Modeling and prediction of nonstationary spatial processes

Montserrat Fuentes[1]

SUMMARY

We present an approach for the spectral analysis of non-stationary spatial processes, that is based on the concept of *spatial spectra*, this means spectral functions which are space dependent. This notion of spatial spectra generalizes the definition of spectra for stationary processes, and under certain conditions, the spatial spectrum at each location can be estimated from a single realization of the spatial process. We propose a nonstationary nonparametric and various parametric approaches to estimate the spectral density of a nonstationary spatial process defined on a continuous space, and we develop new fitting algorithms. The methods are based on the spatial spectral approach. We also study the asymptotic properties of the proposed estimates via *shrinking asymptotics*, assuming the distance between neighboring observations tends to zero as the size of the observation region grows without bound.

We also consider the problem of testing a given spatial process for stationarity and isotropy. The approach is based again on the spatial spectral analysis, and the proposed method consists essentially in testing the homogeneity of a set of spatial spectra evaluated at different locations. In addition to testing stationarity and isotropy, the analysis provides also a method for testing whether the observed process fits a uniformly modulated model, and a test for randomness (constancy of spectra). Applications include modeling and testing for non-stationary of sulfur dioxide concentrations over different geo-political boundaries.

[1]Montserrat Fuentes is an assistant 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. This research was sponsored by a National Science Foundation grant DMS 0002790, and by a US EPA grant R-8287801.

Key words: Anisotropy, covariance, edge effect, Fourier transform, geostatistics, periodogram, spatial statistics, variogram.

1 Introduction

Spectral methods are powerful tools for studying the spatial structure of stationary processes. It is widely recognized that in real applications spatial processes are rarely stationary and isotropic, then an important extension of these spectral algorithms is to processes that are nonstationary. In this work, we present some new spectral approaches and tools to estimate the spatial structure of a nonstationary process. More specifically, we propose a nonstationary nonparametric and various parametric approaches to estimate the spectral density of a nonstationary spatial process defined on a continuous space. We also study the asymptotic properties of the proposed estimates via *shrinking asymptotics*, assuming the distance between neighboring observations tends to zero as the size of the observation region grows without bound. With this type of asymptotic model we can uniquely determine the spectral density, avoiding the aliasing problem.

Use and properties of spatial nonparametric spectral estimates 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. The spectral representation of a stationary process Z is interpreted as its representation in the form of superposition of sine and cosine waves of different frequencies. Spatial processes in environmental sciences, oceanography, soil science, and many other disciplines are generally nonstationary, in the sense that the spatial structure depends on location. This lack of stationarity thwarts the use of the standard spectral methods mentioned above. Dahlhaus (1996a) uses the notion of local stationarity to study asymptotic properties of nonstationary time series. The asymptotic model proposed by Dahlhaus (1996a, and 1996b) is for discrete processes observed at some fixed times, $t = 0, \ldots, T$. This asymptotic model assumes that by increasing the sample size ($T \to \infty$) more and more data of each local structure become available allowing for meaningful asymptotic investigations of statistical procedures. This means an increase of the bandwidth of stationarity with an increase of the sample size. Dahlhaus and Sahm (2001) use the same asymptotic framework for random fields. This asymptotic framework is appropriate for time series but unrealistic for most spatial processes. Generally, in a spatial analysis the subregions of stationarity correspond to some geographic areas and the bandwidth of stationarity should not be altered by increasing the sample size. The models by Dahlhaus and Sahm (2001), and Dahlhaus (1996b) are for discrete processes. However, in most spatial settings the underlying spatial process is continuous. The work presented here is for spatial continuous processes. This continuity assumption makes estimating the spectrum a more challenging problem due to the aliasing phenomenon, that occurs when we try to approximate the spectrum of a continuous process using observations on a lattice.

What is novel about this paper is the introduction of spectral methods to study the spatial structure of a nonstationary process spatial process, using a "spatial spectra"; that is, spectral functions which are space dependent. New fitting algorithms are developed using parametric spectral estimates within a Bayesian framework. We also introduce a new estimate of the spectral density for spatial processes and we study its asymptotic properties using a novel asymptotic model for spatial processes. In cases where the data are observed on a d-dimensional lattice, it is argued that these spectral approaches have computational benefits compared with maximum likelihood on the space domain.

All the work that has been done so far to test the non-stationarity of a random process is mainly in one dimension. Several authors have proposed methods for testing whether or not a given time-series may be regarded as stationary, some of these methods are designed to detect non-stationary "trends". Priestley and Rao (1969) used the concept of the "evolutionary (time-dependent) spectrum" to test the lack of stationarity of a time series, and this is the key idea for the work developed in this article. The method to be described in this paper may be used to test the *overall* stationarity of the complete second-order properties of a spatial processes in two or higher dimensions. A further advantage of our method is that it enables one to test not only the overall stationarity of the spatial process, but also to examine the character of the non-stationarity (when it exists) and the potential anisotropy. To implement our test for nonstationarity we need the approximations provided by the asymptotic results obtained in this article.

This article is organized as follows. In Section 2, we introduce a *spatial spectral* representation for a nonstationary spatial process, using spectral functions which are space dependent, and we define a new asymptotic model called *shrinking asymptotics*, which assumes the distance between neighboring observations tends to zero and the size of the observation region grows without bound. In Section 3, we propose parametric and nonparametric approaches to estimate the spectral density of a nonstationary spatial process, and we study the properties of the proposed estimates. In section 4, we consider the problem of testing a given spatial process for stationarity. The approach is based on the spatial spectral analysis. Section 5 is an application of the methodology presented in this paper to atmospheric air pollution data obtained from EPA. The source of air pollution data in this article is from the regional scale air quality models, known as Models-3, that are run by EPA and the U.S. States. Models-3 provide pollutants concentrations and fluxes in regular grids in parts of the US (see Figure 1). The lattice structure of Models-3 is perfectly suited to the spectral methods presented in this article. In Section 6 we present some conclusions and final remarks.

2 Representation of a Nonstationary process

2.1 Spatial spectra

A random field Z in $\mathbb{R}^2$ is called *weakly* stationary (or stationary), if it has finite second moments, its mean function is constant and it possesses an autocovariance function C, such that $C(\mathbf{x}-\mathbf{y}) = \text{cov}\{Z(\mathbf{x}), Z(\mathbf{y})\}$. If $C(h) = C_0(|h|)$ for some function C_0, then the process is called isotropic.

The spectral representation of a random process $Z(\mathbf{x})$ is always interpreted as its representation in the form of superposition of sine and cosine waves of different frequencies $\boldsymbol{\omega}$. If Z is a stationary random field with autocovariance C, then we can represent the process in the form of the following Fourier-Stieltjes integral:

$$Z(\mathbf{x}) = \int_{\mathbb{R}^2} \exp(i\mathbf{x}^T\boldsymbol{\omega})dY(\boldsymbol{\omega}) \tag{1}$$

where Y are random functions with uncorrelated increments, (see Yaglom (1987), Cramér and Leadbetter (1967) for example). Using the spectral representation of Z and proceeding formally,

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

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

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

If F has a density with respect to Lebesgue measure, this density is the spectral density, f, which is the Fourier transform of the autocovariance function:

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

Each random process (not necessarily a stationary process) having a covariance function of the form (2) where F is a function of bounded variation is *harmonizable*, i.e. such a process is representable as the Fourier-Stieltjes integral (1), where Z is a random function whose spectral function coincides with F.

There have been several attempts to define a spectrum for non-stationary one dimensional processes, generally the object has been to obtain a single function whose properties depended on the behavior of the process over the whole parameter space. Cramér (1960) considered the class of time series which are harmonizable. In the approach we propose here, we define a spectral quantity whose physical interpretation is similar to that of the spectrum of a stationary process. The nonstationary process is represented in terms of a common underlying process with an amplitude varying with location. This is an extension of the "evolutionary spectra" for time series introduced by Priestley (1965 and 1966) to spatial processes.

We define a new class of generalized spectral representations for nonstationary processes that we call spatial spectral representation. This is an extension of the evolutionary spectra in time series to spatial processes. For this new class of spatial models the spectral representation itself and the corresponding spectral distribution function (or spectral density) can change slowly on space. Let Z be a nonstationary process, we propose the following representation

$$Z(\mathbf{x}) = \int_{\mathbb{R}^2} \exp(i\mathbf{x}^T\boldsymbol{\omega}) \phi_{\mathbf{x}}(\boldsymbol{\omega}) dY(\boldsymbol{\omega}) \tag{3}$$

where Y is an orthogonal process, so it satisfies the relation

$$E\left[Y(\boldsymbol{\omega})Y^c(\boldsymbol{\omega}')\right] = \delta(\boldsymbol{\omega} - \boldsymbol{\omega}')d\mu(\boldsymbol{\omega}')$$

where δ is the delta Dirichlet function, and Y^c denotes the conjugate of Y. We assume that the measure $\mu(\boldsymbol{\omega})$ is absolutely continuous with respect to Lebesgue measure, and that $\phi_{\mathbf{x}}(\boldsymbol{\omega})$ are "slowly varying" functions of $\mathbf{x}$ satisfying the condition

$$\int_{\mathbb{R}^2} |\phi_{\mathbf{x}}(\boldsymbol{\omega})|^2 d\mu(\boldsymbol{\omega}) < \infty \tag{4}$$

for all $\mathbf{x}$.

It is easy to see that then the covariance function, C, of the nonstationary process $Z(\mathbf{x})$ is given by the formula

$$\operatorname{cov}\{Z(\mathbf{x}_1), Z(\mathbf{x}_2)\} = C(\mathbf{x}_1, \mathbf{x}_2) = \int_{\mathbb{R}^2} \exp\{i(\mathbf{x}_1 - \mathbf{x}_2)^T \boldsymbol{\omega})\} \phi_{\mathbf{x}_1}(\boldsymbol{\omega}) \phi^c_{\mathbf{x}_2}(\boldsymbol{\omega}) d\mu(\boldsymbol{\omega}) \tag{5}$$

In particular

$$\operatorname{var}\{Z(\mathbf{x})\} = C(\mathbf{x}, \mathbf{x}) = \int_{\mathbb{R}^2} |\phi_{\mathbf{x}}(\boldsymbol{\omega})|^2 d\mu(\boldsymbol{\omega}) \tag{6}$$

so that condition (4) is necessary for the variance of $Z(\mathbf{x})$ to be finite at all $\mathbf{x}$, i.e. for the existence of a covariance function $C(\mathbf{x}_1, \mathbf{x}_2)$. The representation may be interpreted as representation of the process Z in the form of a superposition of sinusoidal oscillations with different frequencies $\boldsymbol{\omega}$ and random amplitudes $\phi_{\mathbf{x}}(\boldsymbol{\omega}) dY(\boldsymbol{\omega})$ varying over space, i.e. $Z(\mathbf{x})$ is a superposition of amplitude modulated random oscillations. According to this interpretation, (6) describes the distribution of the "total power" of the process $Z(\mathbf{x})$ at location $\mathbf{x}$ over the frequencies $\boldsymbol{\omega}$, hence the contribution from the frequency $\boldsymbol{\omega}$ is $|\phi_{\mathbf{x}}(\boldsymbol{\omega})|^2 d\mu(\boldsymbol{\omega})$. Therefore the function $F_{\mathbf{x}}(\boldsymbol{\omega})$ defined by the relation

$$dF_{\mathbf{x}}(\boldsymbol{\omega}) = |\phi_{\mathbf{x}}(\boldsymbol{\omega})|^2 d\mu(\boldsymbol{\omega}) \tag{7}$$

will be called the spatial spectral distribution function of the process Z, and $f_{\mathbf{x}}(\boldsymbol{\omega}) = |\phi_{\mathbf{x}}(\boldsymbol{\omega})|^2 h(\boldsymbol{\omega})$ is the spatial spectral density of Z, where h is the density associated to the measure μ.

Since $\phi_{\mathbf{x}}(\boldsymbol{\omega})$ is a slowly varying function of space, it is clear that the process Z may be regarded as being "approximately stationary" within subregions in D. If, however, we examine the behavior of Z within two subregions which are sufficiently far apart, we could find that although Z is practically stationary in both subregions, the spectral distribution function of the two "portions" of Z will, in general, be different (i.e., the spectral distribution of the power of Z varies on space).

There exits different representations of the form (3) for a nonstationary process Z, each representation based on a different family of $\phi_{\mathbf{x}}(\boldsymbol{\omega})$ functions. This problem is similar to the selection of a basis for a vector space. Apart from that, it would not be physically meaningful to interpret $\boldsymbol{\omega}$ as the frequency in all cases. In the physical theory of oscillations the function $A_{\mathbf{x}}(\boldsymbol{\omega}) = \phi_{\mathbf{x}}(\boldsymbol{\omega})\exp(i\mathbf{x}^T\boldsymbol{\omega})$ is said to describe the amplitude modulated oscillation of frequency $\boldsymbol{\omega}$ only if the "amplitude" $\phi_{\mathbf{x}}(\boldsymbol{\omega})$ is a slowly varying compared to $\exp(i\mathbf{x}^T\boldsymbol{\omega})$ function, i.e. if the Fourier transform of $\phi_{\mathbf{x}}(\boldsymbol{\omega})$ as a function of $\mathbf{x}$ includes mainly frequencies much lower than $\boldsymbol{\omega}$, it is even often assumed that this transform must be concentrated in a neighborhood of zero frequency. We restrict the permissible variability of the function $\phi_{\mathbf{x}}(\boldsymbol{\omega})$ of $\mathbf{x}$ by considering only functions $\phi_{\mathbf{x}}(\boldsymbol{\omega})$ that admit a generalized Fourier representation

$$\phi_{\mathbf{x}}(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} e^{i\mathbf{x}^T\boldsymbol{\tau}} dH_{\boldsymbol{\omega}}(\boldsymbol{\tau}) \tag{8}$$

with $|dH_{\boldsymbol{\omega}}(\boldsymbol{\tau})|$ having its maximum at $\boldsymbol{\tau} = \mathbf{0}$ for any fixed $\boldsymbol{\omega}$. This condition guarantees that the Fourier transform of $\phi_{\mathbf{x}}(\boldsymbol{\omega})$ as a function of $\mathbf{x}$ includes mainly frequencies much lower than any $\boldsymbol{\omega}$, and it has been suggested by Priestley in the time series context. Since the functions $\phi_{\mathbf{x}}(\boldsymbol{\omega}) = 1$ clearly satisfy the conditions to be imposed on $\phi_{\mathbf{x}}(\boldsymbol{\omega})$, the representation (3) certainly includes all the stationary processes having a finite variance.

Alternatively, we suggest here a different constraint for the two-dimensional case. We impose the following condition on $dH_{\boldsymbol{\omega}}(\boldsymbol{\tau})$, instead of assuming that that $|dH_{\boldsymbol{\omega}}(\boldsymbol{\tau})|$ has its maximum at $\boldsymbol{\tau} = \mathbf{0}$. We assume $dH_{\boldsymbol{\omega}}(\boldsymbol{\tau})$ is of bounded variation and

$$\int\int\int |dH_{\boldsymbol{\omega}}(\boldsymbol{\tau})||dH_{\boldsymbol{\omega}}(\boldsymbol{\tau})|dF(\boldsymbol{\omega}) < \infty \tag{9}$$

then the corresponding covariance function is necessarily a harmonizable covariance function.

2.2 Relation to other approaches

A broad class of stationary Gaussian processes may be represented in the form

$$Z(\mathbf{x}) = \int K(\mathbf{x} - \mathbf{u})W(\mathbf{u})d\mathbf{u},$$

with $K(\cdot)$ some positive kernel function and $W(\cdot)$ a constant-variance Gaussian white noise process. Higdon, Swall and Kern (1999) considered extensions of the form

$$Z(\mathbf{x}) = \int K_{\mathbf{x}}(\mathbf{u})W(\mathbf{u})d\mathbf{u}, \tag{10}$$

where the kernel $K_{\mathbf{x}}$ depends on position $\mathbf{x}$. The idea of Higdon *et al.* was to model $K_{\mathbf{x}}(\mathbf{u})$ as an unknown function in terms of specific parameters which can then be estimated in a hierarchical Bayes framework. In the case where $K_{\mathbf{x}}$ is a Gaussian kernel for each $\mathbf{x}$, this leads to tractable expressions for the covariance function and hence the likelihood function for the process. Ver Hoef and Barry (1998) used a similar moving average specification of a Gaussian process for cokriging and multivariable spatial prediction. The development of Higdon *et al.* relies heavily on the Gaussian form of kernel function and it is not clear how restrictive this is. The approach proposed here has a quite different motivation, but it also includes model (10) as a particular case. In model (3), if we treat $A_{\mathbf{x}}(\boldsymbol{\omega})$ as a positive kernel function and Y as a Wiener process, we obtain representation (10). However, our representation is more general, $A_{\mathbf{x}}$ is a complex function and Y a process with orthogonal increments (not necessarily a Wiener process).

The nonstationary model in (10) is quite different idea from earlier approaches for nonstationarity; see for example Sampson and Guttorp (1992), Guttorp and Sampson (1994), Haas (1995), and Nychka *et al.* (1999). But it also has the disadvantage of not being easily related to traditional spatial models. In the approach proposed by Fuentes (2001), the nonstationary process Z is modeled locally as a stationary isotropic random field with some parameters that describe the local spatial structure. These parameters are allowed to vary across space and reflect the lack of stationarity of the process. Thus, the space domain of interest D is divided into small subgrids $S_1, \ldots, S_M$ centered at the nodes $\mathbf{s}_1, \ldots, \mathbf{s}_M$, and Z is represented as a weighted average of orthogonal stationary processes:

$$Z(\mathbf{x}) = \sum_{i=1}^{M} Z_{\boldsymbol{\theta}(\mathbf{s}_i)}(\mathbf{x}) K(\mathbf{x} - \mathbf{s_i}) \tag{11}$$

where $Z_{\boldsymbol{\theta}(\mathbf{s}_i)}$ is a Gaussian stationary process with a covariance function $C_{\boldsymbol{\theta}(\mathbf{s}_i)}$, that explains the spatial structure of Z in the subregion S_i, and $K(\mathbf{x} - \mathbf{s_i})$ is a nonnegative weight function. The parameter function $\boldsymbol{\theta}$ is allowed to vary across space to reflect the lack of stationarity of the process. This representation of a nonstationary process was first introduced by Fuentes (2001), and then Fuentes (2002a) used it to propose a new spatial periodogram for nonstationary processes. Fuentes and Smith (2001) generalized this representation of a nonstationary process by allowing the parameter function $\boldsymbol{\theta}$ to vary continuously across space.

We derive the spectral representation of the nonstationarity model (11) and we compare it to the spectral model (3) presented here. The processes $\{Z_{\boldsymbol{\theta}(\mathbf{s}_i)}\}$ are assumed to be orthogonal and stationary with absolutely summable covariance functions, so they admit a representation

$$Z_{\boldsymbol{\theta}(\mathbf{s}_i)}(\mathbf{x}) = \int \exp(i\boldsymbol{\omega}^T\mathbf{x})dY_{\boldsymbol{\theta}(\mathbf{s}_i)}(\boldsymbol{\omega}). \tag{12}$$

and $E|dY_{\boldsymbol{\theta}(\mathbf{s}_i)}(\boldsymbol{\omega})|^2 = \phi^2_{\boldsymbol{\theta}(\mathbf{s}_i)}(\boldsymbol{\omega})d\boldsymbol{\omega}$. The processes $dY_{\boldsymbol{\theta}(\mathbf{s}_i)}(\boldsymbol{\omega})$ for $i = 1, \ldots, M$ are pairwise orthogonal because the processes $\{Z_{\boldsymbol{\theta}(\mathbf{s}_i)}\}$ are orthogonal. We obtain the spectral representation of Z

$$Z(\mathbf{x}) = \int \exp(i\boldsymbol{\omega}^T\mathbf{x})dY_{\mathbf{x}}(\boldsymbol{\omega}), \tag{13}$$

where

$$dY_{\mathbf{x}}(\boldsymbol{\omega}) = \sum_{i=1}^{M} K(\mathbf{x}-\mathbf{s_i})dY_{\boldsymbol{\theta}(\mathbf{s}_i)}(\boldsymbol{\omega}).$$

Then for each $\mathbf{x}$, $Y_{\mathbf{x}}(\boldsymbol{\omega})$ is a process with orthogonal increments; that is, $E\left[dY_{\mathbf{x}}(\boldsymbol{\omega})dY^c_{\mathbf{x}}(\boldsymbol{\omega}')\right] = 0$, and $E|dY_{\mathbf{x}}(\boldsymbol{\omega})|^2 = \phi^2_{\mathbf{x}}(\boldsymbol{\omega})d\boldsymbol{\omega}$, where $\{\phi^2_{\mathbf{x}}(\boldsymbol{\omega})\}$ can be defined in terms of the kernel functions as

$$\phi^2_{\mathbf{x}}(\boldsymbol{\omega}) = \sum_{i=1}^{M} K^2(\mathbf{x}-\mathbf{s_i})\phi^2_{\boldsymbol{\theta}(\mathbf{s}_i)}(\boldsymbol{\omega}). \tag{14}$$

We also have

$$E\left[dY_{\mathbf{x}}(\boldsymbol{\omega})dY^c_{\mathbf{y}}(\boldsymbol{\omega})\right] = \sum_{i=1}^{M} K(\mathbf{x}-\mathbf{s_i})K(\mathbf{y}-\mathbf{s_i})\phi^2_{\mathbf{s}_i}(\boldsymbol{\omega})d\boldsymbol{\omega} \tag{15}$$

and $E\left[dY_{\mathbf{x}}(\boldsymbol{\omega})dY^c_{\mathbf{y}}(\boldsymbol{\omega}')\right] = 0$.

We define

$$\phi_{\mathbf{x},\mathbf{y}}(\boldsymbol{\omega})d\boldsymbol{\omega} = E\left[dY_{\mathbf{x}}(\boldsymbol{\omega})dY^c_{\mathbf{y}}(\boldsymbol{\omega})\right]; \tag{16}$$

then, we have

$$\phi_{\mathbf{x},\mathbf{y}}(\boldsymbol{\omega})d\boldsymbol{\omega} = \sum_{i=1}^{M} K(\mathbf{x}-\mathbf{s_i})K(\mathbf{y}-\mathbf{s_i})\phi^2_{\mathbf{s}_i}(\boldsymbol{\omega})d\boldsymbol{\omega}.$$

We can derive a spectral representation for the covariance of Z in terms of the functions $\{\phi_{\mathbf{x},\mathbf{y}}(\boldsymbol{\omega})\}$:

$$\begin{aligned}
\mathrm{cov}(Z(\mathbf{x}), Z(\mathbf{y})) &= \sum_{i=1}^{M} K(\mathbf{x}-\mathbf{s_i})K(\mathbf{y}-\mathbf{s_i})C_{\boldsymbol{\theta}(\mathbf{s}_i)}(\mathbf{x}-\mathbf{y}) \\
&= \sum_{i=1}^{M} K(\mathbf{x}-\mathbf{s_i})K(\mathbf{y}-\mathbf{s_i}) \int \exp\{i\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y})\}\phi^2_{\boldsymbol{\theta}(\mathbf{s}_i)}(\boldsymbol{\omega})d\boldsymbol{\omega} \\
&= \int \exp\{i\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y})\}E\left[dY_{\mathbf{x}}(\boldsymbol{\omega})dY^c_{\mathbf{y}}(\boldsymbol{\omega})\right] \\
&= \int \exp\{i\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y})\}\phi_{\mathbf{x},\mathbf{y}}(\boldsymbol{\omega})d\boldsymbol{\omega},
\end{aligned} \tag{17}$$

where the third equality is obtained from (15) and the last expression in (17) is obtained from the definition of $\phi_{\mathbf{x},\mathbf{y}}(\boldsymbol{\omega})$ in (16). Therefore, by comparing (17) to (5) we conclude that the representation (11) is not reducible to one of the new form (3) presented here ($\phi_{\mathbf{x},\mathbf{y}}$ in (17) is not a separable function of $\mathbf{x}$ and $\mathbf{y}$ as in (5)). Fuentes (2002b) provides more detail about the spectrum of model (11).

The spatial spectral representation (3) proposed here is a promising approach for spatial processes in two or higher dimensions. One of the attractive features of the nonstationary model (3) is that it provides a natural framework to easily fit and estimate the spatial structure of a nonstationary process (Section 3) and to test for stationarity (Section 4). Apart from all the enormous computational benefits obtained by the use of the Fourier transform.

3 Estimating the covariance

3.1 Parametric estimate

Consider a nonstationary process Z, in this Section we present a parametric approach to estimate the spectral density and the covariance of Z, based on the concept of "spatial spectra". Since Z is nonstationary the spectral density of the process Z is space dependent. A class of practical variograms and autocovariance functions can be obtained from the Matérn class of spectral densities. We denote $f_i(\boldsymbol{\omega})$ the spectral density of $Z(\mathbf{x}_i)$,

$$f_i(\boldsymbol{\omega}) = \phi_i(\alpha_i^2 + |\boldsymbol{\omega}|^2)^{\left(-\nu_i - \frac{d}{2}\right)} \tag{18}$$

with parameters $\nu_i > 0$, $\alpha_i > 0$ and $\phi_i > 0$ (d is the dimensionality). Here, the vector of covariance parameters is $\boldsymbol{\theta}_i = (\phi_i, \nu_i, \alpha_i)$. The parameter α_i^{-1} can be interpreted as the autocorrelation range. The parameter ν_i measures the degree of smoothness of the process Z at $\mathbf{x_i}$. The parameter ϕ_i is proportional to the ratio of the variance σ_i^2 (also called the *sill* parameter) and the range (α_i^{-1}) to the $2\nu_i^{\text{th}}$ power, $\phi_i = \sigma_i^2 \alpha_i^{2\nu}$. The parameter $\boldsymbol{\theta}_i$ changes with location.

The smoothness of a random field, the parameter ν_i in the Matérn class, plays a critical role in interpolation problems (Stein, 1999). A number of commonly used models for the covariance structure, including spherical, exponential and Gaussian structures assume that the smoothness parameter is known a priori.

The first step in our analysis will be to estimate this parameter from the data, instead of considering it known.

We assume a parametric model for the spectral density, a Matérn with parameter $\boldsymbol{\theta}_i$ changing with location. We could model $\boldsymbol{\theta}$ as a function of location, $\boldsymbol{\theta}(\mathbf{s})$ for $\mathbf{s} \in \mathbf{D}$. This parameter function $\boldsymbol{\theta}(\mathbf{s})$ for $\mathbf{s} \in \mathbf{D}$, measures the lack of stationarity of the process Z. Then, it would be natural to treat $\boldsymbol{\theta}(\mathbf{s})$ as a stochastic process, with correlated errors. Alternatively, we could use a spline approach to model $\boldsymbol{\theta}(\mathbf{s})$, though the approach presented here allows splines as a special case. We model the parameter function $\boldsymbol{\theta}$ as a spatial process with correlated errors,

$$\boldsymbol{\theta}(\mathbf{s}) = \mu_{\boldsymbol{\theta}}(\mathbf{s}) + \epsilon_{\boldsymbol{\theta}}(\mathbf{s}) \tag{19}$$

$\mu_{\boldsymbol{\theta}}(\mathbf{s})$ represents the spatial trend (the large scale structure) and the process $\epsilon_{\boldsymbol{\theta}}(\mathbf{s})$ represents some spatially correlated zero-mean noise. We assume here $\mu_{\boldsymbol{\theta}}(\mathbf{s})$ is a polynomial function in $\mathbf{s}$ with coefficients, β_0, and the process $\epsilon_{\boldsymbol{\theta}}(\mathbf{s})$ is Gaussian with mean zero and a Matérn stationary covariance, $\mathrm{cov}(\epsilon_{\boldsymbol{\theta}}(\mathbf{x}+\mathbf{y}), \epsilon_{\boldsymbol{\theta}}(\mathbf{x})) = C_{\boldsymbol{\tau}_0}(\mathbf{y})$, with parameters $\boldsymbol{\tau}_0$.

We consider a hierarchical Bayesian approach to model and take into the account the spatial structure of the parameter $\boldsymbol{\theta}(\mathbf{s})$ when estimating $f_{\mathbf{x}}$.

Stage 1. β_0, $\boldsymbol{\tau}_0$ are the hyperparameters, we need to specify the priors.

Stage 2. $[\boldsymbol{\theta}|\beta_0, \boldsymbol{\tau}_0]$ is Gaussian

Stage 3. $[Z|\boldsymbol{\theta}]$ is Gaussian with covariance (5).

The parameter function $\boldsymbol{\theta}$ could be also modeled using a Bayesian version of the median polish approach (e.g. Cressie, 1993). Assume the process Z is observed in a grid $n_1 \times n_2$, then we propose the following model for the parameter function $\boldsymbol{\theta}$

$$\boldsymbol{\theta}(\mathbf{s}) = a + r_i + c_j + \epsilon_{\boldsymbol{\theta}}(\mathbf{s}) \tag{20}$$

where $\mathbf{s} = (s_i, s_j)$, for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$. The hyperparameters a, r_i for $i = 1, \ldots, n_1$, and c_j for $j = 1, \ldots, n_2$, are unknown and correspond to the β_0 parameter function in *Stage 1* of the hierarchical model. Once we determine the values of $\boldsymbol{\theta}(\mathbf{x})$ for $\mathbf{x} \in D$ using a likelihood or a Bayesian approach, the spectral density (and the corresponding covariance) of Z would be completely specified.

3.2 Nonparametric estimate

Consider a spatial process Z. We observe the process at N equally spaced locations in a regular grid D $(n_1 \times n_2)$, where $N = n_1 n_2$, and the spacing between observations is Δ. Let $dH_{\boldsymbol{\omega}}(\boldsymbol{\theta})$ denote the Fourier transform of $\phi_{\mathbf{x}}(\boldsymbol{\omega})$ as a function of $\mathbf{x}$, so that we may write, for each $\boldsymbol{\omega}$,

$$\phi_{\mathbf{x}}(\boldsymbol{\omega}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\{i\mathbf{x}^T\boldsymbol{\theta}\} dH_{\boldsymbol{\omega}}(\boldsymbol{\theta})$$

We define the characteristic width of the process Z,

$$B_Z = [\sup_{\boldsymbol{\omega}}\{B_H(\boldsymbol{\omega})\}]^{-1}$$

where

$$B_H(\boldsymbol{\omega}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |\boldsymbol{\theta}||dH_{\boldsymbol{\omega}}(\boldsymbol{\theta})|.$$

The characteristic width of Z may be interpreted roughly as the maximum area over which the process may be treated as approximately stationary. Dahlhaus (1996b) allows the characteristic width to tend to infinity as we increase the sample size. We could use a similar approach here, but we prefer to treat the characteristic width as a geographic area of stationarity which should not be altered by increasing the sample size.

We propose a nonparametric estimate of the spectral density. We first define $J_{\mathbf{x}}(\boldsymbol{\omega}_0)$,

$$J_{\mathbf{x}}(\boldsymbol{\omega}_0) = \Delta \sum_{u_1=0}^{n_1}\sum_{u_2=0}^{n_2} g(\mathbf{x} - \Delta\mathbf{u}) Z(\Delta\mathbf{u}) \exp\{-i\Delta\mathbf{u}^T\boldsymbol{\omega}_0\}, \tag{21}$$

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

$$J_N = \{\lfloor -(n_1-1)/2 \rfloor, \ldots, n_1 - \lfloor n_1/2 \rfloor\} \times \{\lfloor -(n_2-1)/2 \rfloor, \ldots, n_2 - \lfloor n_2/2 \rfloor\}. \tag{22}$$

where $\lfloor u \rfloor$ denotes the largest integer less or equal than u. The spectral estimate $|J_{\mathbf{x}}(\boldsymbol{\omega})|^2$ is an approximately unbiased estimate of $f_{\mathbf{x}}(\boldsymbol{\omega})$ (see Theorem 1), 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 satisfies

C.1 $W_\rho(\mathbf{s}) \geq 0$, for all $\mathbf{s}, \rho$.

C.2 $W_\rho(\mathbf{s})$ decays to zero as $|\mathbf{s}| \to \infty$, for all ρ.

C.3

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_\rho(\mathbf{s})d\mathbf{s} = 1, \qquad \text{for all } \rho.$$

C.4

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \{W_\rho(\mathbf{s})\}^2 d\mathbf{s} < \infty, \qquad \text{for all } \rho.$$

C.5 We write

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

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 effective ”width” of $\{W_\rho\}$ and it is chosen, such that the width of $\{W_\rho\}$ is larger than B_g.

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

$$\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_\rho(\mathbf{x}-\mathbf{s})\left|J_{\mathbf{s}}(\boldsymbol{\omega}_0)\right|^2 d\mathbf{s} \tag{23}$$

In (23) every entry requires an integration. Since each such integration is actually an expectation with respect to a uniform distribution, in practice we propose to evaluate (23) by Monte Carlo integration. We draw an independent set of locations $\mathbf{s}_m$, $m = 1, 2, ..., M$ over the domain of interest D, and approximate (23) with

$$M^{-1}\sum_{m=1}^{M} W_\rho(\mathbf{x}-\mathbf{s}_m)\left|J_{\mathbf{s}_m}(\boldsymbol{\omega}_0)\right|^2 . \tag{24}$$

Thus, $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)$ can be interpreted as an average of the total energy of the process contained within a band of frequencies in the region of $\boldsymbol{\omega}_0$ and an region in space in the neighborhood of $\mathbf{x}$.

In the application presented in this article we consider $\{g(\mathbf{s})\}$ for $\mathbf{s} = (s_1, s_2)$ to be a multiplicative filter, i.e. the tensor product of two 1 dimensional filters, $g(\mathbf{s}) = g_1(s_1)g_1(s_2)$, where g_1 is of the form

$$g_1(s) = \begin{cases} 1/\{2\sqrt{h\pi}\} & |s| \le h \\ 0 & |s| > h \end{cases} \tag{25}$$

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

$$W_{1,\rho}(s) = \begin{cases} 1/\rho & -1/2\rho \le s \le 1/2\rho \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

corresponding to the Daniell window.

We will study next the asymptotic properties of $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)$ using a *shrinking asymptotics* model described later in this section, but first we discuss the aliasing phenomenon in the spectral domain.

3.2.1 Aliasing

The spatial spectral functions $f_{\mathbf{x}}$ and $F_{\mathbf{x}}$ are closely related to the spectral decomposition of the process Z into a superposition of harmonic oscillations. However, it is easy to see that such a decomposition cannot be uniquely restored from observations of a continuous process Z on a lattice. Assume we observe $Z(\mathbf{x})$ at locations $\mathbf{x} \in \Delta\mathbb{Z}^2$, where Δ is the distance between neighboring observations, and $\mathbb{Z}^2$ is the integer lattice. Indeed,

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

for any $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{Z}^2$. We simply cannot distinguish an oscillation with an angular frequency $\boldsymbol{\omega}$ from all the oscillations with frequencies $\boldsymbol{\omega} + 2\pi\mathbf{v}/\Delta$. 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 $-\boldsymbol{\pi}/\Delta \leq \boldsymbol{\omega} < \boldsymbol{\pi}/\Delta$. 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. Therefore, we must have a dense sample of observed values (small Δ).

The spectral density $f_{\Delta,\mathbf{x}}$ of the process on the lattice, can be written in terms of the spectral density $f_{\mathbf{x}}$ of the continuous process Z,

$$f_{\Delta,\mathbf{x}}(\boldsymbol{\omega}) = \sum_{\mathbf{Q}\in\mathbb{Z}^2} f_{\mathbf{x}}\left(\boldsymbol{\omega} + \frac{2\pi\mathbf{Q}}{\Delta}\right). \tag{27}$$

Similarly, we can obtain a representation of $\phi_{\Delta,\mathbf{x}}$ in terms of $\phi_{\mathbf{x}}$.

The spatial spectral representation (see (3)) of a nonstationary process $Z(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ is given by

$$Z(\mathbf{x}) = \int_{\mathbb{R}^2} \exp(i\mathbf{x}^T\boldsymbol{\omega})\phi_{\mathbf{x}}(\boldsymbol{\omega})dY(\boldsymbol{\omega}), \tag{28}$$

where Y is a Wiener process. Thus, by combining (27) and (28), we derive the spectral representation of the process on the lattice $Z(\Delta\mathbf{u})$ for $\mathbf{u} \in \mathbb{Z}^{\mathbf{2}}$,

$$Z(\Delta\mathbf{u}) = \int_{\Pi_{\Delta}{}^2} \exp(i\Delta\mathbf{u}^T\boldsymbol{\omega})\phi_{\Delta,\Delta\mathbf{u}}(\boldsymbol{\omega})dY(\boldsymbol{\omega}), \tag{29}$$

where $\Pi_\Delta^2 = [-\pi/\Delta, \pi/\Delta]^2$.

3.2.2 Shrinking asymptotics for 2-d periodogram

In some instances, consistency demands that the length of the grid $(n_1 \times n_2)$ over which the process is observed increase as the number of observations increases. This ensures that the amount of information in the data increases. We should also ask that the spacing between neighboring observations, Δ, goes to 0 as the number of observations increases. This guarantees that an accurate picture of the covariance function and spectral density can be developed non parametrically, assuming only smoothness conditions. Therefore, our asymptotic model supposes that the observed data represent a realization of $Z(\mathbf{u}\Delta)$ for $\mathbf{u} = (u_1, u_2)$, where $0 \leq u_1 \leq n_1$, and $0 \leq u_2 \leq n_2$. We assume that Δ goes to zero as $n_1 \to \infty$, $n_2 \to \infty$, and $n_1/n_2 \to \lambda$, for a constant $\lambda > 0$, and that $\Delta n_1 \to \infty$, $\Delta n_2 \to \infty$. We call this type of asymptotics *shrinking asymptotics*. This is a mixture of increasing-domain (eg. Cressie 1993) and fixed-domain asymptotics (Stein, 1995) in that the distance between neighboring observations tends to zero and the size of the observation region grows without bound. Constantine and Hall (1994) used this asymptotic model for one dimensional processes.

Throughout this paper, we study asymptotic properties using the *shrinking asymptotics* model described above. In the next theorem we study the asymptotic behavior of the spatial periodogram.

We introduce now some notation, for each $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, we write $\|\mathbf{x}\|$ to denote the Euclidean norm $\sqrt{x_1^2 + x_2^2}$.

Theorem 1

Consider a Gaussian process $Z(\mathbf{x})$, with covariance $\operatorname{cov}(Z(\mathbf{x}), Z(\mathbf{y})) = C(\mathbf{x}, \mathbf{y})$ and spatial spectral density $f_{\mathbf{x}}(\boldsymbol{\omega})$. The process Z is observed at N equally spaced locations in a regular grid $(n_1 \times n_2)$, where $N = (1 + n_1)(1 + n_2)$ and the spacing between observations is Δ. We define for $\boldsymbol{\omega} \in [-\pi/\Delta, \pi/\Delta]^2$ a spatial periodogram function, $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$, as in (23) to estimate the spectral density of the process observed on the lattice $Z(\Delta\mathbf{u})$. We consider two filters $\{g(\mathbf{s})\}$ and $\{W(\mathbf{s})\}$ satisfying conditions B.1-B.2 and C.1-C.5 respectively.

We assume:

A.1 The rate of decay of the spectral density $f_{\mathbf{x}}(\boldsymbol{\omega})$ at high frequencies is proportional to $\|\boldsymbol{\omega}\|^{-\tau_{\mathbf{x}}}$, where $\tau_{\mathbf{x}} > 2$, for all $\mathbf{x}$.

A.2 The Fourier transform of $f_{\mathbf{x}}(\boldsymbol{\omega})$ as a function of $\boldsymbol{\omega}$, $C_{\mathbf{x}}(\mathbf{s})$, satisfies the following inequality:

$$\sup_{\mathbf{x}} \int ||\mathbf{s}|| |C_{\mathbf{x}}(\mathbf{s})| d\mathbf{s} < \infty.$$

A.3 $\Delta \to 0$, $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \lambda$ for a constant $\lambda > 0$, $\Delta n_1 \to \infty$, and $\Delta n_2 \to \infty$.

A.4 We assume $B_g/B_Z \to 0$, i.e. the spectrum is changing very slowly over the effective range of the filter $\{g(\mathbf{u})\}$.

A.5 We assume $B_g/\rho \to 0$, so that the "width" of $W_\rho(\mathbf{s})$ is much larger than B_g.

Then we get:

- (i) The expected value of the smoothed spatial periodogram, $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$, for $\boldsymbol{\omega} \in [-\pi/\Delta, \pi/\Delta]^2$, is asymptotically $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \tilde{f}_{\mathbf{x}}(\boldsymbol{\omega} + \boldsymbol{\theta}) |\Gamma(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}$ where

$$\tilde{f}_{\mathbf{x}}(\boldsymbol{\omega} + \boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega} + \boldsymbol{\theta}) W_\rho(\mathbf{s}) d\mathbf{s}$$

- (ii) The asymptotic variance of $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ is

$$(C/\rho^2) \tilde{f}^2_{\mathbf{x}}(\boldsymbol{\omega}) \left\{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\Gamma(\boldsymbol{\theta})|^4 d\boldsymbol{\theta} \right\},$$

  where

$$\tilde{f}^2_{\mathbf{x}}(\boldsymbol{\omega}) = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f^2_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}) \{W_\rho(\mathbf{s})\}^2 d\mathbf{s}}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \{W_\rho(\mathbf{s})\}^2 d\mathbf{s}.}$$

- (iii) The spatial periodogram values $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ and $\hat{f}_{\mathbf{y}}(\boldsymbol{\omega}')$ are asymptotically uncorrelated, if either:

  $||\boldsymbol{\omega} \pm \boldsymbol{\omega}'||$ is sufficiently large so that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\Gamma(\boldsymbol{\theta} + \boldsymbol{\omega})|^2 |\Gamma(\boldsymbol{\theta} + \boldsymbol{\omega}')|^2 d\boldsymbol{\omega} \sim 0, \tag{30}$$

  i.e. if $||\boldsymbol{\omega} \pm \boldsymbol{\omega}'|| \gg$ bandwidth of $|\Gamma(\boldsymbol{\theta})|^2$ or

  $||\mathbf{x} - \mathbf{y}|| \gg$ effective "width" of the function $\{W_\rho(\mathbf{u})\}$.

Condition A.1 is always satisfied by the Matérn spectral density (18), we refer to Section 3.1 for more information and discussion about the Matérn spectral density. Condition A.2 implies that the spectral density

$f_{\mathbf{x}}$ has a bounded derivative for all $\mathbf{x}$, condition A.3 describes the asymptotic model used here (*shrinking asymptotics*). Condition A.4 ensures that the spectrum is changing very slowly over the effective range of the filter $\{g(\mathbf{s})\}$, and condition A.5 assumes that the "width" of $W_\rho(\mathbf{s})$ is much larger than the "width" of $\{g(\mathbf{s})\}$. The bandwidth of $|\Gamma(\boldsymbol{\theta})|^2$ is related to the magnitude of $1/B_g$ and it can be approximated by the distance between the "half-power" points on the main lobe of $|\Gamma(\omega_1)||^2$ (Priestley and Rao, 1969).

As a consequence of Theorem 1. If the "bandwidth" of $|\Gamma(\boldsymbol{\theta})|^2$ is small compared with the "frequency domain bandwidth" of $f_{\mathbf{x}}(\boldsymbol{\omega})$, that is if $f_{\mathbf{x}}(\boldsymbol{\omega})$, is smooth compared with $|\Gamma(\boldsymbol{\theta})|^2$ for each $\mathbf{x}$. Then, the expected value of the smoothed spatial periodogram, $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$, for $\boldsymbol{\omega} \in [-\pi/\Delta, \pi/\Delta]^2$, is approximately $f_{\mathbf{x}}(\boldsymbol{\omega})$.

If the "bandwidth" of $|W_\rho(\mathbf{s})|^2$ is small compared with the "space-domain bandwidth" of $f_{\mathbf{x}}(\boldsymbol{\omega})$, Then, $\tilde{f}^2_{\mathbf{x}}(\boldsymbol{\omega})$, is approximately $f^2_{\mathbf{x}}(\boldsymbol{\omega})$, and the asymptotic variance of $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ is

$$(C/\rho^2) f^2_{\mathbf{x}}(\boldsymbol{\omega}) \left\{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\Gamma(\boldsymbol{\theta})|^4 d\boldsymbol{\theta} \right\}.$$

4 Test for non-stationarity

Using the novel asymptotic results obtained in the previous section, 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 (1968) to spatial processes. We first estimate 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

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

We have obtained (Theorem 1) that asymptotically $E(\epsilon(\mathbf{s}_i, \boldsymbol{\omega})) = 0$ and $\mathrm{var}\{\epsilon(\mathbf{s}_i, \boldsymbol{\omega})\} = \sigma^2$ where

$$\sigma^2 = (C/\rho^2) \left\{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\Gamma(\boldsymbol{\theta})|^4 d\boldsymbol{\theta} \right\}, \tag{31}$$

for $\boldsymbol{\omega} \notin \delta\Pi^2_\Delta$, where $\Pi_\Delta = [-\pi, \pi]$ and $\delta\Pi^2_\Delta$ denotes the boundary of the region Π^2_Δ. 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 $\|\mathbf{x} - \mathbf{y}\| \gg$ effective "width" of the function $\{W_\rho(\mathbf{u})\}$.

Assuming that the process Z is Gaussian, 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

$$U_{ij} = U(\mathbf{s}_i, \boldsymbol{\omega}_j),$$

$$\log \hat{f}_{\mathbf{s}_i}(\boldsymbol{\omega}_j) = f_{ij},$$

$$\epsilon_{ij} = \epsilon(\mathbf{s}_i, \boldsymbol{\omega}_j).$$

Then we have the model

$$U_{ij} = f_{ij} + \epsilon_{ij}. \tag{32}$$

Equation (32) 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. Wince we know the value of $\sigma^2 = \mathrm{var}\{\epsilon_{ij}\}$, 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. We give now an example of a uniformly modulated spatial process. Supposed that the lack of stationarity of a process Z is caused by the change with location of the *sill* parameter $\sigma^2(\mathbf{x})$, this means that the variance of the process depends on space. Then, $Z(\mathbf{x})/\sigma(\mathbf{x})$ is a stationary process and consequently $Z(\mathbf{x})$ would be an uniformly modulated process.

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 (the *frequency* effect) is significant.

We could test for "complete randomness", this means constant spectra, by testing the "main-effect" β, either at all locations 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.

5 Application

Our goal is to understand and quantify the weekly spatial structure of air pollutants using the output of the regional scale air quality models (Models-3). Models-3 estimates hourly concentrations and fluxes of different pollutants. The primary objective of Models-3 is to improve the environmental management community's ability to evaluate the impact of air quality management practices for multiple pollutants at multiple scales, as part of the regulation process of the air pollutants standards. As an example we examine nitric acid. The spatial domain, D, is a regular grid (69×75), the dimensions of each cell in the grid are 36km $\times$ 36km.

The 69×75 lattice for EPA Models-3 is a two dimensional grid obtained taking into account the effects of the earth's curvature. Models-3 provides the estimated concentration for the middle point of each cell, i.e. $\Delta = 36km$, where Δ is the spacing between observed values. In this example we analyze the spatial structure of a process Z, which is the hourly averaged nitric acid concentrations for the week starting July 11, 1995. Statistical models for output of air quality models are of urgent need. Some applications include the evaluation of the performance of numerical models (by interpolating the model output at the locations where we have observations), the construction of reliable maps of air pollution and metereological variables by combining numerical models output with observations, e.g. by using the method proposed by Fuentes and Raftery (2001), data assimilation and estimation of spatial and temporal patterns of air pollutants.

First, we use the methods proposed in Section 3 to estimate the spatial structure of a process Z. For the purpose of illustrating the techniques presented in this paper, 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 $S_1, \ldots, S_9$, shown in Figure 1. We estimated the parameters of the 9 spectral density Matérn models (18) $\{f_{\mathbf{x_i}}\}$ for $i = 1, \ldots, 9$ by fitting parametric models to the empirical estimates $\{\hat{f}_{\mathbf{x_i}}\}$, we used a non linear least squares approach in the log scale i.e. we fitted a parametric model to $\log\left(\hat{f}_{\mathbf{x_i}}(\boldsymbol{\omega})\right)$ for $i = 1, \ldots, 9$, where $\boldsymbol{\omega} \in [-\pi/\Delta, \pi/\Delta]^2$. Figure 2 shows the fitted models for the spectral densities, $f_{\mathbf{x_i}}$, for $i = 1, \ldots, 9,$. The smoothing parameter represents the rate of decay at high frequencies. We observe a relatively fast rate of decay in Region 6, with smoothness parameter .8, compared for example to Region 5 that has smaller smoothing parameter (.4). The smoothing parameter for each stationary model in this application does not seem to change too much with location, probably because we are analyzing preprocessed spatial data, the output of a physical model (Models-3). When the range parameter is large (eg. Region 8, Region 1), assuming the other parameters are fixed, there is a faster decay at short frequencies, e.g. compare the spectral density in Region 1 (large range) to the spectral density in Region 2 (small range). 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 in Region 8. There is higher spatial variability (large sill) mainly on the great lakes area (Region 8) due to downwinds from sources of pollution, primary Chicago. 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 due to transition from land to ocean. The change with location of the estimate values for the spectral density parameters indicates that this is clearly a nonstationary process.

Now, we implement the test for stationarity. Using the asymptotic results obtained in this paper, we select values of locations $\mathbf{x}$ and frequencies $\boldsymbol{\omega}$ that are sufficiently apart. The estimates, $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ were obtained using expression (23) in which $W_\rho(\mathbf{u})$ is given by (26) with $\rho = 20$ units (1 unit = 36 km), and $g(\mathbf{u})$ is of the form (25) with $h = 3$. The window $|\Gamma(\omega_1)||^2$ has a bandwidth of approximately $\pi/h = \pi/3$. We used the distance between the "half-power" points on the main lobe of $|\Gamma(\omega_1)||^2$ to approximate 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 $\mathbf{x_i}$ are at least 20 units, the sample points $\mathbf{x}_1, \ldots, \mathbf{x}_9$ are the centroids of the 9 equally-dimensioned regions $S_1, \ldots, S_9$, showed in Figure 1. 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 $\boldsymbol{\omega}_9 = (15\pi/20, 15\pi/20)$.

$\mathbf{x}\backslash\boldsymbol{\omega}$	$\boldsymbol{\omega}_1$	$\boldsymbol{\omega}_2$	$\boldsymbol{\omega}_3$	$\boldsymbol{\omega}_4$	$\boldsymbol{\omega}_5$	$\boldsymbol{\omega}_6$	$\boldsymbol{\omega}_7$	$\boldsymbol{\omega}_8$	$\boldsymbol{\omega}_9$
$\mathbf{x}_1$	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 $f_{\mathbf{x}}(\boldsymbol{\omega})$.

We need to calculate σ^2 (see Equation (31)) 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.

In Figure 2 we fitted *isotropic* parametric Matérn models, assuming then that $f_{\mathbf{x}}(\boldsymbol{\omega})$ depends on its vector argument $\boldsymbol{\omega}$ only through its length $\|\boldsymbol{\omega}\|$. 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. 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 $S_1, \ldots, S_9$.

6 Conclusions and final remarks

In this paper we develop a spectral theory for a class of non-stationary processes by introducing the concept of "spatial spectra", this means spectral functions that are space-dependent. We show that the spatial spectrum at each location could be estimated from a sample of a single realization. We propose parametric and nonparametric estimates of the spectral density. We also consider the problem of testing a given spatial process for stationarity.

However, before the spectral estimate $\hat{f}(\boldsymbol{\omega}_0)$ can be evaluated we have to choose the form of the filter $\{g(\mathbf{s})\}$ and the form of the weight function $\{W_\rho(\mathbf{s})\}$. Generally, $\{g(\mathbf{s})\}$ will be chosen from the standard collection of "windows" and will involved a parameter h, so that by adjusting the value of h we can vary the values of B_g. Similarly, we may choose $\{W_\rho(\mathbf{s})\}$ from the same collection of windows with a parameter ρ. Suppose now that we have chosen the mathematical forms of $\{g(\mathbf{s})\}$ and $\{W_\rho(\mathbf{s})\}$. The problem arises as to how we should choose the parameters h and ρ so that the estimate $\hat{f}(\boldsymbol{\omega}_0)$ possesses certain required properties. If there are no requirements for specified degrees of resolvability in either domain, we may choose h and ρ so as to jointly minimize the relative mean-square error of $\hat{f}(\boldsymbol{\omega}_0)$. Thus, subject to the conditions $B_g \ll B_Z$ and $\rho \ll \delta D$, where δD is the diameter of D, we choose h and ρ to minimize the mean square error of $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)$, we could allow the parameters h and ρ to be both space and frequency selective. Note that as

$$\rho \uparrow, \mathrm{var}\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)\} \downarrow, \mathrm{bias}\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)\} \uparrow$$

and as

$$\rho \downarrow, \mathrm{var}\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)\} \uparrow, \mathrm{bias}\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)\} \downarrow .$$

Also, as

$$B_g \uparrow, \mathrm{var}\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)\} \uparrow, \mathrm{bias}\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)\} \downarrow$$

and as

$$B_g \downarrow, \mathrm{var}\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)\} \downarrow, \mathrm{bias}\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}_0)\} \uparrow .$$

The asymptotic results obtained in this paper (Theorem 1) provide an approximation for the mean squared error (mse) of $\hat{f}(\boldsymbol{\omega}_0)$. We suggests using a plug-in approach and replace $f(\boldsymbol{\omega}_0)$ with $\hat{f}(\boldsymbol{\omega}_0)$ in the expression for the mse, to obtain the values of h and ρ that minimize the relative mse of $\hat{f}(\boldsymbol{\omega}_0)$.

An alternative method to choose h and ρ is using the statistical model presented in Section 3.1 and putting priors on the hyperparameters h and ρ. This method adds another stage to the Hierarchical Bayesian parametric approach proposed in Section 3.1.

Appendix: Proofs

Proof of Theorem 1:

Proof of part (i):

In part (i) of Theorem 1 we seek to obtain an asymptotic expression for the expected value of the estimate $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ (Equation (23)).

First, we define,

$$\Gamma_{\mathbf{x},\boldsymbol{\lambda}}(\boldsymbol{\theta}) = \Delta^2 \sum_{v_1=0}^{n_1} \sum_{v_2=0}^{n_2} g(\mathbf{x}-\Delta\mathbf{v}) \left\{ f_{\Delta,\Delta\mathbf{v}}(\boldsymbol{\lambda}) / f_{\Delta,\mathbf{x}}(\boldsymbol{\lambda}) \right\}^{1/2} \exp\{-i(\mathbf{x}-\Delta\mathbf{v}^T)\boldsymbol{\theta}\}$$

The function $J_{\mathbf{x}}(\boldsymbol{\omega})$ is written in equation (21) in terms of the process $Z(\Delta\mathbf{u})$ for $\mathbf{u}$ in the lattice, then by using the representation of $Z(\Delta\mathbf{u})$ given in (29), we get

$$J_{\mathbf{x}}(\boldsymbol{\omega}) = \Delta \int_{\Pi_\Delta} \int_{\Pi_\Delta} \Gamma_{\mathbf{x},\boldsymbol{\omega}+\boldsymbol{\theta}}(\boldsymbol{\theta}) f^{1/2}_{\Delta,\mathbf{x}}(\boldsymbol{\omega}+\boldsymbol{\theta}) \exp\{i\Delta\boldsymbol{\omega}^T\mathbf{x}\} dY(\boldsymbol{\omega}+\boldsymbol{\theta})$$

since the process Y has orthogonal increments, we have

$$\begin{aligned} E\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_\rho(\mathbf{s}) E\left\{ |J_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})|^2 \right\} d\mathbf{s} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_\rho(\mathbf{s}) \int_{\Pi_\Delta}\int_{\Pi_\Delta} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta}) |\Gamma_{\mathbf{x}-\mathbf{s},\boldsymbol{\omega}+\boldsymbol{\theta}}(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} d\mathbf{s}. \end{aligned} \tag{33}$$

By Priestley (1965) Theorem 8.1

$$\int_{\Pi_\Delta}\int_{\Pi_\Delta} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma_{\mathbf{s},\boldsymbol{\omega}+\boldsymbol{\theta}}(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} = \int_{\Pi_\Delta}\int_{\Pi_\Delta} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma_\Delta(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} + 0(B_g/B_Z)$$

were the residual term is uniform in $\mathbf{x}$, Δ and n_1, n_2 and

$$\Gamma_\Delta(\boldsymbol{\theta}) = \Delta^2 \sum_{v_1=0}^{n_1}\sum_{v_2=0}^{n_2} g(\Delta\mathbf{v})\exp\{-i\Delta\mathbf{v}^T\boldsymbol{\theta}\}.$$

Then, we obtain

$$E\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{\Pi_\Delta}\int_{\Pi_\Delta} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma_\Delta(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}\right] W_\rho(\mathbf{s})d\mathbf{s} + 0(B_g/B_Z).$$

Now we study the term in brackets $\int_{\Pi_\Delta}\int_{\Pi_\Delta} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma_\Delta(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}$, which is the expected value for the periodogram I_g of a tapered process $Z_g(\Delta\mathbf{u}) = Z_{\mathbf{x}-\mathbf{s}}(\Delta\mathbf{u})g(\Delta\mathbf{u}-\mathbf{x})$, where the process on the lattice, $Z_{\mathbf{x}-\mathbf{s}}(\Delta\mathbf{u})$, has spectral density $f_{\Delta,\mathbf{x}-\mathbf{s}}$ (which does not depend on $\mathbf{u}$) and covariance $C_{\Delta,\mathbf{x}-\mathbf{s}}$. We define

$$J_g(\boldsymbol{\omega}) = \Delta\sum_{u_1=0}^{n_1}\sum_{u_2=0}^{n_2} Z_g(\Delta\mathbf{u})\exp\{-i\Delta\mathbf{u}^T\boldsymbol{\omega}\}. \tag{34}$$

for $\boldsymbol{\omega}\in[-\pi/\Delta,\pi/\Delta]^2$, then the periodogram of Z_g is $I_g(\boldsymbol{\omega}) = J_g(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega})$. We could write $f_{\Delta,\mathbf{x}-\mathbf{s}}$ in terms of $C_{\Delta,\mathbf{x}-\mathbf{s}}$,

$$f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}) = \Delta^2(2\pi)^{-2}\sum_{\tau_1\in\mathbb{Z}}\sum_{\tau_2\in\mathbb{Z}} C_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\tau})\exp\{-i\Delta\boldsymbol{\tau}^T\boldsymbol{\omega}\} \tag{35}$$

where $\boldsymbol{\tau} = (\tau_1,\tau_2)$. By combining (34) and (35) we obtain the following expression for the expected value of I_g in terms of $C_{\Delta,\mathbf{x}-\mathbf{s}}$ (see Brillinger (1970) for example):

$$\begin{aligned} E\{J(\boldsymbol{\omega})J^c(\boldsymbol{\omega})\} &= \Delta^2 \sum_{\tau_1=-(n_1-1)}^{n_1-1}\sum_{\tau_2=-(n_2-1)}^{n_2-1}\left(\sum_{t_1=\tau_1-n_1}^{\tau_1}\sum_{t_2=\tau_2-n_2}^{\tau_2} g(\mathbf{t}-\boldsymbol{\tau})g(\mathbf{t})\right) C_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\tau})\exp\{-i\Delta\boldsymbol{\tau}^T\boldsymbol{\omega}\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} + \epsilon_{N,\Delta}. \end{aligned} \tag{36}$$

We prove now that as $n_i\to\infty$ and $\Delta n_i\to\infty$ for $i=1,2$, the term $\epsilon_{N,\Delta}$ goes uniformly to zero. First, by Lemma P4.3 in Brillinger (1981, pag. 403), the residual term in (36) can be bounded as follows:

$$|\epsilon_{N,\Delta}| \le \frac{1}{n_1 n_2}\sum_{\tau_1=-2(n_1-1)}^{2(n_1-1)}\sum_{\tau_2=2(n_2-1)}^{2(n_2-1)} \|\Delta\boldsymbol{\tau}\||C_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\tau})| \tag{37}$$

where $\boldsymbol{\tau} = (\tau_1,\tau_2)$. By the definition of $C_{\Delta,\mathbf{x}-\mathbf{s}}$ and under assumption A.2,

$$\frac{1}{n_1 n_2}\sum_{\tau_1=-2(n_1-1)}^{2(n_1-1)}\sum_{\tau_2=2(n_2-1)}^{2(n_2-1)} \|\Delta\boldsymbol{\tau}\||C_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\tau})| \le \frac{1}{N\Delta^2}\int_{\mathbb{R}^2}\|\mathbf{y}\||C_{\mathbf{x}-\mathbf{s}}(\mathbf{y})|d\mathbf{y} = 0(\Delta^{-2}N^{-1}). \tag{38}$$

Thus, by combining results (36),(37) and (38), we obtain that for $\boldsymbol{\omega} \in [-\pi/\Delta, \pi/\Delta]^2$, $E\{J_g(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega})\}$ converges uniformly to $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}$,

$$E\{I_g(\boldsymbol{\omega})\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} + 0(\Delta^{-2}N^{-1}).$$

Thus, we obtain that $\int_{\Pi_\Delta}\int_{\Pi_\Delta} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma_\Delta(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}$ converges uniformly in $\boldsymbol{\omega}$ and in $\mathbf{x}-\mathbf{s}$ to $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}$.

Now, we obtain that as $\Delta \to 0$, and $n_i \to \infty$ for $i = 1, 2$, the spectral density of the sampled data $Z_{\mathbf{x}-\mathbf{s}}(\Delta\mathbf{u})$ for $\mathbf{u} \in \mathbb{Z}^2$, becomes the spectral density of $Z_{\mathbf{x}-\mathbf{s}}(\mathbf{y})$ for $\mathbf{y} \in \mathbb{R}^2$, i.e. $f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})$ converges to $f_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})$. We also address the question of how fast the convergence $f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})$ to $f_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})$ occurs.

Define $\mathbb{Z}^* = \mathbb{Z} - \{0\}$, and $\mathbb{Z}^{*^2} = \mathbb{Z}^2 - \{0\}$, then

$$f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}) - f_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}) = \sum_{\mathbf{Q}\in\mathbb{Z}^2} f_{\mathbf{x}-\mathbf{s}}\left(\boldsymbol{\omega} + \frac{2\pi\mathbf{Q}}{\Delta}\right) - f_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}) = \sum_{\mathbf{Q}\in\mathbb{Z}^{*2}} f_{\mathbf{x}-\mathbf{s}}\left(\boldsymbol{\omega} + \frac{2\pi\mathbf{Q}}{\Delta}\right)$$

Under condition A.1, for Δ small enough the previous expression is proportional to

$$\sum_{\mathbf{Q}\in\mathbb{Z}^{*2}} \left\|\boldsymbol{\omega} + \frac{2\pi\mathbf{Q}}{\Delta}\right\|^{-\tau_{\mathbf{x}-\mathbf{s}}} \tag{39}$$

Under condition A.1 we have $\tau_{\mathbf{x}} > 2$ for all $\mathbf{x}$, then by the integral test (see Stein (1967), pag. 319) the series (39) converges, we get that for $\boldsymbol{\omega} \in [-\pi/\Delta, \pi/\Delta]^2$,

$$\sum_{\mathbf{Q}\in\mathbb{Z}^{*2}} \left\|\boldsymbol{\omega} + \frac{2\pi\mathbf{Q}}{\Delta}\right\|^{-\tau_{\mathbf{x}-\mathbf{s}}} < \sum_{\mathbf{Q}\in\mathbb{Z}^{*2}} \left\|\boldsymbol{\omega} + \frac{2\pi\mathbf{Q}}{\Delta}\right\|^{-2} < \Delta^2 K$$

where K does not depend on $\boldsymbol{\omega}$ or $\mathbf{x}-\mathbf{s}$.

Therefore,

$$E\{J_g(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega})\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} + 0(\Delta^{-2}N^{-1}) + O(\Delta^2). \tag{40}$$

Thus, we obtain that $f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})$ converges to $f_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})$, uniformly in $\mathbf{x}-\mathbf{s}$. This result and the convergence of $\int_{\Pi_\Delta}\int_{\Pi_\Delta} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma_\Delta(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}$ to $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}$, lead to the proof of Theorem 1 part (i), we get

$$E\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}+\boldsymbol{\theta})|\Gamma(\boldsymbol{\theta})|^2 d\boldsymbol{\theta}\right] W_\rho(\mathbf{s})d\mathbf{s} + 0(B_g/B_Z) + 0(\Delta^{-2}N^{-1}) + O(\Delta^2)$$

Proof of part (ii):

The variance of $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ can be written in terms of the covariance for $|J_{\mathbf{x}}(\boldsymbol{\omega})|^2$,

$$\operatorname{var}\left\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})\right\} = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} W_\rho(\mathbf{s})W_\rho(\mathbf{y})\operatorname{cov}\left\{|J_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})|^2, |J_{\mathbf{x}-\mathbf{y}}(\boldsymbol{\omega})|^2\right\} d\mathbf{s}d\mathbf{y} \tag{41}$$

where by Priestley (1966)

$$\operatorname{var}\left\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})\right\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \operatorname{var}(I_g(\boldsymbol{\omega}))\{W_\rho(\mathbf{s})\}^2 d\mathbf{s}\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|\Gamma(\boldsymbol{\theta})|^4 d\boldsymbol{\theta}\right\} + O(B_g/B_Z) + O(B_g/\rho)$$

with the residual terms uniform in n_1, n_2 and Δ. It can be shown that

$$\operatorname{var}(I_g(\boldsymbol{\omega})) = f^2_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}) + O(\Delta^{-2}N^{-1}) \tag{42}$$

To derive (42) we write first the variance of I_g in terms of J_g,

$$\begin{aligned}\operatorname{var}(I_g(\boldsymbol{\omega})) &= \operatorname{cov}\left(J_g(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega}), J_g(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega})\right)\\ &= \operatorname{cov}\left(J_g(\boldsymbol{\omega}), J_g(\boldsymbol{\omega})\right)\operatorname{cov}\left(J_g^c(\boldsymbol{\omega}), J_g^c(\boldsymbol{\omega})\right) + \\ &\quad \operatorname{cov}\left(J_g(\boldsymbol{\omega}), J_g^c(\boldsymbol{\omega})\right)\operatorname{cov}\left(J_g^c(\boldsymbol{\omega}), J_g(\boldsymbol{\omega})\right)\\ &= E\{J_g(\boldsymbol{\omega})J(\boldsymbol{\omega})\}E\{J_g^c(\boldsymbol{\omega})J^c(\boldsymbol{\omega})\} + |E\{J_g(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega})\}|^2.\end{aligned} \tag{43}$$

By combining (36), (37) and (38), as a generalization of Theorem 5.2.4 in Brillinger (1981, pag. 125), we get

$$E\{J_g(\boldsymbol{\omega})J_g(\boldsymbol{\omega})\}E\{J_g^c(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega})\} + |E\{J_g(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega})\}|^2 = f^2_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}) + 0(\Delta^{-2}N^{-1})$$

the residual term is uniform in $\boldsymbol{\omega}$. This leads to (42).

Then, we get

$$\operatorname{var}\left\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})\right\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f^2_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})\{W_\rho(\mathbf{s})\}^2 d\mathbf{s}\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|\Gamma(\boldsymbol{\theta})|^4 d\boldsymbol{\theta}\right\} + (\Delta^{-2}N^{-1})$$

Thus, we can write the variance of the periodogram in terms of the spectral density, $f_{\Delta,\mathbf{x}-\mathbf{s}}$, of the process on the lattice. Now, we prove that as $\Delta \to 0$, $\Delta n_i \to \infty$ and $n_i \to \infty$ for $i = 1, 2$, the function $f^2_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})$, converges to $f^2_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})$.

$$\begin{aligned}f^2_{\Delta,\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}) - f^2_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}) &= \left(\sum_{\mathbf{Q}\in\mathbb{Z}^2} f_{\mathbf{x}-\mathbf{s}}\left(\boldsymbol{\omega} + \frac{2\pi\mathbf{Q}}{\Delta}\right)\right)^2 - f^2_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})\\ &= \sum_{\mathbf{Q}_1\in\mathbb{Z}^{*2}} f_{\mathbf{x}-\mathbf{s}}\left(\boldsymbol{\omega} + \frac{2\pi\mathbf{Q}_1}{\Delta}\right)\sum_{\mathbf{Q}_2\in\mathbb{Z}^{*2}} f_{\mathbf{x}-\mathbf{s}}\left(\boldsymbol{\omega} + \frac{2\pi\mathbf{Q}_2}{\Delta}\right)\end{aligned} \tag{44}$$

By A.1, for Δ small enough the previous expression becomes

$$\sum_{\mathbf{Q}_1\in\mathbb{Z}^{*2}} \left\|\boldsymbol{\omega}+\frac{2\pi\mathbf{Q}_1}{\Delta}\right\|^{-\tau_{\mathbf{x}-\mathbf{s}}} \sum_{\mathbf{Q}_2\in\mathbb{Z}^{*2}} \left\|\boldsymbol{\omega}+\frac{2\pi\mathbf{Q}_2}{\Delta}\right\|^{-\tau_{\mathbf{x}-\mathbf{s}}} \tag{45}$$

If $\tau_{\mathbf{x}} > 2$ for all $\mathbf{x}$ (assumption A.1) the series (45) converges. We obtain

$$\sum_{\mathbf{Q}_1\in\mathbb{Z}^{*2}} \left\|\boldsymbol{\omega}+\frac{2\pi\mathbf{Q}_1}{\Delta}\right\|^{-\tau_{\mathbf{x}-\mathbf{s}}} \sum_{\mathbf{Q}_2\in\mathbb{Z}^{*2}} \left\|\boldsymbol{\omega}+\frac{2\pi\mathbf{Q}_2}{\Delta}\right\|^{-\tau_{\mathbf{x}-\mathbf{s}}} < \sum_{\mathbf{Q}_1\in\mathbb{Z}^{*2}} \left\|\boldsymbol{\omega}+\frac{2\pi\mathbf{Q}_1}{\Delta}\right\|^{-2} \sum_{\mathbf{Q}_2\in\mathbb{Z}^{*2}} \left\|\boldsymbol{\omega}+\frac{2\pi\mathbf{Q}_2}{\Delta}\right\|^{-2} < \Delta^4 K_1$$

where K_1 does not depend on $\boldsymbol{\omega}$ or $\mathbf{x}-\mathbf{s}$. Thus,

$$\operatorname{var}\left\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})\right\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f^2_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})\{W_\rho(\mathbf{s})\}^2 d\mathbf{s}\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|\Gamma(\boldsymbol{\theta})|^4 d\boldsymbol{\theta}\right\} + O(B_g/B_Z) + O(\Delta^{-2}N^{-1}) + O(\Delta^4). \tag{46}$$

Therefore, by (46) and condition (C.5), the variance of $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ is asymptotically:

$$(C/\rho^2)\tilde{f}^2{}_{\mathbf{x}}(\boldsymbol{\omega})\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|\Gamma(\boldsymbol{\theta})|^4 d\boldsymbol{\theta}\right\}$$

Proof of part (iii):

The covariance of $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ can be written in terms of the covariance for $|J_{\mathbf{x}}(\boldsymbol{\omega})|^2$,

$$\operatorname{cov}\left\{\hat{f}_{\mathbf{x}}(\boldsymbol{\omega}), \hat{f}_{\mathbf{x}'}(\boldsymbol{\omega}')\right\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_\rho(\mathbf{s})W_\rho(\mathbf{y})\operatorname{cov}\left\{|J_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})|^2, |J_{\mathbf{x}'-\mathbf{y}}(\boldsymbol{\omega}')|^2\right\} d\mathbf{s}d\mathbf{y} \tag{47}$$

and we have

$$\operatorname{cov}\left\{|J_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega})|^2, |J_{\mathbf{x}-\mathbf{s}}(\boldsymbol{\omega}')|^2\right\} = \operatorname{cov}\{I_g(\boldsymbol{\omega}), I_g(\boldsymbol{\omega}')\}$$

by (38) and Theorem 5.2.4 in Brillinger (1981, pag. 125), we get

$$\operatorname{cov}\{I_g(\boldsymbol{\omega}), I_g(\boldsymbol{\omega}')\} = E\{J_g(\boldsymbol{\omega})J_g(\boldsymbol{\omega}')\}E\{J_g^c(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega}')\} + |E\{J_g(\boldsymbol{\omega})J_g^c(\boldsymbol{\omega}')\}|^2 = 0(\Delta^{-2}N^{-1}). \tag{48}$$

Therefore, by expression (48) and condition (A.4), if $\|\mathbf{x}-\mathbf{x}'\| \gg$ than the "width" of the function $W_\rho(\mathbf{u})$ the the covariance between $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ and $\hat{f}_{\mathbf{x}'}(\boldsymbol{\omega})$ would be approximately zero.

The covariance between $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ and $\hat{f}_{\mathbf{x}'}(\boldsymbol{\omega}')$ as a function of the frequency $(\boldsymbol{\omega})$ is proportional to (Priestley, 1966):

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\Gamma(\boldsymbol{\theta}+\boldsymbol{\omega})|^2|\Gamma(\boldsymbol{\theta}+\boldsymbol{\omega}')|^2 d\boldsymbol{\omega} \tag{49}$$

Thus, the covariance between $\hat{f}_{\mathbf{x}}(\boldsymbol{\omega})$ and $\hat{f}_{\mathbf{x}'}(\boldsymbol{\omega}')$ would be also approximately zero if $\|\boldsymbol{\omega}\pm\boldsymbol{\omega}'\| \gg$ bandwidth of $|\Gamma(\boldsymbol{\theta})|^2$, because then, expression (49) would be effectively zero.

References

Brillinger, D. R. (1970).The frequency analysis of relations between stationary spatial series, *Proceedings of the Twelfth Biennial Seminar of the Canadian Mathematical Congress,* (ed. R. Pyke), Montreal, Canadian Math. Congress, 39-81.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* Expanded edition. Holden-Day, Inc, San Francisco.

Constantine, A. G, and Hall, P. (1994). Characterizing surface smoothness via estimation of effective fractal dimension. *Journal of the Royal Statistical Society. Series B,* **56**, 97-113.

Cramér, H. (1960). On some classes of non-stationary processes. *Proc. Fourth Berkeley Symposium Math. Statist. and Prob.,* **2**, 57-78. University of California Press.

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

Cressie, N. (1993), *Statistics for Spatial Data.* John Wiley, New York.

Dahlhaus, R. (1996a). On the Kullback-Leibler information divergence of locally stationary processes. *Stoch. Proc. Appl.* **62**, 139-168.

Dahlhaus, R. (1996b). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In: *Athens Conference on Applied Probability and Time Series Vol II.* (P.M. Robinson and M. Rosenblatt, Eds.), 145-159, Lecture Notes in Statistics 115, Springer, New York.

Dahlhaus, R. and Sahm, M. (2001). Local likelihood methods for nonstationary time series and random fields. Technical report, Universität Heidelberg.

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

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

Fuentes, M. (2002b). Modeling and prediction of nonstationary spatial processes. *Statistical Modeling: An International Journal,* invited paper, to appear December 2002.

Fuentes, M. and Raftery, A. (2001). Model Validation and Spatial interpolation by combining observations with outputs from numerical models via Bayesian melding. Technical Report no. 403, Dept. of

Statistics, University of Washington (under review by JASA).

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, (under review by JASA).

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.

Guttorp, P. and Sampson, P. (1994), Methods for estimating heterogeneous spatial covariance functions with environmental applications. In *Handbook of Statistics 12*, eds. G.P. Patil and C.R. Rao, Elsevier Science B.V., 661–689.

Haas, T.C. (1995), Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *J. Amer. Statist. Assoc.* **90**, 1189–1199.

Higdon, D., Swall, J. and Kern, J. (1999). Non-stationary spatial modeling. In *Bayesian Statistics 6*, eds. J.M. Bernardo *et al.*, Oxford University Press, pp. 761–768.

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

Nychka, D., Wikle, C. and Royle, J.A. (1999), Large spatial prediction problems and nonstationary random fields. Preprint, Geophysical Statistical Program, National Center for Atmospheric Research.

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

Priestley, M. B. (1965). Evolutionary spectral and non-stationary processes. *Journal of the Royal Statistical Society. Series B.* **27**, 204-237.

Priestley, M. B. (1966). Design relations for non-stationary processes. *Journal of the Royal Statistical Society. Series B.* **28**, 228-240.

Priestley, M. B. and Rao, T. S. (1969). A test for non-stationarity of time-series. *Journal of the Royal Statistical Society. Series B.* **31**, 140-149.

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

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. *J. Amer. Statist. Assoc.* **87**, 108-119.

Stein, S. K. (1967). *Calculus in the first three dimensions.* New York: McGraw-Hill, Inc.

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.* Springer-Verlag, New York.

Ver Hoef, J.M. and R.D. Barry. (1998). Modeling crossvariograms for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference,* **69**, 275-294.

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

Yaglom, A. M. (1987). *Correlation theory of stationary and related random functions.* Springer-Verlag, New York.

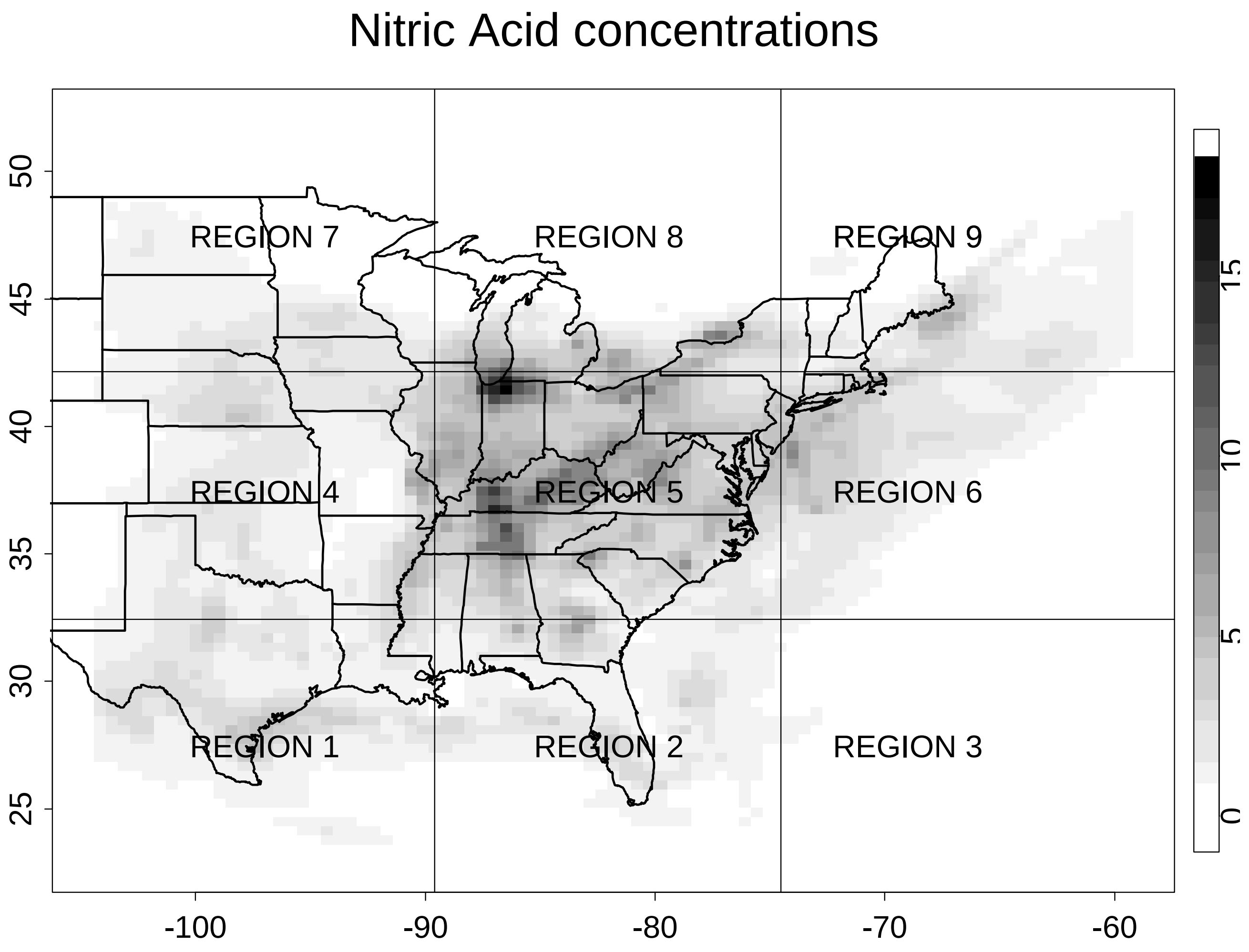

Figure 1: Output of EPA Models-3, showing the hourly average concentrations of nitric acid (ppb) for the week starting July 11, 1995. The resolution is 36 km^2.

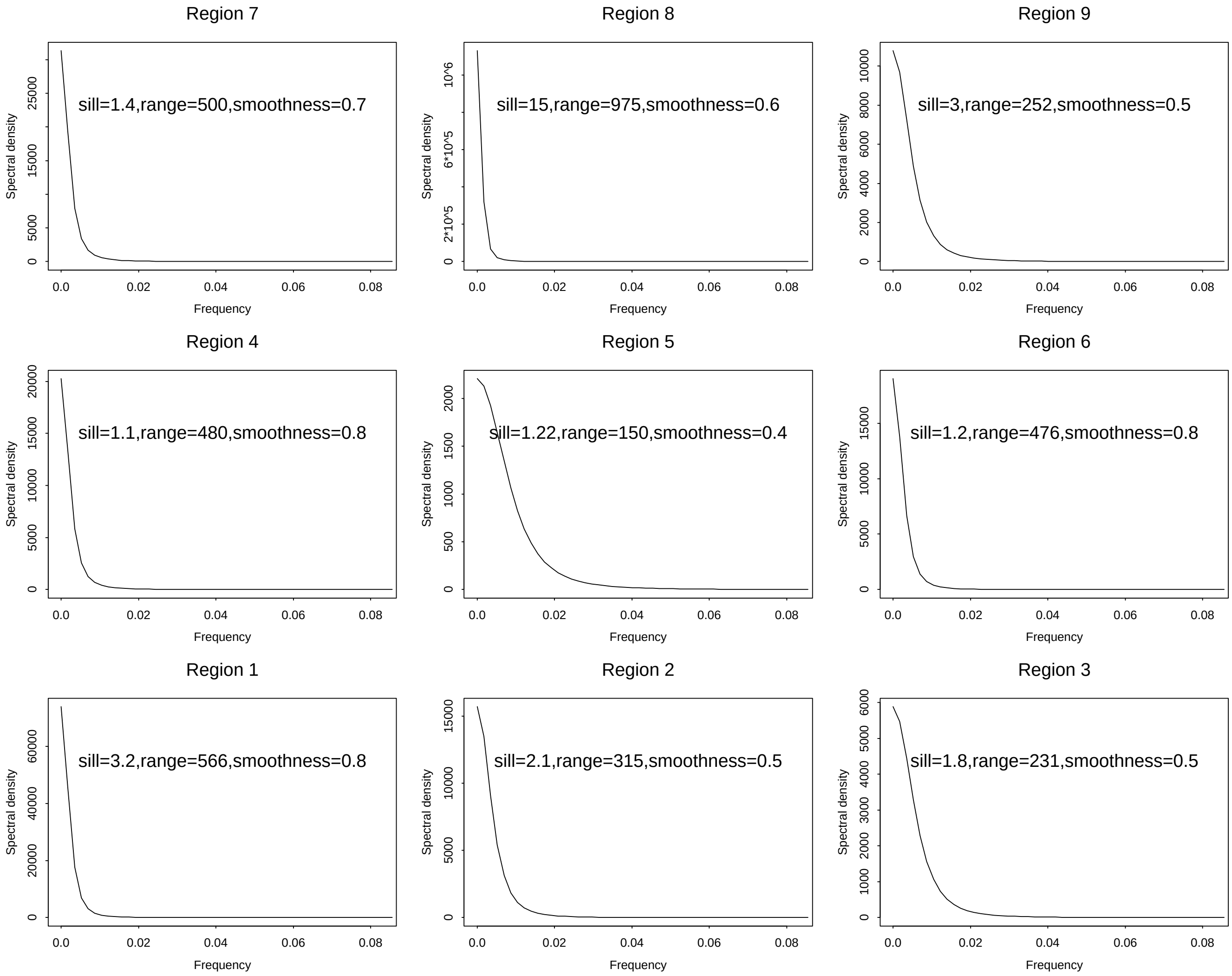

Figure 2: This figure shows the estimated Matérn spectral densities $\{f_{\mathbf{x_i}}\}$ of the nonstationary process Z at 9 locations, $\mathbf{x}_1, \ldots, \mathbf{x}_9$, which are the centroids of the 9 equally-dimensioned subregions $S_1, \ldots, S_9$, shown in Figure 1, e.g. location 1 ($\mathbf{x_1}$) corresponds to the centroid of the subregion in the lower left (TX) of Figure 1. We estimated the parameters of the 9 spectral density Matérn models by fitting parametric models to the empirical estimates $\{\hat{f}_{\mathbf{x_i}}\}$ in the log scale, we used a non linear least squares approach. Each spectral density represents the local spatial structure of the process Z at a different location. The sill is in km, and Δ (the spacing between observed values) is 36 km.