring sites and their representation in a deformed coordinate system reflecting spatial covariance

Figure 23: Posterior mean spatial deformation representing nonstationary spatial covariance structure.

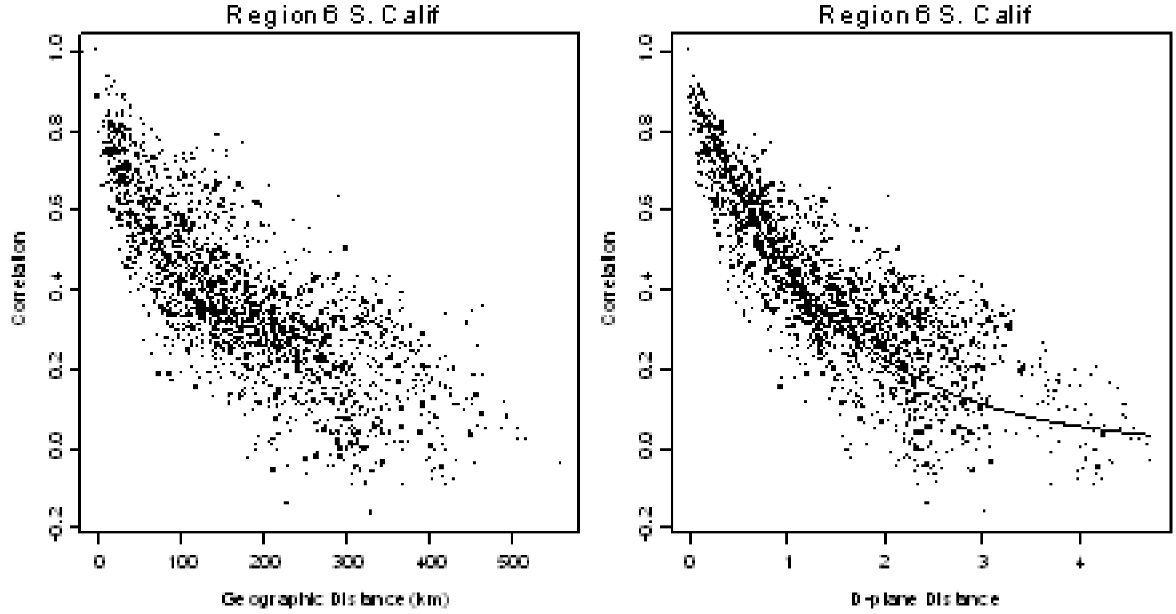


Figure 24: Spatial correlations vs (a) distance in the geographic plane and (right) distance in the deformed coordinate system (D-plane) of Figure 4. An estimated exponential spatial correlation function is drawn on the D-plane scatter.

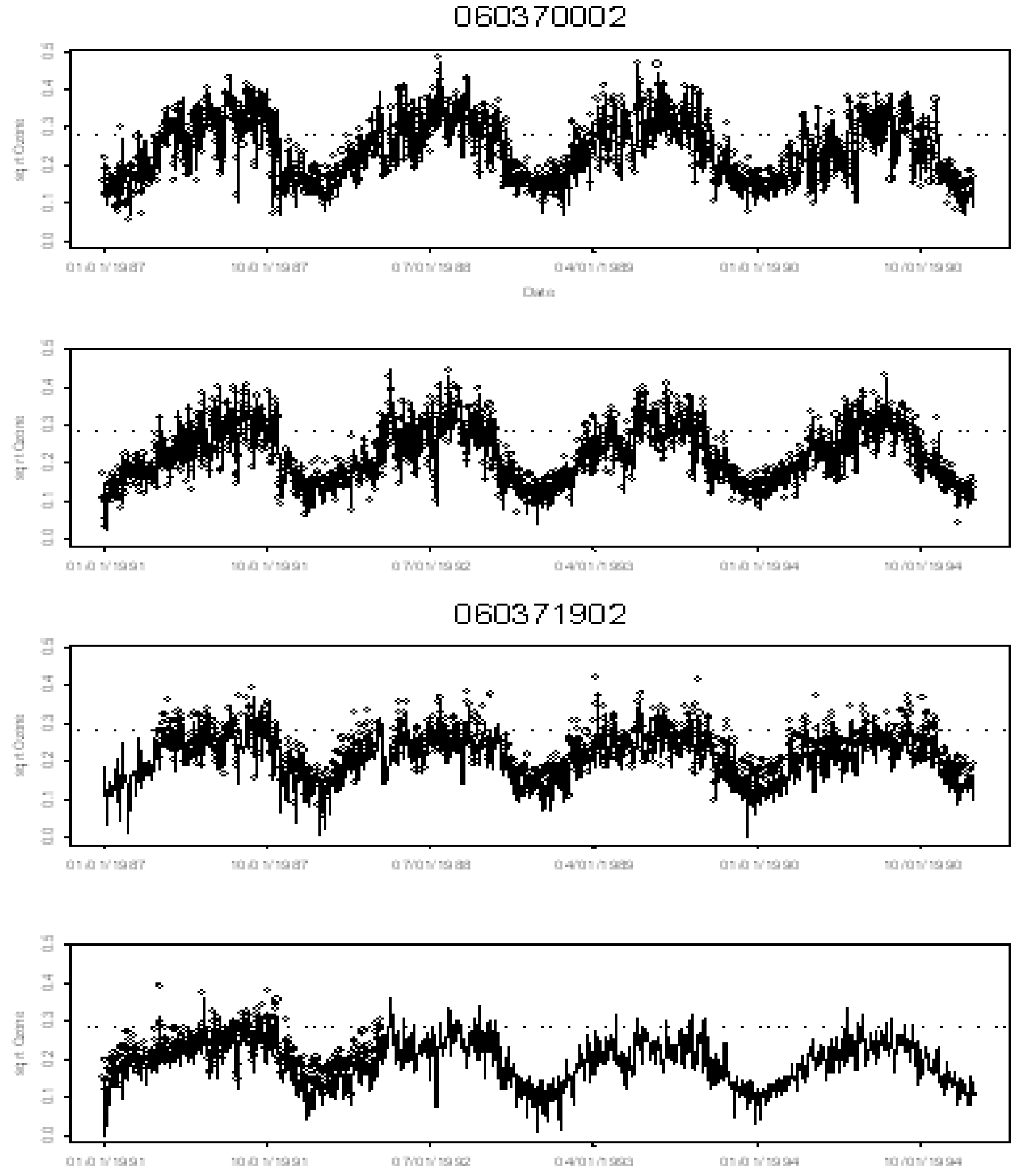


Figure 25: Cross-validation predictions of daily (square root) ozone concentrations at two monitoring sites in Los Angeles county.

where Y is a random function that has uncorrelated increments with complex symmetry except for the constraint, $dY(\boldsymbol{\omega},\tau) = dY^c(-\boldsymbol{\omega},-\tau)$, needed to ensure $Z(\mathbf{x},t)$ is real-valued. Y^c denotes the conjugate of Y. Using the spectral representation of Z and proceeding formally,

$$C(\mathbf{x},t) = \int_{\mathbb{R}^d}\int_{\mathbb{R}} \exp(i\boldsymbol{\omega}^T\mathbf{x} + i\tau t)F(d\boldsymbol{\omega}, d\tau), \tag{48}$$

where the function F is a positive finite measure and is called the spectral measure or spectrum for Z. The spectral measure F is the mean square value of the process Y,

$$E\{|Y(\boldsymbol{\omega},\tau)|^2\} = F(\boldsymbol{\omega},\tau).$$

It is easy to see that for any finite positive measure F, the function given in (48) is positive definite. If F has a density with respect to Lebesgue measure, it is the spectral density, f, which is the Fourier transform of the spatial-temporal covariance function:

$$f(\boldsymbol{\omega},\tau) = \frac{1}{(2\pi)^{d+1}}\int_{\mathbb{R}^d}\int_{\mathbb{R}} \exp(-i\boldsymbol{\omega}^T\mathbf{x} - i\tau t)C(\mathbf{x},t)d\mathbf{x}dt, \tag{49}$$

and the corresponding covariance function is given by

$$C(\mathbf{x},t) = \int_{\mathbb{R}^d}\int_{\mathbb{R}} \exp(i\boldsymbol{\omega}^T\mathbf{x} + i\tau t)f(\boldsymbol{\omega},\tau)d\boldsymbol{\omega}d\tau. \tag{50}$$

When $f(\boldsymbol{\omega},\tau) = f^{(1)}(\boldsymbol{\omega})f^{(2)}(\tau)$, we obtain

$$\begin{aligned} C(\mathbf{x},t) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}} \exp(i\boldsymbol{\omega}^T\mathbf{x} + i\tau t)f^{(1)}(\boldsymbol{\omega})f^{(2)}(\tau)d\boldsymbol{\omega}d\tau \\ &= \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}^T\mathbf{x})f^{(1)}(\boldsymbol{\omega})d\boldsymbol{\omega}\int_{\mathbb{R}} \exp(i\tau t)f^{(2)}(\tau)d\tau \\ &= C^{(1)}(\mathbf{x})C^{(2)}(t), \end{aligned}$$

which means the corresponding spatial-temporal covariance is separable.

3.2.2 A new class of nonseparable space-time covariances

We propose the following spatial-temporal spectral density that has a separable model as a particular case,

$$f(\boldsymbol{\omega},\tau) = \gamma(\alpha^2\beta^2 + \beta^2|\boldsymbol{\omega}|^2 + \alpha^2\tau^2 + \epsilon|\boldsymbol{\omega}|^2\tau^2)^{-\nu}, \tag{51}$$

where γ, α and β are positive, $\nu > \frac{d+1}{2}$ and $\epsilon \in [0,1]$. The function in (51) is a valid spectral density. First, $f(\boldsymbol{\omega},\tau) > 0$ everywhere. Second, $f(\boldsymbol{\omega},\tau) \leq \gamma(\alpha^2\beta^2 + \beta^2|\boldsymbol{\omega}|^2 + \alpha^2\tau^2)^{-\nu}$, and

$$\begin{aligned} &\int_{\mathbb{R}^d}\int_{\mathbb{R}} \exp(i\boldsymbol{\omega}^T\mathbf{x} + i\tau t)\gamma(\alpha^2\beta^2 + \beta^2|\boldsymbol{\omega}|^2 + \alpha^2\tau^2)^{-\nu}d\boldsymbol{\omega}d\tau \\ = &\frac{\pi^{\frac{d+1}{2}}\gamma}{2^{\nu-\frac{d+1}{2}-1}\Gamma(\nu)\alpha^{2\nu-d}\beta^{2\nu-1}}\left(\alpha\sqrt{(\frac{\beta}{\alpha}t)^2 + |\mathbf{x}|^2}\right)^{\nu-\frac{d+1}{2}} \times \\ &\mathcal{K}_{\nu-\frac{d+1}{2}}\left(\alpha\sqrt{(\frac{\beta}{\alpha}t)^2 + |\mathbf{x}|^2}\right), \end{aligned} \tag{52}$$

Therefore, $\int_{\mathbb{R}^d}\int_{\mathbb{R}} \exp(i\boldsymbol{\omega}^T\mathbf{x} + i\tau t)f(\boldsymbol{\omega},\tau)d\boldsymbol{\omega}d\tau$ exists.

In the representation (51), the parameter α^{-1} explains the rate of decay of the spatial correlation, β^{-1} explains the rate of decay for the temporal correlation, and γ is a scale parameter. The parameter ν measures the degree of smoothness of the process Z. The parameter ϵ indicates the interaction between the spatial and temporal components. Next, we discuss two particular case of ϵ.

Two particular cases:

$\underline{\epsilon = 1}$

When $\epsilon = 1$, the equation (51) can be written as

$$\begin{aligned} f(\boldsymbol{\omega},\tau) &= \gamma(\alpha^2\beta^2 + \beta^2|\boldsymbol{\omega}|^2 + \alpha^2\tau^2 + |\boldsymbol{\omega}|^2\tau^2)^{-\nu} \\ &= \gamma(\alpha^2 + |\boldsymbol{\omega}|^2)^{-\nu}(\beta^2 + \tau^2)^{-\nu}. \end{aligned}$$

Therefore the corresponding spatial-temporal covariance is separable. Moreover, in the expression of this covariance, both the spatial component and the temporal component are the Matérn type covariances.When $\gamma = \alpha = \beta = d = 1$ and $\nu = 3/2$, a contour plot of corresponding separable spatial-temporal covariance is given in Figure 8. From the plot, there are ridges along the lines where spatial lag is 0 and temporal lag is 0.

$\underline{\epsilon = 0}$

When $\epsilon = 0$,

$$f(\boldsymbol{\omega},\tau) = \gamma(\alpha^2\beta^2 + \beta^2|\boldsymbol{\omega}|^2 + \alpha^2\tau^2)^{-\nu}. \tag{53}$$

The function in (53) is an extension of the traditional Matérn spectral density. It treats time as an additional component of space, but it does have a different rate of decay. In the spectral density (53), the parameter α^{-1} explains the rate of decay of the spatial correlation. For the temporal correlation, the rate of decay is explained by the parameter β^{-1}. γ is a scale parameter. The parameter ν measures the degree of smoothness of the process Z. The higher value of ν, the smoother the process Z will be. The corresponding spatial-temporal covariance is give by (52), which is a Matérn type covariance, we have

$$C(\mathbf{x},t) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}\left(\frac{\|(\mathbf{x},\rho t)\|}{r}\right)\mathcal{K}_\nu\left(\frac{\|(\mathbf{x},\rho t)\|}{r}\right), \tag{54}$$

where σ^2, r, ρ and ν are all positive. $\|\cdot\|$ denotes the Euclidean distance.In the representation (54), the parameter r measures how the correlation decays with distance; generally this parameter is called range. The parameter σ^2 is the variance of the process Z. The parameter $\nu > 0$ measures the degree of smoothness of the process Z. The parameter ρ is a scale factor to take into account the change of units between the spatial and temporal domains. Therefore, this parametric model for C corresponds to a $d+1$ dimensional Matérn type covariance with an extra parameter ρ, which can be explained as a conversion factor between the units in the space and time domains. When $\gamma = \alpha = \beta = d = 1$ and $\nu = 3/2$, a contour plot of corresponding separable spatial-temporal covariance is given in Figure 27. It has a very smooth surface.

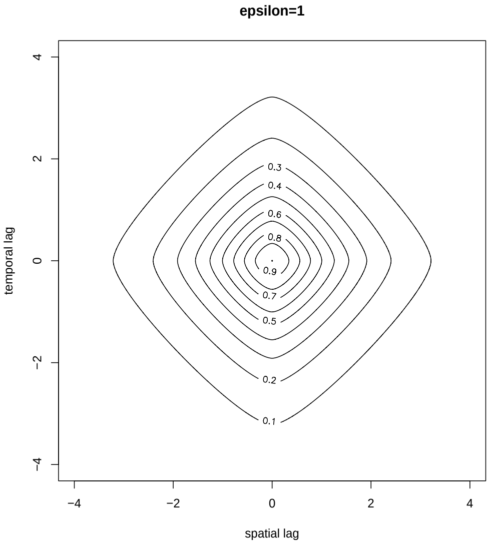


Figure 26: The contour plot for a separable spatial-temporal covariance.

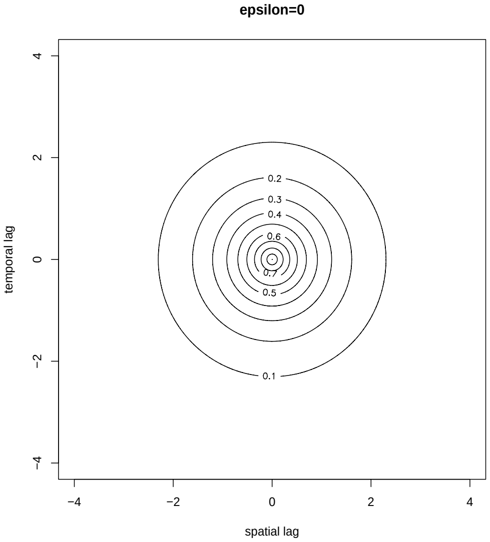


Figure 27: The contour plot for a nonseparable spatial-temporal covariance.

When $\epsilon \in (0,1)$ (Figure 10), we are not able to write down the exact expression of spatial-temporal covariance, which correponds to the spectral density in (51). But we can calculate this numerically, since

$$\begin{aligned} C(\mathbf{x}, t) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp(i\boldsymbol{\omega}^T \mathbf{x} + i\tau t) f(\boldsymbol{\omega}, \tau) d\boldsymbol{\omega} d\tau \\ &= \int_{\mathbb{R}} \exp(i\tau t) g(\mathbf{x}, \tau) d\tau, \end{aligned} \tag{55}$$

where $g(\mathbf{x}, \tau) = \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}^T \mathbf{x}) f(\boldsymbol{\omega}, \tau) d\boldsymbol{\omega}$. The function $g(\mathbf{x}, \tau)$ is available from the integration, therefore $C(\mathbf{x}, t)$ can be computed by numerically carrying out a one-dimensional Fourier transformation of g. This can be quickly approximated using fast Fourier transform. A separate transform needs to be done for every value of $(\mathbf{x}, t)$ of interest, but this is feasible. The expression of g is given by

$$g(\mathbf{x}, \tau) = \frac{\pi^{d/2}\gamma}{2^{\nu - \frac{d}{2} - 1}\Gamma(\nu)} (\beta^2 + \epsilon\tau^2)^{-\nu} \left(\frac{|\mathbf{x}|}{\theta(\tau)} \right)^{\nu - \frac{d}{2}} \mathcal{K}_{\nu - \frac{d}{2}}(\theta(\tau)|\mathbf{x}|),$$

where $\theta(\tau) = \sqrt{\frac{\alpha^2(\beta^2+\tau^2)}{\beta^2+\epsilon\tau^2}}$. When $\gamma = \alpha = \beta = d = 1$ and $\nu = 3/2$, contour plots of corresponding separable spatial-temporal covariances with ϵ =0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 0.9 are given in Figure 28. The ridge is getting more obviously when ϵ is getting large, say close to 1. In fact, when ϵ is close to 1, $\epsilon|\boldsymbol{\omega}|^2\tau^2$ is close to $|\boldsymbol{\omega}|^2\tau^2$. Therefore the spectral density in (51) is turning close to a separable form from a nonseparable form as $\epsilon \to 1$.

In summary, the new class spectral density in (51) is nonseparable for $0 \le \epsilon < 1$, and separable for $\epsilon = 1$. Therefore, the parameter ϵ plays a role for separability. It controls the interaction between the spatial component and the temporal component. Note that the degree of smoothness is same for the spatial component and the temporal component. Therefore a more general class is proposed to allow different degree of smoothness for space and time,

$$f(\boldsymbol{\omega}, \tau) = \gamma \left\{ c_1(a_1^2 + |\boldsymbol{\omega}|^2)^{\alpha_1} + c_2(a_2^2 + \tau^2)^{\alpha_2} + \epsilon(a_3^2 + |\boldsymbol{\omega}|^2\tau^2)^{\alpha_3} \right\}^{-\nu}, \tag{56}$$

where a_1, a_2, a_3, α_1, α_2, α_3 and c_1, c_2 are positive; $\epsilon \in [0,1]$ and $\frac{d}{\alpha_1\nu} + \frac{1}{\alpha_2\nu} < 2$. This is a valid spectral density. The spectral density in (51) is a special case of this more general class. If $\alpha_1 = \alpha_3 = 1$ and $\frac{d}{\nu} + \frac{1}{\alpha_2\nu} < 2$ in (56), we have

$$\begin{aligned} &\int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}^T \mathbf{x}) f(\boldsymbol{\omega}, \tau) d\boldsymbol{\omega} \\ &\quad = \frac{\pi^{d/2}\gamma}{2^{\nu - \frac{d}{2} - 1}\Gamma(\nu)} (c_1 + c_3\tau^2)^{-\nu} \left(\frac{|\mathbf{x}|}{\rho(\tau)} \right)^{\nu - \frac{d}{2}} \mathcal{K}_{\nu - \frac{d}{2}}(\rho(\tau)|\mathbf{x}|), \end{aligned} \tag{57}$$

where $\rho(\tau) = \left(\frac{c_2(a_2^2+\tau^2)^{\alpha_2} + c_1a_1^2 + c_3a_3^2}{c_1+c_3\tau^2} \right)^{1/2}$. So the corresponding spatial-temporal covariance $C(\mathbf{x}, t)$ can be quickly approximated by using the fast Fourier transformation of (57).

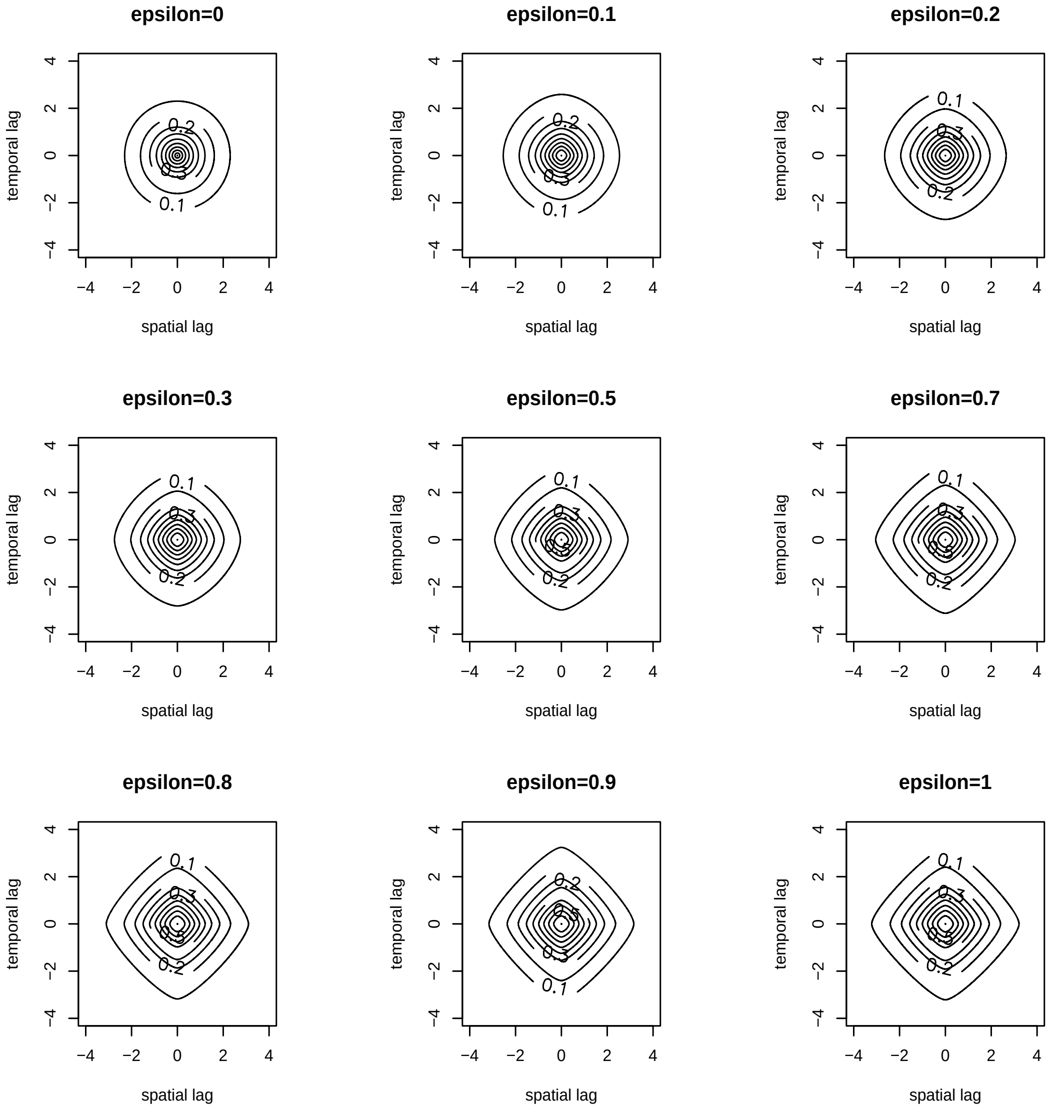


Figure 28: Contour plots for some nonseparable spatial-temporal covariances.

3.2.3 An example in meteorology

In the analysis presented here we study and model the spatial temporal structure of wind fiels using MM5 model output fields from July 21, 2002 for the full 24-h period. The complicated flow patterns over the region during this time are evident in Figure 29. The arrows indicate the direction from which the winds are coming, while the length of the stem indicates wind speed. An easterly wind (winds from the east, southeast and northeast) tends to dominate over the majority of the region for this time period. The 3AM (7-h forecast) plot clearly shows an area of confluence (an area where the wind vectors tend to come together) over the Bay. At 9AM (13-h forecast) this area of confluence has been replaced by an area of diffluence (an area where the wind vectors tend to spread apart). Diffluence seems to persist over the majority of the Bay at noon (16-h forecast) and for the rest of the period studied. There is little evidence in these plots to suggest that MM5 was capturing the sea breeze circulation, which observations show to be present.

As a first empirical attempt to deal with the nonstationarity inherent in these kinds of environmental data, we divided the spatial domain into two broad categories: land and water. Five subregions of nonstationarity were found and they are shown in Figure 30, these subregions were identified using the test for nonstationary. This final regional arrangement of clusters appears reasonable considering atmospheric and oceanic processes that are occurring in the boundary layer on this day.

We model the nonstationary Z as a mixture of (independent) local stationary space-time processes Z_i for $i = 1, \ldots, k$ that explain the space-time dependence structure in the 5 subregions of stationarity $S_1, \ldots, S_5$,

$$Z(\mathbf{x}, t) = \sum_{i=1}^{5} K(\mathbf{s} - \mathbf{s}_i) Z_i(\mathbf{x}, t), \tag{58}$$

where $\mathbf{s}_i = (\mathbf{x}_i, t_i)$ is the centroid of the $\mathrm{i}^t h$ subregion, and for $i = 1, \ldots, 5$. Z_i explains the spatial-temporal structure of Z in a subregion of stationarity S_i. The corresponding covariance function for the process Z is

$$\mathrm{cov}\{Z(\mathbf{x}_1, t_1), Z(\mathbf{x}_2, t_2)\} = \sum_{i=1}^{5} K(\mathbf{s}_1 - \mathbf{s}_i) K(\mathbf{s}_2 - \mathbf{s}_i) C_i(\mathbf{x}_1 - \mathbf{x}_2, t_1 - t_2) \tag{59}$$

where $\mathbf{s}_1 = (\mathbf{x}_1, t_1)$, $\mathbf{s}_2 = (\mathbf{x}_2, t_2)$, and each C_i is a space-time covariance that explains the space-time dependency in a subregion of stationarity S_i.

The weight function in (59) is modeled as $K(\mathbf{x} - \mathbf{x}_i) = \frac{1}{h_i^2} K_0(\frac{\mathbf{x} - \mathbf{x}_i}{h_i})$, where the location $\mathbf{x}_i$ is the centroid of the i-th subregion. The bandwith h_i is defined as half of the maximum distance for the i-th subregion. The function K_0 is modeled as $K_0(\mathbf{u}) = \frac{3}{4}(1 - u_1^2)_+ + \frac{3}{4}(1 - u_2^2)_+$, which is a quadratic weight function for $\mathbf{u} = (u_1, u_2)$. We fit the parameters for the spatial-temporal covariance of Z using a Bayesian framework. The priors for the sill parameter, spatial range parameter and temporal range parameter are Inverse Gamma with infinite variance. The prior distribution for the smoothness parameter is a uniform distribution with support $(0, 2]$. The support for the smoothness parameter is a conservative interval based on our previous experience analyzing similar datasets. The prior for ϵ gives all the mass to the values 0 and 1: $p(\epsilon = 0) = .5$, $p(\epsilon = 1) = 1$. The posterior distributions for ϵ_i (Table 4) suggest the separability for each subregion except for subregion 4.

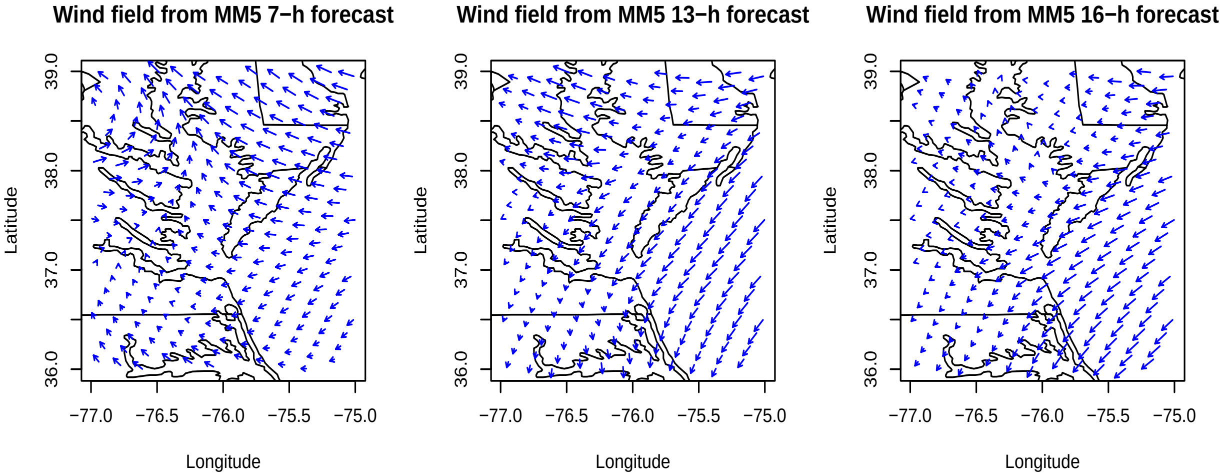


Figure 29: *Wind field maps, showing wind direction and speed over the Chesapeake Bay at 3am, 9am and noon on July 21, 2002.*

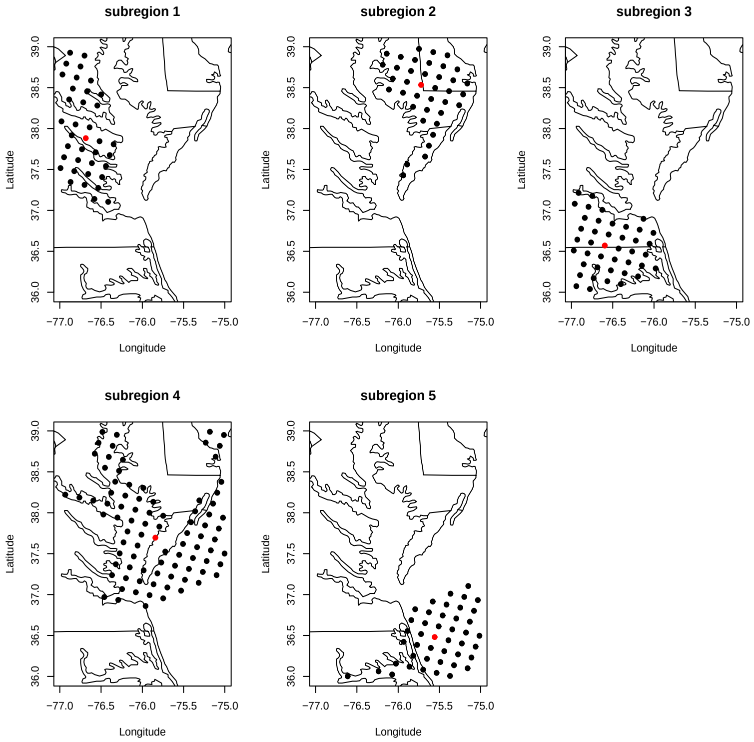


Figure 30: *Subregions of stationarity.*

	$P(\epsilon_i = 0)$	$P(\epsilon_i = 1)$
Subregion 1	7.62e-48	1
Subregion 2	1.03e-91	1
Subregion 3	3.76e-94	1
Subregion 4	1	0
Subregion 5	2.56e-150	1

Table 1: Table 4. The posterior distribution for ϵ_i.

These seems to indicate that in subregion 4 we have a nonseparable covariance and in the rest of the subregions we have separability.

3.2.4 Testing for separability

In this section we introduce a spectral analog of the assumption of separability.

Consider $\{Z(\mathbf{s};t) : \mathbf{s} \in D \subset \mathbb{R}^d; t \in \mathbb{Z}\}$ a spatial temporal zero-mean process observed at N space-time coordinates $(\mathbf{s}_1;t_1),\ldots,(\mathbf{s}_N;t_N)$. We start by assuming that the covariance function is stationary in space and time,

$$\mathrm{cov}(Z(\mathbf{s}_1;t_1), Z(\mathbf{s}_2;t_2)) = C(\mathbf{h};u)$$

for $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$, $u = t_1 - t_2$.

We can write the covariance C in terms of the spectral density g of the spatial temporal process Z (Section 1),

$$C(\mathbf{h},u) = \int\int \exp\{i\mathbf{h}^T\boldsymbol{\omega} + iu\tau\} g(\boldsymbol{\omega};\tau) d\boldsymbol{\omega} d\tau,$$

where

$$\begin{aligned} g(\boldsymbol{\omega},\tau) &= (2\pi)^{-d-1} \sum_{u=-\infty}^{u=\infty} \int \exp\{-i\mathbf{h}^T\boldsymbol{\omega} - iu\tau\} C(\mathbf{h};u) d\mathbf{h} \\ &= (2\pi)^{-d} \int \exp\{-i\mathbf{h}^T\boldsymbol{\omega}\} f(\mathbf{h};\tau) \mathbf{dh} \end{aligned} \tag{60}$$

for any fixed $\mathbf{h}$, $f(\mathbf{h};\tau)$ is the cross-spectral density function of the time processes $Y_1(t) = Z(\mathbf{s};t)$ and $Y_2(t) = Z(\mathbf{s}+\mathbf{h};t)$, and we have,

$$f(\mathbf{h};\tau) = (2\pi)^{-1} \sum_{u=-\infty}^{u=\infty} \exp\{-iu\tau\} C(\mathbf{h},u). \tag{61}$$

If C is a separable covariance then we can write

$$C(\mathbf{h},u) = C_1(\mathbf{h})C_2(u)$$

where C_1 is a positive-definite function in $\mathbb{R}^d$, and C_2 is a positive-definite function in $\mathbb{R}$. Thus, $f(\mathbf{h};\tau)$ is the product of a function of $\mathbf{h}$ and a function of τ,

$$f(\mathbf{h};\tau) = (2\pi)^{-1} \sum_{u=-\infty}^{u=\infty} \exp\{-iu\tau\} C(\mathbf{h},u) = (2\pi)^{-1} \sum_{u=-\infty}^{u=\infty} \exp\{-iu\tau\} C_1(\mathbf{h})C_2(u) = C_1(\mathbf{h})\kappa(\tau) \tag{62}$$

where κ is an integrable and positive function, and C_1 for each fixed τ is a covariance function of $\mathbf{h}$ and an integrable function of $\mathbf{h}$. We can obtain a nonseparable covariance function by making C_1 depend on τ. Thus, we get

$$C(\mathbf{h},u) = \int \exp\{iu\tau\} C_1(\mathbf{h};\tau)\kappa(\tau) \mathbf{d}\tau.$$

Cressie and Huang (1999) use this spectral representation to generate parametric models of non-separable spatio-temporal stationary covariance functions.

Thus, if Z is separable,

$$f(\mathbf{h};\tau) = C_1(\mathbf{h})\kappa(\tau). \tag{63}$$

In principle f could be a complex function, but when Z is separable and stationary f is real.

Testing for sparability

We propose a test for separability of spatial-temporal processes. The beauty of this method is that the mechanics of the test can be reduced to those of a simple two-way ANOVA procedure. We test for separability by studying if the coherence R is a function of τ.

We consider the standardized asymptotic distribution of the variance stabilizing transformation of R, i.e.

$$\phi_{\mathbf{a},\mathbf{b}}(\tau) = \tanh^{-1}(\tilde{R}_{\mathbf{a},\mathbf{b}}(\tau)),$$

and we estimate it with

$$\hat{\phi}_{\mathbf{a},\mathbf{b}}(\tau) = \tanh^{-1}(\hat{R}_{\mathbf{a},\mathbf{b}}(\tau)).$$

where the coherency is estimated by replacing f with $\hat{f}_{\mathbf{ab}}(\omega)$ a tapered second-order periodogram function proposed by Fuentes (2005b),

$$\hat{f}_{\mathbf{ab}}(\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_\rho(\mathbf{a}-\mathbf{s})g_\rho(\mathbf{b}-\mathbf{s})I^*_{\mathbf{a}+\mathbf{s},\mathbf{b}+\mathbf{s}}(\omega)d\mathbf{s}, \tag{64}$$

where

$$I^*_{\mathbf{ab}}(\omega) = 2\pi/T\sum_{t=0}^{T-1} W^{(T)}(\omega - 2\pi t/T)I_{\mathbf{ab}}(2\pi t/T), \tag{65}$$

I is the second-order periodogram, and W and g are two filter functions. Thus, $\hat{f}_{\mathbf{ab}}(\omega)$ can be interpreted as an average of the total energy of the process contained within a band of frequencies in the region of ω and a region in space in the neighborhood of $\mathbf{a}$, and $\mathbf{b}$. We evaluate $\hat{\phi}_{(\mathbf{a}_i,\mathbf{b}_i)}(\tau)$, at k pairs at pairs of locations $\{(\mathbf{a}_i,\mathbf{b}_i)\}_{i=1}^k$ and a set of frequencies $\tau_1,\tau_2,\ldots,\tau_n$ that cover the domain. We write

$$\hat{\phi}_{(\mathbf{a}_i,\mathbf{b}_i)}(\tau_j) = \phi_{(\mathbf{a}_i,\mathbf{b}_i)}(\tau_j) + \epsilon((\mathbf{a}_i,\mathbf{b}_i),\tau_j).$$

Asymptotically $E\{\epsilon((\mathbf{a}_i,\mathbf{b}_i),\tau_j)\} = 0$ and $\mathrm{Var}\{\epsilon((\mathbf{a}_i,\mathbf{b}_i),\tau_j)\} = \sigma^2$ where σ^2 is independent of $(\mathbf{a}_i,\mathbf{b}_i)$, and ω_j.

Assuming the $(\mathbf{a}_i,\mathbf{b}_i)$ and τ_j are spaced "sufficiently wide apart," then the $\epsilon((\mathbf{a}_i,\mathbf{b}_i),\tau_j)$ will be approximately uncorrelated, this is based on the asymptotic properties of $\hat{f}_{(\mathbf{a}_i\mathbf{b}_i)}(\tau_j)$ (see Fuentes (2005b). We write

$$U_{ij} = \hat{\phi}_{(\mathbf{a}_i,\mathbf{b}_i)}(\tau_j), \tag{66}$$

$$m_{ij} = \phi_{(\mathbf{a}_i,\mathbf{b}_i)}(\tau_j),$$

and

$$\epsilon_{ij} = \epsilon((\mathbf{a}_i, \mathbf{b}_i), \tau_j).$$

Then, we have the model

$$U_{ij} = m_{ij} + \epsilon_{ij},$$

that becomes the usual "two-factor analysis of variance" model, and can be rewritten:

$$H_1 : U_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, n$. The parameters $\{\alpha_i\}$, $\{\beta_j\}$ represent the main effects of the space and frequency factors. Then, we test for separability by using the standard techniques to test the model ($\beta_j = 0$):

$$H_{0a} : U_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

against the more general model H_1.

3.2.5 An example in air quality

In this example we analyze the spatial temporal structure of a process Z, which is the hourly averaged ozone fluxes (kg/hectare) in June 1996 (June 2-June 4), using the output of air quality numerical models (Models-3) (see Figure 31). We have 72 observations over time. Before applying our tests for stationarity and separability we need to remove the spatial-temporal trend.

In order to remove the spatial trend, we calculated at each location the ozone anomaly, that is the corresponding ozone value minus the mean over time (using the 72 observations over time at each location). We removed the temporal trend using a cosine and sine function with a period of 24 hours. The diurnal cycle of the ozone anomalies appeared to be the same everywhere. We implemented our test to the ozone anomalies (after removing this diurnal cycle).

The estimates, $\hat{f}_{\mathbf{ab}}(\boldsymbol{\omega})$, were obtained using (64), the cross-spectral function proposed by Fuentes (2005b), for the pair $\{\mathbf{a}, \mathbf{b}\}$ in D, in which $W(\alpha)$ is given by (**??**) with a bandwidth of $2\pi B$ with $B = 1/12$, and $g_\rho(\mathbf{u})$ is of the form (**??**) with $\rho = 5.5$ units (1 unit = 36 km), $m = BT$. Thus, in order to obtain approximately uncorrelated estimates, the frequencies ω_j and pairs $\{(\mathbf{a}_i, \mathbf{b}_i)\}_{i=1}^k$ should be chosen so that the spacings between the ω_j are at least $\pi/6$ and the distance between any pairs $(\mathbf{a}_i, \mathbf{b}_i)$ and $(\mathbf{a}_j, \mathbf{b}_j)$ for $i \neq j$ is at least 5.5 grid cells (198km).

The ω_j were chosen as follows $\omega_j = \pi j/17$ with $j = 1\,(3)\,16$, corresponding to a uniform spacing of $3\pi/17$ (which just exceeds $\pi/6$). We evaluate $\hat{f}_{\mathbf{ab}}(\omega)$ at the following frequencies, $\omega_1 = \pi/17$, $\omega_2 = 4\pi/17$, $\omega_3 = 7\pi/17$, $\omega_4 = 10\pi/17$, $\omega_5 = 13\pi/17$, and $\omega_6 = 16\pi/17$. We consider six pairs,$\{(\mathbf{a}_i, \mathbf{b}_i)\}_{i=1}^6$, such that the distance between pairs is at least 216 km (which just exceeds ρ) for all i, where $\|.\|$ denotes the Euclidean distance. Figure 11 shows the locations of $\mathbf{a}_i$ for $i = 1, \ldots, 6$, which correspond to the sites $A, B, \ldots,$ and F, respectively. The locations of $\mathbf{b}_i$ for $i = 1, \ldots, 6$, are other six sites at the same latitude as $A, B, \ldots,$ and F but 72 km, 144 km, 180 km, 72 km, 144 km, and 180 km further east, respectively. Table 6 shows the the results of the test for separability, using these 6 pairs.

Item	Degrees of freedom	Sum of squares	F Value	Pr(F)
Between spatial points	5	10.4784	6.1440	0.0007
Between frequencies	5	25.5483	14.9802	0.0000
Residuals	25	8.5273		

Table 6. Analysis of variance (all 6 sites).

The "between spatial locations" effect is highly significant (p-value < 0.01), confirming that there is clear evidence of lack of separability. The coefficient of determination is .81.

In the next table we study separability in a smaller subregion (the eastern part of our domain) using only pairs $(\mathbf{a}_3, \mathbf{b}_3)$ and $(\mathbf{a}_6, \mathbf{b}_6)$. Table 7 shows that the "between spatial locations" effect is not significant, suggesting that then there is no evidence of lack of stationarity (in the eastern part of our domain using pairs $(\mathbf{a}_3, \mathbf{b}_3)$ and $(\mathbf{a}_6, \mathbf{b}_6)$). However, the "between frequencies" effect is highly significant, confirming that the process is nonseparable. The coefficient of determination is .97. Even in a smaller subregion the assumption of separability is still unrealistic (Table 2). The distance between the two components in both pairs is the same, so we could use these two pairs to test for stationarity. It appears that stationarity in that smaller subregion is a reasonable assumption.

Item	Degrees of freedom	Sum of squares	F Value	Pr(F)
Between spatial points	1	0.0129	0.1872	0.6832
Between frequencies	5	11.1377	32.1995	0.0008
Residuals	5	0.3459		

Table 7. Analysis of variance (using sites **c** and **f**).

Table 8 shows the $|\hat{R}_{\mathbf{ab}}(\omega)|^2$ values at each one of the frequencies ω_j for $j = 1, \ldots, 6$ for the pairs $(\mathbf{a}_3, \mathbf{b}_3)$ and $(\mathbf{a}_6, \mathbf{b}_6)$. $|R|^2$, can be interpreted as a coefficient of correlation: values close to 1 indicate high correlation between the two time series. Both pairs seem to have a very similar coherency function, which supports the assumption of stationarity. However, the coherence is clearly changing with frequency. Thus, it seems that locally it might reasonable to assume stationarity for ozone fields, but separability seems to be unrealistic even for small geographic areas.

Pairs:	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$(\mathbf{a}_3, \mathbf{b}_3)$	0.769	0.648	0.891	0.960	0.994	0.999
$(\mathbf{a}_6, \mathbf{b}_6)$	0.931	0.660	0.840	0.978	0.994	0.998

Table 8. Coherence for pairs $(\mathbf{a}_3, \mathbf{b}_3)$ and $(\mathbf{a}_6, \mathbf{b}_6)$ at six frequencies.

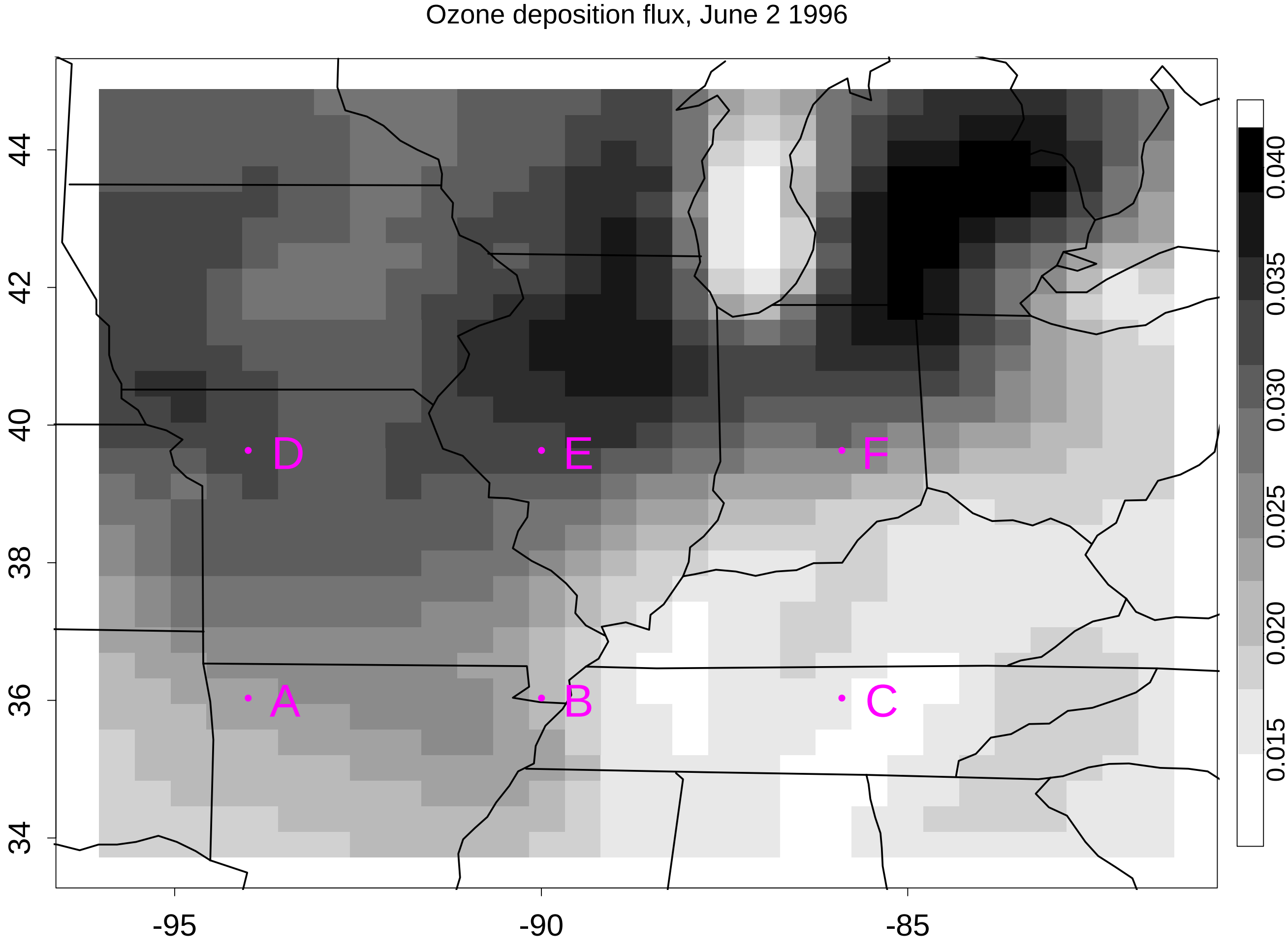


Figure 31: This graph shows the ozone deposition flux (kg/hectare) in a region in the Midwest. The values in this graph are the output of the EPA regional scale air quality model (Models-3) on June 2 at 12pm central time.

References

Abramowitz, M., and Stegun, I. A. (1965). *Handbook of Mathematical Functions,* Dover Publications, New York.

Akaike, H. (1974). A new look at statistical model identification. *IEEE Trans. Auto. Contr.* **AC -19**, 716-22.

Beran, J. (1994). *Statistics for Long Memory Processes.* Monographs on Statistics and Applied Probability, **61**. Chapman & Hall. New York.

Bookstein, F.L. (1989). Principal warps – Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6),. 567-585.

Bracewell, R. (1999) *The Fourier Transform and Its Applications*, 3rd ed. New York: McGraw-Hill.

Clark, R. M. (1977). Non-parametric estimation of smooth regresion function. *Journal of the Royal Statistical Society,* **39**, 107-113.

Craigmile, P.F. and Guttorp, P. and Percival, D.B. (2004). Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics*, **15**, 313-355.

Cole, J. and Dunbar, R. and McClanahan, T. and Muthiga, N. (2000). Tropical pacific forcing of decadal variability in the western indian ocean over the past two centuries. *Science,* **287**, 617-619.

Cramér, H. and Leadbetter, M. R. (1967). *Stationary and related stochastic processes. Sample function properties and their applications.* Wiley, New York.

Dahlhaus, R. and Küsch, H. (1987), Edge effects and efficient parameter estimation for stationary random fields. *Biometrika,* **74** 877-882.

Damian, D., Sampson, P.D., and Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures, *Environmetrics*, **12**, 161-178.

Damian, D., Sampson, P.D., and Guttorp, P. (2003). Variance modeling for nonstationary spatial processes with temporal replications. *Journal of Geophysical Research-Atmospheres*, 108(D24), Art. No. 8778.

Daubechies, I. (1992). Ten Lectures on Wavelets. *SIAM, CBMS-NSF Conference Series in Applied Mathematics*, **61**.

Fuentes, M. (2001). A new high frequency kriging approach for nonstationary environmental processes. *Environmetrics,* **12**, 469-483.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, **89**, 197-210.

Fuentes, M. (2005a). A formal test for nonstationarity of spatial stochastic processes. *Journal of Multivariate Analysis,* to appear.

Fuentes, M. (2005b). Testing for separability of spatial-temporal covariance functions. *Journal of Statistical Planning and Inference,* in press.

Fuentes, M. and Smith, R. (2001). A new class of nonstationary models. Tech. report at North Carolina State University, Institute of Statistics Mimeo Series #2534.

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika,* **69**, 95–105.

Guyon, X. (1992). *Champs aléatoires sur un réseau,* Paris: Masson.

Handcock, M. S., and Wallis, J. R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields (with discussion). *J. Amer. Statist. Assoc.*, **89**, 724-731.

Haas, T.C. (1998), Statistical assessment of spatio-temporal pollutant trends and meteorological transport models. *Atmospheric Environment* **32**, 1865-1879.

Jenkins, G. M. (1961). General considerations in the estimation of spectra. *Technometrics,* **3**, 133-166.

Kitanidis, P. K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resour. Res.* **19**, 909-21.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674-693.

Mardia, K. V. & Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-46.

Mardia, K.V., Kent, J.T. & Bibby, J.M. (1979). *Multivariate Analysis.* New York: Academic Press.

Matérn, B. (1960). *Spatial Variation.* Meddelanden från Statens Skogsforskningsinstitut, **49**, No. 5. Almaenna Foerlaget, Stockholm. Second edition(1986), Springer-Verlag, Berlin.

Nychka, D. and Wikle C. and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modeling,* **2**, 315-332.

Pawitan, Y. and O'Sullivan F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *Journal of the American Statistical Association,* **89**, 600-610.

Percival, D.B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis.* Cambridge University Press, Cambridge, UK.

R Development Core Team, (2004). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Ripley, B. D. (1981). *Spatial Statistics,* New York: John Wiley.

Renshaw, E. and Ford, E. D. (1983). The Interpretation of Process from Pattern using Two-dimensional Spectral Analysis: Methods and Problems of Interpretation. *Applied Statistics*, **32**, 51-63.

Renshaw and Ford (1983)

Rosenblatt, M. R. (1985). *Stationary Sequences and Random Fields,* Boston: Birkhäuser.

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

Stein, M. L. (1995). Fixed-domain asymptotics for spatial periodograms. *Journal of the American Statistical Association,* **90**, 1277-1288.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* New York: Springer-Verlag.

Stein, M. L., and Handcock, M. S. (1989). Some asymptotic properties of kriging when the covariance function is misspecified. *Mathematical Geology,* **21**, 171-190.

Stein, Chi and Welty. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society, Series B.* **66**, 275-296.

Whittle, P. (1954). On stationary processes in the plane, *Biometrika,***41**, 434-449.