STRUCTURAL BREAK DETECTION FOR GEOSTATISTICAL DATA

by

Weiyu Zhou
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Statistical Science

Committee:

______________________	Dr. Pramita Bagchi, Dissertation Director
______________________	Dr. Anand Vidyashankar, Committee Member
______________________	Dr. David Straus, Committee Member
______________________	Dr. Seiyon Lee, Committee Member
______________________	Dr. Jiayang Sun, Department Chair
Date: ________________	Summer Semester 2022 George Mason University Fairfax, VA

Structural Break Detection for Geostatistical Data

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Weiyu Zhou
Bachelor of Economics
Nankai University, 2016

Director: Dr. Pramita Bagchi, Professor
Department of Statistics

Summer Semester 2022
George Mason University
Fairfax, VA

Copyright © 2022 by Weiyu Zhou
All Rights Reserved

Dedication

I dedicate this dissertation to my families and friends.

Acknowledgments

First of all, I want to express my gratitude to Professor Bagchi, who introduces me the problem of structure break detection for spatial data, and has been providing me all kinds of supports and guidance during the past three years. This thesis would not be possible without her. I also want to thank my committee members: Professor Vidyashankar, Professor Straus, for their lectures and guidance; Professor Lee, for his kind words and suggestions. In addition, I want to thank Professor Qiao, Professor Slawski, Professor Bruce and Professor Diao for their lectures and helps. I would also like to thank Professor Rosenberger, who let me know the exciting news that I was admitted to GMU statistics department. Thanks to Professor Lowder and Professor Xue for the research opportunities they provide.

There are many other people I appreciate. Most importantly, my families, friends and school mates. Life will be so hard without you. I wish everyone a good health and a bright future.

Table of Contents

List of Tables

List of Figures

Abstract

STRUCTURAL BREAK DETECTION FOR GEOSTATISTICAL DATA

Weiyu Zhou, PhD

George Mason University, 2022

Dissertation Director: Dr. Pramita Bagchi

This thesis proposes a method for investigating structural breaks in a non-stationary spatial random field and provide a piecewise stationary approximation that best describes the process.

Suppose a random field is observed only once over a regular grid. We study the covariance structure of this field in the frequency domain. In the first part the thesis, we define a spectral difference statistic, a spatially varying quantity showing the difference between local spatial spectral density integrated over a range of frequencies. This spectral difference process is expected to be uniformly close to zero if and only if the underlying field is stationary. This intuitive behavior is justified by a rigorous derivation of the asymptotic behavior of this process under an increasing domain asymptotic scheme. This result is then utilized to construct a consistent and asymptotic level alpha test for the hypothesis of stationarity using the maximum of the spectral difference process over locations and range of frequencies. A field plot of this spectral difference process, called discrepancy map, further provides insight to the nature of nonstationarity presents in the observed field. Next, we propose a method to construct a piece-wise approximation of the observed random field, where the pieces are spatial regions with linear boundaries by an iterative search. A hierarchical clustering algorithm is used to appropriately merge the initial partition to produce the final

approximation. A computationally efficient implementation of this methodology has been outlined. The accuracy and performance of the proposed methods are demonstrated via extensive simulations and two case studies on real data. The later part of the thesis outlines strategies of extending this methodology for a random field observed at irregularly spaced locations. The efficiency of this extension is investigated by some numerical experiments.
rectangular regions is constructed by an iterative search. A hierarchical clustering algorithm is then used to determine the optimal number of clusters and appropriately merge the initial partition to produce the final approximation. A computationally efficient implementation of this methodology has been outlined. The accuracy and performance of the proposed methods are demonstrated via extensive simulations and two case studies on real data.

Chapter 1: Introduction

In this thesis, we propose a pipeline for investigating, understanding, and modeling the nonstationarity of a spatial random field.

Geo-referenced or spatial data are prevalent in many fields, for example geology, ecology and econometrics (see, e.g., Anselin, 2001; Magnussen, 1993; Mets et al., 2017; Rollinson et al., 2021). These data usually exhibit spatial correlation, prohibiting the use of many traditional statistical methods as they assume independent observations. Understanding the spatial correlation in the data is necessary to develop statistical models, inference, and prediction methods for spatial data. In this thesis, we focus on the geostatistical process, a spatial process defined on a continuous geographical region and usually modeled as a random field. In most practical situations, our data constitute one partial observation of a single random field instead of multiple copies of the underlying object of interest.

Modeling dependence in spatial data is often challenging and differs significantly from other classical dependence models, such as clustered observations or time-series data. Modeling the dependence between the observed spatial data using a general covariance matrix is practically impossible due to the curse of dimension and difficulty of implementation due to high memory and computational cost. However, a geostatistical process often exhibits a more systematic correlation structure, with observations close-by exhibiting higher dependence than observations further apart. Although this dependence structure is similar to time series data, the time has a clear definition of direction, which spatial data lacks. Due to this, it is often tricky to mathematically define a valid and realistic covariance structure for such data.

The most common strategy to deal with such inconvenience is to impose simplifying assumptions on the second-order structure. Stationarity is one of the most popular simplifying assumptions for modeling spatial data, which assumes the joint second-order behavior only depends on the relative position of two locations. The assumption of stationarity allows us to define an auto-covariance function, a function of the spatial lag, instead of a full covariance between every pair of locations. However, stationarity is not a realistic assumption in many practical situations, and there has been a growing recognition of the need for nonstationary spatial covariance functions in various disciplines. The atmospheric and environmental sciences are two prominent areas of application; for example, Holland et al. (2003) reviews the use of nonstationary spatial covariance functions for the statistical analysis of air quality data. C. Paciorek and Schervish (2003) discusses the use of nonstationary models for Gaussian process regression in machine learning applications.

In practice, validation of stationarity is essential in choosing the appropriate model for spatial data. If the underlying model is, in fact, stationary, models and inference methods designed explicitly for a stationary process are computationally efficient and often more powerful. On the other hand, a misspecified stationarity assumption can lead to loss of power and wrong inference. There are two effective ways to check stationarity: visualization and statistical testing. Early tests for spatial stationarity are developed in the context of linear array data, such as Bose and Steinhardt (1996) and Ephraty et al. (1996). For data observed on a regular grid, Fuentes, 2005 proposes a test that partitions the space into multiple separate regions and compares the spectral density in each region. For irregular spaced data, two special tests are Jun and Genton, 2012, and Bandyopadhyay and Rao, 2017. The former compares the estimated covariance of two separate regions at various lags, while the latter works in the frequency domain.

In Chapter 3 of this thesis, we propose a statistical test for stationarity of a spatial random field. We use a frequency domain approach and define statistics that can be regarded as the gradient of the spectral density. We use the maximum of these statistics over space and frequency range as our test statistic. We derive a rigorous asymptotic distribution of this proposed statistic under an increasing domain set-up. We used that distribution to propose a consistent test that asymptotically preserves the specified level and develop a computationally efficient algorithm to implement this test. The performance of the proposed test has been investigated with extensive numerical simulations and two case studies with actual data.

When the assumption of stationarity is likely violated, one could fit the data with nonstationary models, and there is a need for interpretability and computationally tractability. C. Paciorek and Schervish (2003) describes a method for producing explicit expressions for mathematically valid spatial covariance functions with locally varying geometric anisotropies. However, this approach does not allow the other aspects of the covariance to vary spatially. Higdon (1998) and Fuentes (2002) described methods for generating nonstationary covariance functions as integrals. Nychka et al. (2002) proposed a wavelet approach for producing nonstationary spatial covariance functions. Sampson and Guttorp (1992) proposed explicit covariance functions with locally varying geometric anisotropies using spatial deformations. Stein (1995) extended this idea to produce covariance functions that allow both the local geometric anisotropy and the degree of differentiability to vary spatially.

The nature of nonstationarity guides the choice of the nonstationary covariance model and inference method in the data. It is thus essential to investigate the dynamic nature of the second-order structure. We propose a visualization of the gradient of spectral density statistic over the space, called a disparity map. The disparity map shows the spatial dynamics of the spectral density of the process and provides insight into the nature of nonstationarity in the data. This visualization can be an essential tool in guiding the choice of covariance model and subsequent spatial data analysis.

Further, we develop an algorithm to partition a spatial random field into rectangular regions. This partitioning algorithm provides the best piecewise stationary approximation for the given random field. The work in Fuglstad et al., 2015 can justify the usefulness of our partition method, which shows that a piecewise stationary model consisting of two stationary regions may perform as well as nonstationary models while being less computationally demanding. Although piecewise stationary approximation has been studied in time series, such as Adak, 1998, it is still an open problem under active research for the spatial random field. Moreover, while a field can be roughly stationary over some regions, it may contain some more suitable regions to be modeled as nonstationary.

Existing partition methods include those based on trees (e.g., Gramacy and Lee, 2008), Voronoi tessellation (e.g., Kim et al., 2005), clustering (e.g., Morris, 2021), or closed curves (e.g., Masotti et al., 2021). All the existing methods use a parametric and Bayesian approach, which is sensitive to the choice of model and prior distribution. Among the existing methods, the tree-based methods usually generate partition boundaries parallel to coordinate axes, which can be too restrictive sometimes but are computationally efficient. Voronoi tessellation is more flexible by partitioning a field into convex polygons, but it may over-partition if some underlying regions are non-convex. To this end, Pope et al., 2021 suggests merging the regions from the initial partition so that the final partition can include non-convex stationary regions, therefore this is more flexible than the Voronoi tessellation. In Chapter 4, we introduce the spatial partition procedure, which makes use of the spectral difference statistics defined in in Chapter 3 to create the initial partition. After the field is partitioned, we then check if there are any connected sub-fields that are similar in spectral density and can be merged together. To decide the number of partitions to retain, we define the dendrogram, the scree plot, and the gap statistics for our case. We then try to expand our methods to irregular spaced data in Chapter 5. Finally, we present two case studies in Chapter 6 to show how to apply our methods in practice.

Chapter 2: Spectral Analyses of Spatial Random Fields

A spatial random field is a stochastic process $Z(\mathbf{s})$ indexed by $s \in S$, where $S \subset \mathbb{R}^d$ is a continuous spatial region. In this thesis we focus on the case where $d = 2$, but the discussion and the methodology described can be extended to $d > 2$ in a straightforward manner.

A spatial random field is usually characterized using the mean

$$m(\mathbf{s}) = \mathbf{E}(Z(\mathbf{s})), \quad \mathbf{s} \in S$$

and the covariance function

$$C(\mathbf{s}_1, \mathbf{s}_2) = \mathrm{Cov}(Z(\mathbf{s}_1), Z(\mathbf{s}_2)), \quad \mathbf{s}_1, \mathbf{s}_2 \in S.$$

where the covariance function satisfies

1. Non-negative definiteness: $\forall m \geq 1, a \in \mathbb{R}^m$ and $\{\mathbf{s}_1, ..., \mathbf{s}_m\} \subseteq S$,

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j C(\mathbf{s}_i, \mathbf{s}_j) \geq 0.$$

2. Symmetry: $C(\mathbf{s}_1, \mathbf{s}_2) = C(\mathbf{s}_2, \mathbf{s}_1)$, $\forall \mathbf{s}_1, \mathbf{s}_2 \in S$.

The usual statistical problems related to spatial random fields involve estimating these infinite dimensional quantities based on a single sample of partial observation of the random field. Therefore it is important to impose additional structure to the underlying process in order to do any meaningful inference.

2.1 Stationarity

Stationarity is the most popular simplifying structural assumption for stochastic processes. Stationarity is a property of self-replication of a stochastic process. It implies the lack of importance of absolute coordinates. There are three types of stationarity:

- **Strong stationarity**: A random field is *strict* stationary if the spatial distribution only depends on the *relative position* of the locations, i.e., it is invariant under arbitrary translations of the locations by a vector $\mathbf{h}$, that is,

  $$\mathbb{P}(Z(\mathbf{s}_1) \leq z_1, \cdots, Z(\mathbf{s}_k) \leq z_k) = \mathbb{P}(Z(\mathbf{s}_1 + \mathbf{h}) \leq z_1, \cdots, Z(\mathbf{s}_k + \mathbf{h}) \leq z_k)$$

  for all integers k and all $\mathbf{s}_1, \ldots, \mathbf{s}_k \in S$.

- **Weak stationarity**: Z is a *second-order or weak stationary random field* on S if it has constant mean and covariance between two sites $C(\mathbf{s}_1, \mathbf{s}_2)$ depends on their relative position $\mathbf{s}_1 - \mathbf{s}_2$. Mathematically, we can write

  $$m(\mathbf{s}) = m \text{ and } C(\mathbf{s}_1, \mathbf{s}_2) = C(\mathbf{s}_1 - \mathbf{s}_2) = C(\mathbf{h}), \quad \forall \mathbf{s}_1, \mathbf{s}_2 \in S$$

  where $C : S \to \mathbb{R}$ is the stationary auto-covariance function of Z. The autocorrelation function $\rho(\mathbf{h}) = C(\mathbf{h})/C(\mathbf{0})$ of Z is a function of $\mathbf{h}$.

  Note that the assumption $m(\mathbf{s}) = m$ is too strong, but it can be weakened relatively easily. For example, we can model the mean as a function of the location and consider a weak stationary covariance structure of the demeaned spatial process.

  If Z is strictly stationary and if $Z \in L^2$, i.e., the random functions $Z(s)$'s are square-integrable (this means that the variance and covariance function exist as finite quantities) then Z is second-order stationary. The converse is generally not true but both notions represent the same thing if Z is a Gaussian process.

- **Intrinsic stationary**: Z is an intrinsically stationary process (or intrinsic process) if the increments of the process are weak stationary, i.e., for each $\mathbf{h} \in S$, the process

$\Delta Z^{(\mathbf{h})} = \{\Delta Z^{(\mathbf{h})}(\mathbf{s}) = Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s}) : s \in S\}$ is weak stationary. It is charecterized by its linear drift

$$m(\mathbf{s}_1 - \mathbf{s}_2) = \mathbb{E}\left[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\right]$$

and its variogram

$$2\gamma(\mathbf{s}_1 - \mathbf{s}_2) = \mathrm{Var}(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)).$$

The function $\gamma : S \to \mathbb{R}$ is called the semivariogram. In fact one can show, for such a process $m(\mathbf{h})$ is linear in $\mathbf{h}$, i.e., $m(\mathbf{h}) = a^T\mathbf{h}$ for some vector $a \in \mathbb{R}^d$.

For weak stationary process, $\gamma(\mathbf{h}) = C(\mathbf{0}) - \mathbf{C}(\mathbf{h})$.

2.2 Spectral Density of Stationary Spatial Random Fields

Harmonic analysis of a spatial process 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.

Consider a weak stationary random field $Z(\mathbf{s}), \mathbf{s} \in S$ with zero mean and autocovariance function C_0. Following Fuentes (2002), the spectral density of a stationary random field is defined as:

$$f_0(\lambda) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} C_0(\mathbf{h}) \exp(-i\mathbf{h}^T\lambda) d\mathbf{h}, \qquad \lambda \in \mathbb{R}^2. \tag{2.1}$$

The spectral density $\{f_0(\lambda), \lambda \in \mathbb{R}^2\}$ provides a complete characterization of the second-order structure of the spatial random field and gives the decomposition of the total variation of the process across different frequencies.

We additionally assume that the random field satisfies the following weak-dependence assumption,

$$\int_{\mathbb{R}^2} \Big| C_0(\mathbf{h}) \Big| d\mathbf{h} < \infty. \tag{2.2}$$

Under the weak dependence assumption (2.2), the integral in (2.1) can be shown to be absolutely convergent and hence the quantity $f_0(\lambda)$ is well defined. Moreover the spectral density $f_0(\lambda) = f_0(\lambda_1, \lambda_2)$, where $\lambda_1, \lambda_2 \in \mathbb{R}$, is uniformly continuous with respect to both λ_1 and λ_2 (Fuentes (2002)). Furthermore the following inversion formula holds:

$$\int_{\mathbb{R}^2} f_0(\lambda) \exp(i\mathbf{h}^T \lambda) d\lambda = C_0(\mathbf{h}), \qquad \forall \mathbf{h} \in \mathbb{R}^2. \tag{2.3}$$

2.3 Random Field Observed on Regular Grid

The decomposition of the autocovariance function into harmonic oscillations cannot be uniquely determined if the random field is partially observed. For example suppose we observe the random field over a regular grid with spacing Δ, it is not possible to distinguish an oscillation with a spatial frequency λ from all the oscillations with frequencies $\lambda + 2\pi z/\Delta$, where $z \in \mathbb{Z}^2$. The impossibility of distinguishing the harmonic components with frequencies differing by an integer multiple of $2\pi/\Delta$ by observations in the 2-dimensional integer lattice with spacing Δ is called the aliasing effect.

Then, if observation of a continuous process Z is carried out only at uniformly spaced spatial locations Δ units apart, the spectrum is concentrated within the finite frequency 2-dimensional interval $[-\pi/\Delta_1, \pi/\Delta_1] \times [-\pi/\Delta_2, \pi/\Delta_2]$, where $\Delta = (\Delta_1, \Delta_2)$ with Δ_1 and Δ_2 being the horizontal and vertical distance of the grid points respectively. Every frequency not in this interval has an alias in the interval, which is termed its principal alias. Then, the power distribution within each of the intervals distinct from the principal interval is superimposed on the power distribution within the principal interval. 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* $(\pi/\Delta_1, \pi/\Delta_2)$ must necessarily be so high that still higher frequencies make only a negligible contribution to the total power of the process. This means that we need to observe a dense sample of Z (small Δ). Thus when a second-order stationary random field Z is observed on a regular grid such that the observations are Δ units apart, the spectral density f_Δ of the discrete process can be written in terms of the spectral density f_0 of the continuous process as:

$$f_\Delta(\lambda_1, \lambda_2) = \sum_{p\in\mathbb{Z}} \sum_{q\in\mathbb{Z}} f_0 \left(\lambda_1 + \frac{2\pi p}{\Delta_1}, \lambda_2 + \frac{2\pi q}{\Delta_2} \right), \tag{2.4}$$

for $(\lambda_1, \lambda_2) \in [-\pi/\Delta_1, \pi/\Delta_1] \times [-\pi/\Delta_2, \pi/\Delta_2]$.

2.4 Random Field Observed at Irregularly Spaced Locations

When a random field is observed over irregularly spaced locations, the inference becomes much more challenging. One major obstacle is the increasing computational cost in this setting. Although Matsuda and Yajima (2009) argues that using a spectral domain approach can reduce computational cost for analysis of irregularly spaced spatial data, even the spectral analysis is computationally costly as we cannot use fast Fourier transformation. Foundational work for studying spectral density for such random field has been developed by Matsuda and Yajima (2009) and Rao (2018). Both these works assume that the irregular locations are independent, identically distributed random variables which allows it to be extremely irregular.

2.5 Estimation of Spectral Density

The basic estimator of spatial spectral density is the periodogram, a nonparametric estimate of the spectral density. Use and properties of spatial periodograms for stationary processes have been investigated by Whittle (1954), Ripley (2005), Guyon (1982, 1995), Rosenblatt (2012), Stein (1995, 2015) and Fuentes (2002, 2005, 2007), among others.

2.5.1 Regular Grid

Consider a spatial weak stationary process Z with an auto-covariance function C. We observe the process at $m \times n$ equally spaced locations in a two-dimensional regular grid G. For notational convenience, without loss of generality, assume that the vector distance between neighboring observations is (Δ, Δ), i.e., the horizontal and vertical distance are the same. In addition, we assume the origin of the grid is $(0, 0)$. The periodogram is a nonparametric estimate of the spectral density, which is the Fourier transform of the covariance function. We define the discrete Fourier transformation of the data is defined as:

$$\tilde{Z}_{m,n,\Delta}(\lambda_1, \lambda_2) = \frac{\Delta}{2\pi\sqrt{mn}} \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} Z(\Delta p, \Delta q) \exp(-i\Delta(p\lambda_1 + q\lambda_2)) \tag{2.5}$$

The periodogram is then defined as:

$$\begin{aligned} I_{m,n,\Delta}(\lambda_1, \lambda_2) &= \left| \tilde{Z}_{m,n,\Delta}(\lambda_1, \lambda_2) \right|^2 \\ &= \frac{\Delta^2}{(2\pi)^2 mn} \left| \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} Z(\Delta p, \Delta q) \exp(-i\Delta(p\lambda_1 + q\lambda_2)) \right|^2, \end{aligned} \tag{2.6}$$

where $(\lambda_1, \lambda_2) \in \Lambda_{m,n,\Delta}$. $\Lambda_{m,n,\Delta} = \left\{ \left(\frac{2\pi f_1}{\Delta m}, \frac{2\pi f_2}{\Delta n} \right) \right\}$ is the set of frequencies we choose to examine, where:

$$(f_1, f_2) = \left\{ \lfloor -\frac{m-1}{2} \rfloor, \ldots, m - \lfloor \frac{m}{2} \rfloor \right\} \times \left\{ \lfloor -\frac{n-1}{2} \rfloor, \ldots, n - \lfloor \frac{n}{2} \rfloor \right\}. \tag{2.7}$$

Theoretical Properties of Periodogram

The expected value of the periodogram at frequency $\lambda = (\lambda_1, \lambda_2)$ is given by (see Fuentes (2002))

$$\mathbf{E}\left(I_{m,n,\Delta}(\lambda)\right) = \frac{1}{mn(2\pi)^2} \int_{[-\pi/\Delta, \pi/\Delta]^2} f_\Delta(\omega) W_\Delta(\omega - \lambda) d\omega,$$

where

$$W_\Delta(\omega) = W_\Delta(\omega_1, \omega_2) = \frac{\sin^2(m\omega_1/2)}{\sin^2(\omega_1/2)} \times \frac{\sin^2(n\omega_2/2)}{\sin^2(\omega_2/2)}.$$

Due to the nature of the function W_Δ periodogram can have substantial bias in small sample, however for Gaussian random field, the bias is asymptotically negligible. In fact, Brillinger

(1981) showed that under increasing domain asymptotics, the periodogram is an asymptotically unbiased but inconsistent estimator of f_Δ process. Fuentes (2002) established the following similar but stronger result under a mixed asymptotic set-up.

Theorem 2.5.1. *Fuentes (2002) Consider a Gaussian stationary process Z with spectral density $f(\lambda)$ on a lattice G. We assume Z is observed at $m \times n$ equally spaced locations in G and the spacing between observations is Δ. We define the periodogram function, $I_{m,n,\Delta}(\lambda)$, as in* (2.6)*. Assume that*

(a) The rate of decay of $f(\lambda)$ at high frequencies is proportional to $\|\lambda\|^{-\tau}$ for $\tau > 2$.

(b) The autocovariance function satisfies $\int \|\mathbf{h}\| c(\mathbf{h}) d\mathbf{h} < \infty$.

(c) $\Delta \to 0$, $m, n \to \infty, m/n \to c$, for a constant $c > 0$, $\Delta m \to \infty$ and $\Delta n \to \infty$.

Then, we have

(1) The expected value of the periodogram, $I_{m,n,\Delta}(\lambda)$, is asymptotically $f(\lambda)$

(2) The asymptotic variance of the periodogram, $I_{m,n,\Delta}(\lambda)$, is $f^2(\lambda)$.

(3) The periodogram values $I_{m,n,\Delta}(\lambda_1)$ and $I_{m,n,\Delta}(\lambda_2)$, for $\lambda_1 \neq \lambda_2$, are asymptotically independent.

2.5.2 Irregular Grid

When the random field Z in Section 2.5.1 is observed at J irregular-spaced locations $\mathbf{s}_j = (s_j^1, s_j^2) \in [0, L_1] \times [0, L_2]$, following Rao (2018) we define the Fourier transform of the data as:

$$\tilde{Z}_{J,L_1,L_2}(\lambda_1, \lambda_2) = \frac{\sqrt{L_1 L_2}}{2\pi J} \sum_{j=1}^{J} Z(s_j^1, s_j^2) \exp(-i(s_j^1 \lambda_1 + s_j^2 \lambda_2)). \tag{2.8}$$

The periodogram is then defined as:

$$I_{J,L_1,L_2}(\lambda_1, \lambda_2) = \left| \tilde{Z}_{J,L_1,L_2}(\lambda_1, \lambda_2) \right|^2$$

$$= \frac{L_1 L_2}{(2\pi)^2 J^2} \left| \sum_{j=1}^{J} Z(s_j^1, s_j^2) \exp(-i(s_j^1 \lambda_1 + s_j^2 \lambda_2)) \right|^2, \tag{2.9}$$

where $(\lambda_1, \lambda_2) \in \Lambda_{L_1, L_2}$. $\Lambda_{L_1,L_2} = \left\{ \left(\frac{2\pi f}{L_1}, \frac{2\pi f}{L_2} \right) \right\}, f \in \mathbb{Z}$ is the set of frequencies we choose to examine. As argued in Rao (2018) the choice of these frequencies are optimal, and under suitable regularity condition the periodograms at these frequencies are asymptotically unbiased and uncorrelated among themselves. In fact their joint asymptotic behavior is very similar to the regular grid case given in Theorem 2.5.1

2.6 Non-stationary Spatial Random Field

If the random field is non-stationary, we can no longer define an auto-covariance function, and the second-order structure is described by a covariance kernel $C(\mathbf{s}_1, \mathbf{s}_2)$ for $\mathbf{s}_1, \mathbf{s}_2 \in S$. In this situation we cannot define a spectral density that does not depend on the location, and summarizes the second-order dynamics over the whole space. Second-order non-stationarity of random field is difficult to theoretically characterize and causes significant increase in computational complexity for inference purposes. Fuentes (2002) define the spectral density over $\mathbb{R}^4$ of a general non-stationary random field as the Fourier transformation of the covariance kernel. However a more interpretable quantity is a spatially varying spectral density, which provides a decomposition of the local variability across different spatial frequencies at each location. The spatially varying spectral density depends on the nature of non-stationarity in the underlying random field. There are two popular ways of characterizing covariance non-stationarity in spatial random field.

- **Piecewise Stationary:** The process can be partitioned into K stationary pieces.

  $$\mathbf{E}Z(\mathbf{s}) = \mu_i, \quad \mathrm{Var}Z(\mathbf{s}) < \infty,$$

  $$\mathrm{Cov}(Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})) = C_{ij}(\mathbf{h}), \text{ if } \mathbf{s} \in \Omega_i, \mathbf{s}+\mathbf{h} \in \Omega_j, i, j = 1, \ldots, K, \tag{2.10}$$

  where $\{\Omega_1, \Omega_2, \ldots, \Omega_K\}$ is a disjoint partition of S, the spatial region under study.

Sometimes uncorrelatedness across the different regions is imposed, i.e., $C_{ij}(\mathbf{h}) = C_i(\mathbf{h})\mathbf{1}_{ij}$.

Under this piecewise stationary set-up it is meaningful to talk about spectral density within each region, and cross-spectral density for each pair of regions. These quantities can be individually defined and estimated similar to the stationary speactral density.

- **Locally Stationary:** Following Kurisu (2022), the process $\{Z_{S_n}(s_1, s_2) : (s_1, s_2) \in \mathbb{R}^2\}$ is locally stationary if for each rescaled space point $(u_1, u_2) \in [0, 1]^2$ there exists an associated random field $\{Z_{u_1,u_2}(s_1, s_2) : (s_1, s_2) \in \mathbb{R}^2\}$ with the following properties:

  (i) $\{Z_{u_1,u_2}(s_1, s_2) : (s_1, s_2) \in \mathbb{R}^2\}$ is strictly stationary for each $(u_1, u_2) \in [0, 1]^2$.

  (ii) It holds that

$$\left| Z_{S_n}(s_1, s_2) - Z_{(u_1,u_2)}(s_1, s_2) \right| \leq k_1 \left\| \frac{(s_1, s_2)}{|S_n|} - (u_1, u_2) \right\|_2 + k_2 \frac{1}{|S_n|}$$

  for some positive constants k_1 and k_2. This is appropriate for modelling random field continuously evolving over the space.

  Under this set-up the spatially varying spectral density can be defined in terms of the spectral density of the approximating process $Z_{(u_1,u_2)}$. In particular, let $\mathfrak{f}_{(u_1,u_2)}$ be the spectral density of the stationary process $Z_{(u_1,u_2)}$, then the spectral density of the original spatial random field is defined as

$$f_{(s_1,s_2)}(\lambda_1, \lambda_2) = \mathfrak{f}_{(s_1,s_2)/|S_n|}(\lambda_1, \lambda_2).$$

Chapter 3: Test for Stationarity

In this chapter we consider a spatial random field Z observed over a $M \times N$ grid G originated at $(0,0)$, where the horizontal and the vertical spacing between the adjacent observations is Δ. We want to test if the underlying random field Z is stationary, in particular we are interested in testing the following hypothesis:

H_0: The random field Z is second-order stationary; vs.

H_1: The random field Z is not second-order stationary.

Note that in this case, the nonstationarity in mean and covariance are not separable. We assume the random field has a zero mean, so the field is not stationary if the covariance structure of Z varies in space. This assumption is appropriate, for example, when we are dealing with the residual field from a spatial regression model. We will construct a test based on the maximal variation in spectral density, when comparing all adjacent half open half closed rectangular region of size $m\Delta \times n\Delta$ in different directions.

In the sequel, we introduce the notation Z_W as the random field Z restricted in the region W and f_{Z_W} as the spectral density of the random field Z_W. Moreover we expand the notations introduced in Section 2.5.1. Let $W \cap G$ be a $m \times n$ grid with origin (o_1, o_2) and spacing Δ. We write the periodogram (2.6) calculated using the observations from W as $I_{o_1,o_2,m,n,\Delta}(\lambda_1, \lambda_2)$:

$$I_{o_1,o_2,m,n,\Delta}(\lambda_1, \lambda_2) = \frac{\Delta^2}{(2\pi)^2 mn} \left| \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} Z(o_1 + \Delta p, o_2 + \Delta q) exp(-i\Delta(p\lambda_1 + q\lambda_2)) \right|^2 . \tag{3.1}$$

Let $\Lambda_{m,n,\Delta,\omega_1,\omega_2}$ to be the set of frequencies $\left(\frac{2\pi f_1}{\Delta m}, \frac{2\pi f_2}{\Delta n}\right)$, where

$$(f_1, f_2) = \{\lfloor \omega_1 \lfloor -\frac{m-1}{2} \rfloor \rfloor, \ldots, \lfloor \omega_1 (m - \lfloor \frac{m}{2} \rfloor) \rfloor\} \times \{\lfloor \omega_2 \lfloor -\frac{n-1}{2} \rfloor \rfloor, \ldots, \lfloor \omega_2 (n - \lfloor \frac{n}{2} \rfloor) \rfloor\}.$$

Then the periodogram $I_{m,n,\Delta}$ defined in (2.6) can be written as $I_{0,0,m,n,\Delta}$ and the set of frequencies $\Lambda_{m,n,\Delta}$ in Section 2.5.1 can be written as $\Lambda_{m,n,\Delta,1,1}$ under this expanded notations.

3.1 The Spectral Difference Statistics

Let the function $D^d_{m,n,\Delta,\omega_1,\omega_2}(s_1, s_2) : \mathbb{R}^2 \mapsto \mathbb{R}$ be the integrated difference between the spectral density of the two adjacent regions $W^{d_1}_{(s_1,s_2),m,n,\Delta}$ and $W^{d_2}_{(s_1,s_2),m,n,\Delta}$, where d denotes the direction of the comparison, (s_1, s_2) denotes the location of the comparison, Δ denotes the spacing so that $m\Delta$ and $n\Delta$ are the side length of the regions:

$$\begin{aligned} & D^d_{m,n,\Delta,\omega_1,\omega_2}(s_1, s_2) \\ &= \int_{-\omega_1\pi}^{\omega_1\pi} \int_{-\omega_2\pi}^{\omega_2\pi} \left(f_{Z_{W^{d_1}_{(s_1,s_2),m,n,\Delta}}}(\lambda_1, \lambda_2) - f_{Z_{W^{d_2}_{(s_1,s_2),m,n,\Delta}}}(\lambda_1, \lambda_2) \right) \mathrm{d}\lambda_1 \, \mathrm{d}\lambda_2. \end{aligned} \tag{3.2}$$

The frequency window for the integration is $[-\omega_1\pi, \omega_1\pi] \times [-\omega_2\pi, \omega_2\pi], \omega_1, \omega_2 \in [0, 1]$. If $\omega_1, \omega_2 = 1$, $D^d_{m,n,\Delta,1,1}$ can be used to study the spatial variation within the variance structure. To study the spatial variation of the covariance structure, we calculate $D^d_{m,n,\Delta,\omega_1,\omega_2}$ based on the comparisons along two orthogonal directions. For convenience, we let the two orthogonal directions to be parallel to the two axes of the grid G, and call them horizontal direction and vertical direction. Then, the regions for the horizontal comparison are defined as:

$$\begin{aligned} W^{h_1}_{(s_1,s_2),m,n,\Delta} &= [s_1 - m\Delta, s_1) \times [s_2, s_2 + n\Delta) \\ W^{h_2}_{(s_1,s_2),m,n,\Delta} &= [s_1, s_1 + m\Delta) \times [s_2, s_2 + n\Delta), \end{aligned} \tag{3.3}$$

for $(s_1, s_2) \in [m\Delta, (M-m)\Delta] \times [0, (N-n)\Delta]$. The windows for the vertical comparison are defined as

$$W^{v_1}_{(s_1,s_2),m,n,\Delta} = [s_1, s_1 + m\Delta) \times [s_2 - n\Delta, s_2)$$

$$W^{v_2}_{(s_1,s_2),m,n,\Delta} = [s_1, s_1 + m\Delta) \times [s_2, s_2 + n\Delta), \tag{3.4}$$

for $(s_1, s_2) \in [0, (M-m)\Delta] \times [n\Delta, (N-n)\Delta]$. Therefore, the spectral difference statistics can be defined as:

$$D^h_{m,n,\Delta,\omega_1,\omega_2}(s_1, s_2)$$

$$= \int_{-\omega_1\pi}^{\omega_1\pi} \int_{-\omega_2\pi}^{\omega_2\pi} \left(f_{Z_{W^{h_1}_{(s_1,s_2),m,n,\Delta}}}(\lambda_1, \lambda_2) - f_{Z_{W^{h_2}_{(s_1,s_2),m,n,\Delta}}}(\lambda_1, \lambda_2) \right) \mathrm{d}\lambda_1 \, \mathrm{d}\lambda_2 \tag{3.5}$$

$$D^v_{m,n,\Delta,\omega_1,\omega_2}(s_1, s_2)$$

$$= \int_{-\omega_1\pi}^{\omega_1\pi} \int_{-\omega_2\pi}^{\omega_2\pi} \left(f_{Z_{W^{v_1}_{(s_1,s_2),m,n,\Delta}}}(\lambda_1, \lambda_2) - f_{Z_{W^{v_2}_{(s_1,s_2),m,n,\Delta}}}(\lambda_1, \lambda_2) \right) \mathrm{d}\lambda_1 \, \mathrm{d}\lambda_2. \tag{3.6}$$

Note that if the spatial process is stationary, then these differences should be zero as a function of (s_1, s_2) for any choice of $(\omega_1, \omega_2) \in [0,1]^2$. Therefore we will use these $D^d_{m,n,\Delta,\omega_1,\omega_2}$ for $d \in \{h, v\}$ to understand the spatial differences in covariance structure.

In practice we consider a sample version $\hat{D}^d_{m,n,\Delta,\omega_1,\omega_2}$, where the spectral densities in (3.2) are replaced by the periodograms calculated with the observations from the corresponding regions, and the integration is replaced by appropriate averages. We define the horizontal spectral difference statistic to be:

$$\hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(s_1, s_2)$$

$$= \hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(g_1, g_2)$$

$$= \frac{1}{mn\Delta^2} \sum_{(\lambda_1,\lambda_2) \in \Lambda_{m,n,\Delta,\omega_1,\omega_2}} \left[I_{g_1 - m\Delta, g_2, m, n, \Delta}(\lambda_1, \lambda_2) - I_{g_1, g_2, m, n, \Delta}(\lambda_1, \lambda_2) \right], \tag{3.7}$$

where $(g_1, g_2) = (\lceil s_1 \rceil, \lceil s_2 \rceil)$, therefore $(g_1, g_2) \in \{m\Delta, \ldots, (M-m)\Delta\} \times \{0, \ldots, (N-n)\Delta\}$. Similarly the vertical spectral difference statistic is defined as:

$$\begin{aligned}
&\hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(s_1, s_2) \\
=&\hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(g_1, g_2) \\
=&\frac{1}{mn\Delta^2} \sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}} \left[I_{g_1,g_2-n\Delta,m,n,\Delta}(\lambda_1, \lambda_2) - I_{g_1,g_2,m,n,\Delta}(\lambda_1, \lambda_2) \right],
\end{aligned} \tag{3.8}$$

with $(g_1, g_2) = (\lceil s_1 \rceil, \lceil s_2 \rceil) \in \{0, \ldots, (M-m)\Delta\} \times \{n\Delta, \ldots, (N-n)\Delta\}$.

If the underlying random field is second-order stationary we should expect these statistics to be small across the grid for all choices of direction d and frequency ranges determined by ω_1 and ω_2, and should capture the spatial dynamics of the second-order structure if the stationarity assumption is violated.

3.2 Asymptotic Properties of The Spectral Difference Statistics

To develop statistical inference based on this spectral differences, we investigate the asymptotic properties of this process in this section.

Theorem 3.2.1. *Let $Z(s)$ be a Gaussian random field indexed by $s \in S$ where S is a compact subset of $\mathbb{R}^2$ and we observe Z on a a regular $M \times N$ grid G with neighboring grid points separated by Δ units in both horizontal and vertical directions. Assume that Z is weak stationary with zero mean and auto-covariance function C. Under the assumptions*

(i) The auto-covariance function satisfies $\int \|\mathbf{h}\| |c(\mathbf{h})| \, d\mathbf{h} < \infty$.

(ii) The fourth moment of the random field is uniformly bounded by an integrable function, i.e., $\mathbf{E}(Z^4(s_1, s_2)) \leq K(s_1, s_2)$ for all $(s_1, s_2) \in G$ such that $\int \int K(s_1, s_2) ds_1 ds_2 < \infty$

(iii) $m, n \to \infty$, and $\frac{m}{M} \to c_1$, $\frac{n}{N} \to c_2$ for $c_1, c_2 > 0$.

we have

$$\left\{\sqrt{mn}\Delta\left[\hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(s_1,s_2),\hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(s_1,s_2)\right]\right\} \xrightarrow{d} \left\{\mathbb{D}^1_{\omega_1,\omega_2}(s_1,s_2),\mathbb{D}^2_{\omega_1,\omega_2}(s_1,s_2)\right\},$$

uniformly over $(\omega_1,\omega_2,s_1,s_2) \in [0,1]^2 \times S$*, where* $\left\{\mathbb{D}^1_{\omega_1,\omega_2}(s_1,s_2),\mathbb{D}^2_{\omega_1,\omega_2}(s_1,s_2)\right\}_{(\omega_1,\omega_2,s_1,s_2)}$ *is a bivariate centered Gaussian process defined on* $[0,1]^2 \times S$ *with covariance kernel given by equations A.7.*

Proof: To prove this result it is sufficient to show the following two claims (see Theorem 1.5.4 and 1.5.7 in Vaart and Wellner (1996)):

1. For every $k \in \mathbb{N}$, and any $\omega_1^{(1)},\omega_2^{(1)},\ldots,\omega_1^{(k)},\omega_2^{(k)} \in [0,1]$ and $\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_k \in S$,

$$\sqrt{mn}\Delta\begin{pmatrix} \hat{D}^h_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{s}^{(1)}) \\ \hat{D}^v_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{s}^{(1)}) \\ \vdots \\ \hat{D}^h_{m,n,\Delta,\omega_1^{(k)},\omega_2^{(k)}}(\mathbf{s}^{(k)}) \\ \hat{D}^v_{m,n,\Delta,\omega_1^{(k)},\omega_2^{(k)}}(\mathbf{s}^{(k)}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbb{D}^1_{\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{s}^{(1)}) \\ \mathbb{D}^2_{\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{s}^{(1)}) \\ \vdots \\ \mathbb{D}^1_{\omega_1^{(k)},\omega_2^{(k)}}(\mathbf{s}^{(k)}) \\ \mathbb{D}^2_{\omega_1^{(k)},\omega_2^{(k)}}(\mathbf{s}^{(k)}) \end{pmatrix}. \tag{3.9}$$

2. For every $\epsilon,\eta > 0$, there exists $\delta > 0$ such that for large m,n,M,N the probability

$$P\left(\sup_{d_\beta\left((\omega_1^{(1)},\omega_2^{(1)},\mathbf{s}_1),(\omega_1^{(2)},\omega_2^{(2)},\mathbf{s}_2)\right)<\delta}\sqrt{mn}\Delta\left\|\begin{pmatrix} \hat{D}^h_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{s}^{(1)}) - \hat{D}^h_{m,n,\Delta,\omega_1^{(2)},\omega_2^{(2)}}(\mathbf{s}^{(2)}) \\ \hat{D}^v_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{s}^{(1)}) - \hat{D}^v_{m,n,\Delta,\omega_1^{(2)},\omega_2^{(2)}}(\mathbf{s}^{(2)}) \end{pmatrix}\right\| > \eta\right) < \epsilon \tag{3.10}$$

   where the norm d_β is defined as in Theorem 1.5.4 in Vaart and Wellner (1996).

To show (3.9) it is enough to show that all cumulants of the random vector in the left hand side of (3.9) converge to the corresponding cumulants of the vector in the right hand side of the equation. Thus we have to show that for any $\omega_1,\omega_2,\omega_1^{(j)},\omega_2^{(j)} \in [0,1]$ and

$\mathbf{s}, \mathbf{s}^{(j)} \in S$ for $j = 1, 2, \ldots, l$

$$E\left(\hat{D}^d_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s})\right) = o(1), \quad \text{for } d = h, v. \tag{3.11}$$

$$\operatorname{cov}\left(\hat{D}^d_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{s})^{(1)}, \hat{D}^d_{m,n,\Delta,\omega_1^{(2)},\omega_2^{(2)}}(\mathbf{s})^{(2)}\right) = C^d_{\omega_1^{(1)},\omega_2^{(1)},\omega_1^{(2)},\omega_2^{(2)}}(\mathbf{s}^{(1)}, \mathbf{s}^{(2)}), \tag{3.12}$$

where $C^d_{\omega_1^{(1)},\omega_2^{(1)},\omega_1^{(2)},\omega_2^{(2)}}$ is as defined in equation A.7, and

$$\operatorname{cum}\left(\hat{D}^d_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{s}^{(1)}), \ldots, \hat{D}^d_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}(\mathbf{s}^{(l)})\right) = o(1) \tag{3.13}$$

for $l \geq 3$. A detailed derivation of (3.11),(3.12) and (3.13) are presented in the Appendix A.1.

To show (3.10) we note that it is sufficient to show the stochastic equicontinuity of each component $\sqrt{mn}\Delta\hat{D}^h$ and $\sqrt{mn}\Delta\hat{D}^v$ separately. Moreover, we can write

$$\begin{aligned}
\sqrt{mn}\Delta\hat{D}^d_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s}) =& \frac{1}{\sqrt{mn}\Delta} \sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}} I_{s_1 - m\Delta, s_2, m,n,\Delta}(\lambda_1, \lambda_2) \\
&- \frac{1}{\sqrt{mn}\Delta} \sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}} I_{s_1, s_2, m,n,\Delta}(\lambda_1, \lambda_2) \\
=& A_{m,n,\Delta}(\omega_1, \omega_2, s_1, s_2) - B_{m,n,\Delta}(\omega_1, \omega_2, s_1, s_2).
\end{aligned} \tag{3.14}$$

Therefore it is sufficient to show the stochastic equicontinuity of the processes $\{A_{m,n,\Delta}(\omega_1, \omega_2, s_1, s_2)\}$ and $\{B_{m,n,\Delta}(\omega_1, \omega_2, s_1, s_2)\}$. We only present the proof for the first summand in (3.14). The proof for the second term can be shown analogously. Let

$$\begin{aligned}
\mathcal{P} =& \{0, \frac{1}{m}, \ldots, 1 - \frac{1}{m}\} \times \{0, \frac{1}{n}, \ldots, 1 - \frac{1}{n}\} \times \\
& \{m\Delta, 2m\Delta, \ldots, (M - m)\Delta\} \times \{n\Delta, 2n\Delta, \ldots, (N - n)\Delta\}.
\end{aligned}$$

Note that with the choice of our frequency grid $\Lambda_{m,n,\Delta,\omega_1,\omega_2}$ and spatial grid, it is again sufficient to show that

$$P\left(\sup_{B_\beta(\delta)}\left|A_{m,n,\Delta}(\omega_1^{(1)},\omega_2^{(1)},s_1^{(1)},s_2^{(1)})-A_{m,n,\Delta}(\omega_1^{(2)},\omega_2^{(2)},s_1^{(2)},s_2^{(2)})\right|\right) \tag{3.15}$$

is small for large M, N where

$$B_\beta(\delta)=\{x_j:=(\omega_1^{(j)},\omega_2^{(j)},s_1^{(j)},s_2^{(j)})\in\mathcal{P}, j=1,2 \mid d_\beta(x_1,x_2)<\delta\}.$$

For this purpose let $C(u,d_\beta,\mathcal{P})$ denote the covering number of $\mathcal{P}$ with respect to the semi-metric d_β and define the corresponding covering integral of $\mathcal{P}$ by

$$J_{M,N}(\kappa):=\int_0^\kappa\left[\log\left(\frac{C^2(u,d_\beta,\mathcal{P})}{u}\right)\right]^2 du.$$

In subsection A.1.4 we show that

$$\lim_{\kappa\to 0}\lim_{M,N\to\infty} J_{M,N}(\kappa)=0 \tag{3.16}$$

$$\begin{aligned}&\mathbf{E}\left(m^{k/2}n^{k/2}\Delta^k\left(A_{m,n,\Delta}(\omega_1^{(1)},\omega_2^{(1)},s_1^{(1)},s_2^{(1)})-A_{m,n,\Delta}(\omega_1^{(2)},\omega_2^{(2)},s_1^{(2)},s_2^{(2)})\right)^k\right)\\ \leq\;& (4k!)C^k\left[d_\beta\left((\omega_1^{(1)},\omega_2^{(1)},s_1^{(1)},s_2^{(1)}),(\omega_1^{(2)},\omega_2^{(2)},s_1^{(2)},s_2^{(2)})\right)\right]^k,\end{aligned} \tag{3.17}$$

for a constant $C>0$ and even integers k. It then follows by similar arguments as given in Dahlhaus (1988) that the probability in (3.15) can be made smaller than every $\epsilon>0$ by choosing $\delta>0$ sufficiently small and M, N large enough, which proves stochastic equicontinuity.

Remark 3.2.1. *Note that the assumption of Gaussianity is strong, but wide used in modelling geostatistical data. Our proofs and proposed methodology are also valid under a weaker* *assumption that for all $l \in \mathbb{N}$ and all $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_l \in S$,*

$$\sum_{\mathbf{s}_1,\ldots,\mathbf{s}_l} |cum(Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_l))| < \infty$$

3.3 Test Statistic

To test the second-order stationarity of the observed random field we consider this spectral difference process. The main idea behind our proposed test is that this process is uniformly small if and only if the covariance structure is stationary. If there is non-stationarity present, then this process should be high at some spatial region and over some frequency range.

To this end we define the directional maximal difference statistics $T^h_{M,N,m,n,\Delta}$ and $T^v_{M,N,m,n,\Delta}$ as:

$$T^h_{M,N,m,n,\Delta} = \sup_{(s_1,s_2),\omega_1,\omega_2} \left| \hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(s_1, s_2) \right|, \tag{3.18}$$

$$T^v_{M,N,m,n,\Delta} = \sup_{(s_1,s_2),\omega_1,\omega_2} \left| \hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(s_1, s_2) \right|. \tag{3.19}$$

In practice, M, N, Δ are usually given by the data, and we need to choose the window size $m\Delta \times n\Delta$. If we select $m << M$, $n << N$, then we are searching for the structural breaks in a high resolution. While a smaller window can capture more local variation, there will be less observations in a window and the estimation bias could be large. In addition, a smaller window size leads to more comparisons, which could be computationally expensive. To balance the estimation accuracy, computational cost and the detection sensitivity, we define an overall maximal difference test statistics as:

$$\mathcal{T} = \max\left\{\sqrt{mN}\Delta T^h_{M,N,m,N,\Delta}, \sqrt{Mn}\Delta T^v_{M,N,M,n,\Delta}\right\}. \tag{3.20}$$

The motivation of this definition is to reject the null hypothesis when there exists a break in covariance along any one of the comparison direction. When searching for breaks in the horizontal (vertical) direction, we set $n = N$ $(m = M)$ to include more observations, averaging the dynamics across the orthogonal direction of our search direction. Therefore, when conducting the horizontal (vertical) comparison, we only need to choose the value of $m(n)$, which depends on the actual covariance structure of the underlying field.

More insight about these choices are revealed observing a plot of $\hat{D}^d_{m,n,\Delta,\omega_1,\omega_2}(g_1, g_2)$ with different m, n. This plot is termed disparity map, where higher values correspond to more abrupt change in the local spectral density estimated by $m \times n$ observations. When setting a small value for m, n, we can have some comparison results close to the boarders of the field. If there exist clear hot spots or hot band, we can have some ideas about the locations of potential boundaries and choose m, n accordingly.

3.4 Critical Value

To construct a statistical test based on the statistic (3.20) we state the following Corollary of Theorem 3.2.1

Corollary 3.4.1. *Under the assumptions of Theorem 3.2.1 we have*

$$\mathcal{T} = \max\left\{\sqrt{mN}\Delta T^h_{M,N,m,N,\Delta}, \sqrt{Mn}\Delta T^v_{M,N,M,n,\Delta}\right\}$$

$$\xrightarrow{d} \mathcal{D} := \max\left\{\max_{\mathbf{s}} \sup_{\omega_1,\omega_2} \left| \mathbb{D}^1_{\omega_1,\omega_2}(\mathbf{s}) \right|, \max_{\mathbf{s}} \sup_{\omega_1,\omega_2} \left| \mathbb{D}^2_{\omega_1,\omega_2}(\mathbf{s}) \right|\right\},$$

where $\mathbb{D}^1_{\omega_1,\omega_2}$ *and* $\mathbb{D}^2_{\omega_1,\omega_2}$ *are as defined in Theorem 3.2.1.*

Proof. First note that the assumptions of Theorem 3.2.1 remains valid with the choice of $m = M$ or $n = N$. Thus arguments similar to the proof of Theorem 3.2.1 and argmax continuous mapping Theorem then guarantees the stated convergence. □

Remark 3.4.1. *A closed form expression for the covariance kernel of* $(\mathbb{D}^1_{\omega_1,\omega_2}, \mathbb{D}^2_{\omega_1,\omega_2})$ *is derived in the Appendix A.1.2.*

Corollary 3.4.1 and Remark 3.4.1 provides us a way to construct an asymptotic level α test for stationarity. The following result ensures the consistency of a test based on our proposed statistic.

Theorem 3.4.1. *Suppose the underlying random field is Gaussian and piece-wise stationary with zero mean and covariance structure as in* (2.10). *Additionally assume (ii) and (iii) from Theorem 3.2.1 and*

$$\int_{[0,\infty)^2} \|h\| C_{ij}(\mathbf{h}) d\mathbf{h} < \infty, \quad i, j = 1, 2, \ldots, K. \tag{3.21}$$

. Then we have $\mathcal{T} \xrightarrow{p} \infty$.

Proof. The proof follows directly from Lemma A.2.1 □

To obtain the rejection region of the test of stationarity we simulate the quantiles of $\mathcal{D}$ by simulating zero mean Gaussian processes $\mathbb{D}^1$ and $\mathbb{D}^2$ repeatedly. The steps of implementation of our method are described in Algorithm 1.

Algorithm 1: Obtain the reject region

Result: Reject region $[U_h, \infty]$, $[U_v, \infty]$, $[U_{max}, \infty]$

Simulate $\mathbb{D}^1_{\omega_1,\omega_2}(g_1, g_2)$ for $g_1 \in \{o_1 + m\Delta, \ldots, o_1 + (M - m)\Delta\}, g_2 = o_2, \omega_1, \omega_2 \in \{0.1, \ldots, 1\}$ 10,000 times;

Compute the 95% percentile U_h of $\max_{g_1,g_2} \max_{\omega_1,\omega_2} \left|\mathbb{D}^1_{\omega_1,\omega_2}(g_1, g_2)\right|$ based on the simulated data;

Simulate $\mathbb{D}^2_{\omega_1,\omega_2}(g_1, g_2)$ for $g_1 = o_1, g_2 \in \{o_2 + n\Delta, \ldots, o_2 + (N - n)\Delta\}, \omega_1, \omega_2 \in \{0.1, \ldots, 1\}$ 10,000 times;

Compute the 95% percentile U_v of $\max_{g_1,g_2} \max_{\omega_1,\omega_2} \left|\mathbb{D}^2_{\omega_1,\omega_2}(g_1, g_2)\right|$ based on the simulated data;

Let $U_{max} = \max\{U_h, U_v\}$;

We then reject the null hypothesis if the test statistics $\mathcal{T}$ is above the simulate critical value U_{max}, and accept it otherwise. Additionally if $\sqrt{mN}T^h_{M,N,m,N,\Delta} > U_h$ (or $\sqrt{Mn}T^v_{M,N,M,n,\Delta} > U_v$), that indicates there exists breaks in the horizontal (or vertical) direction.

3.5 Simulation

We evaluate the performance of the proposed test under different settings. Let $Exp(\sigma^2, l)$ stands for the exponential covariance: $C(h_1, h_2) = \sigma^2 \exp(-\frac{\sqrt{h_1^2+h_2^2}}{l})$, and $Gauss(\sigma^2, l)$ stands for the Gauss covariance: $C(h_1, h_2) = \sigma^2 \exp(-\frac{h_1^2+h_2^2}{l})$. A visualization of the stationary covariance models used in the simulation can be found in Figure 3.1.

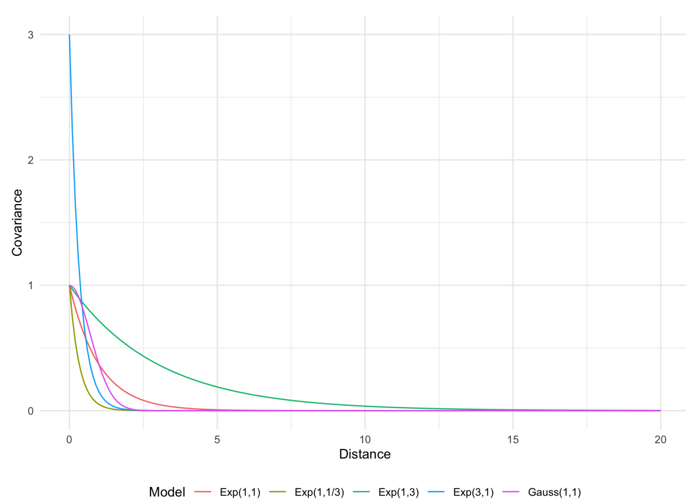


Figure 3.1: Stationary covariance models for the simulation.

For evaluating type I error, we generate equi-spaced data with $\Delta = 1$ from a stationary zero-mean Gaussian random field with covariance $Exp(\sigma^2, l)$ with $M = N = 20, 40, 80, 160, 320$. For each scenario, m and n, the side lengths of the windows are 1/10, 1/5, 2/5 of the corresponding M and N.

We consider two types of nonstationarity in the power evaluation: (1) piecewise stationary with abrupt change and (2) gradual change in covariance structure. For (1), we assume the simulated random field consists of two stationary sub-fields that are mutually exclusive and independent. For (2), we consider the covariance function developed in C. J. Paciorek and Schervish (2006), and the exact form we use is the same as the model NS1 in the Bandyopadhyay et al. (2016), which is:

$$C_{PS}(\mathbf{s}_1, \mathbf{s}_2) = \left| \Sigma(\frac{\mathbf{s}_1}{\rho}) \right|^{1/4} \left| \Sigma(\frac{\mathbf{s}_2}{\rho}) \right|^{1/4} \left| \frac{1}{2}[\Sigma(\frac{\mathbf{s}_1}{\rho}) + \Sigma(\frac{\mathbf{s}_2}{\rho})] \right|^{-1/2} \exp\{-\sqrt{q_\rho(\mathbf{s}_1, \mathbf{s}_2)}\}, \tag{3.22}$$

where $|\cdot|$ denotes the determinant of a matrix, $\mathbf{s} = (s_1, s_2) \in \mathbb{R}^2$,

$$q_\rho(\mathbf{s}_1) = 2(\mathbf{s}_1 - \mathbf{s}_2)^T [\Sigma(\frac{\mathbf{s}_1}{\rho}) + \Sigma(\frac{\mathbf{s}_2}{\rho})]^{-1} (\mathbf{s}_1 - \mathbf{s}_2), \tag{3.23}$$

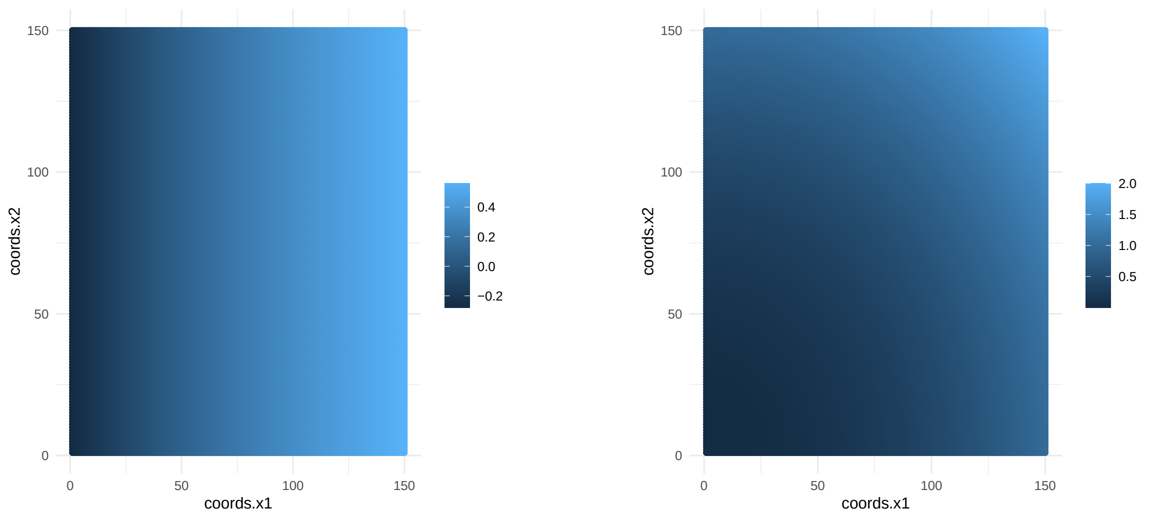
$$\Sigma(\frac{\mathbf{s}}{\rho}) = \Gamma(\frac{\mathbf{s}}{\rho}) \Lambda \Gamma(\frac{\mathbf{s}}{\rho})^T, \tag{3.24}$$

$$\Gamma(\frac{\mathbf{s}}{\rho}) = \begin{pmatrix} \gamma_1(\mathbf{s}/\rho) & -\gamma_2(\mathbf{s}/\rho) \\ \gamma_2(\mathbf{s}/\rho) & \gamma_1(\mathbf{s}/\rho) \end{pmatrix}, \tag{3.25}$$

$$\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}, \tag{3.26}$$

$$\gamma_1(\mathbf{s}/\rho) = \log(s_1/\rho + 0.75), \tag{3.27}$$

$$\gamma_2(\mathbf{s}/\rho) = (s_1/\rho)^2 + (s_2/\rho)^2. \tag{3.28}$$



(a) γ_1 (b) γ_2

Figure 3.2: Visualizations of γ_1 and γ_2 in the definition of C_{PS}, which result in a covariance structure that vary smoothly in the horizontal and anti-diagonal direction

The data are generated from the models below with $M, N = 20, 40, 80, 160$ on a regular grid:

- (NS1) The two sub-fields are separated by a horizontal boundary in the middle. Both sub-fields are zero-mean Gaussian random fields, with different covariance C_1 and C_2.
- (NS2) The two sub-fields are separated by a diagonal boundary. Both sub-fields are zero-mean Gaussian random fields, with different covariance C_1 and C_2.
- (NS3) The field is a zero-mean Gaussian random field with the nonstationary covariance C_{PS}, with $\rho = M = N$. The covariance structure has a gradual change in the horizontal and anti-diagonal directions.

The simulation settings and the results are summarized in Table 3.1 - 3.3, where Rejection Rate (RR) H, Rejection Rate (RR) V, and Rejection Rate (RR) Max stand for the frequency of rejecting the null hypothesis based on the horizontal comparison, vertical comparison, and the test statistics T in each scenario. We set the critical value to be the maximum of the U_h and U_v as defined in Algorithm 1.

In Table 3.1, the rejection rate reflects the type I error of the test. The model S1, S2 and S3 differ in the range parameter, and they are visualized in Figure 3.1. For fixed values of M and N, the rejection rates from the horizontal comparison and vertical comparison get closer to 0.05 , and the rejection rates for T get closer to 0.0975 as m and n increase.

In Table 3.2 and Table 3.3, the rejection rate reflects the power of the test. For fixed values of M and N, all the rejection rates increase to 1 as m and n increase, though the power is quite low in scenario 37-48. The reason for the low power may be that $Exp(1,1)$ and $Gauss(1,1)$ are similar in shape, as can be seen in Figure 3.1. The fields generated in scenario 25-36 are the same as those in scenario 37-48, except the covariance of the subfields are changed. $Exp(3,1)$ and $Exp(1,1)$ are quite different in shape, and the rejection rates are much higher that those in scenario 37-48. In scenario 1-24, the two sub-fields are separated by a vertical boundary, therefore the rejection rate based on the horizontal comparison is much higher than the rate based on the vertical comparison. When the covariance changes smoothly in space rather than having breaks as in Table 3.3, the power is lower, but still goes to 1 as the window size increases.

Table 3.1: Stationary scenarios for type I error evaluation

Scenario	Model	Covariance Model	M, N	m, n	RR H	RR V	RR Max
1	S1	$Exp(1, 1)$	20	2	0.06	0.06	0.11
2	S1	$Exp(1, 1)$	20	4	0.04	0.04	0.07
3	S1	$Exp(1, 1)$	20	8	0.06	0.06	0.10
4	S1	$Exp(1, 1)$	40	4	0.04	0.04	0.07
5	S1	$Exp(1, 1)$	40	8	0.07	0.07	0.11
6	S1	$Exp(1, 1)$	40	16	0.07	0.07	0.12
7	S1	$Exp(1, 1)$	80	8	0.06	0.06	0.11
8	S1	$Exp(1, 1)$	80	16	0.07	0.06	0.13
9	S1	$Exp(1, 1)$	80	32	0.08	0.06	0.13
10	S1	$Exp(1, 1)$	160	16	0.07	0.06	0.12
11	S1	$Exp(1, 1)$	160	32	0.06	0.08	0.13
12	S1	$Exp(1, 1)$	160	64	0.07	0.07	0.14
13	S1	$Exp(1, 1)$	320	32	0.08	0.07	0.14
14	S1	$Exp(1, 1)$	320	64	0.07	0.06	0.13
15	S1	$Exp(1, 1)$	320	128	0.07	0.07	0.13
16	S2	$Exp(1, 3)$	20	2	0.13	0.14	0.22
17	S2	$Exp(1, 3)$	20	4	0.11	0.12	0.18
18	S2	$Exp(1, 3)$	20	8	0.09	0.09	0.14
19	S2	$Exp(1, 3)$	40	4	0.13	0.13	0.22
20	S2	$Exp(1, 3)$	40	8	0.11	0.11	0.18
21	S2	$Exp(1, 3)$	40	16	0.09	0.08	0.14
22	S2	$Exp(1, 3)$	80	8	0.11	0.11	0.18
23	S2	$Exp(1, 3)$	80	16	0.09	0.09	0.15
24	S2	$Exp(1, 3)$	80	32	0.08	0.08	0.14
25	S2	$Exp(1, 3)$	160	16	0.09	0.09	0.17
26	S2	$Exp(1, 3)$	160	32	0.08	0.07	0.14
27	S2	$Exp(1, 3)$	160	64	0.07	0.07	0.13
28	S2	$Exp(1, 3)$	320	32	0.08	0.08	0.15
29	S2	$Exp(1, 3)$	320	64	0.07	0.07	0.13
30	S2	$Exp(1, 3)$	320	128	0.06	0.06	0.12
31	S3	$Exp(1, 1/3)$	20	2	0.00	0.00	0.01
32	S3	$Exp(1, 1/3)$	20	4	0.00	0.00	0.00
33	S3	$Exp(1, 1/3)$	20	8	0.01	0.01	0.03
34	S3	$Exp(1, 1/3)$	40	4	0.00	0.00	0.00
35	S3	$Exp(1, 1/3)$	40	8	0.01	0.01	0.02
36	S3	$Exp(1, 1/3)$	40	16	0.03	0.04	0.06
37	S3	$Exp(1, 1/3)$	80	8	0.00	0.01	0.01
38	S3	$Exp(1, 1/3)$	80	16	0.03	0.03	0.05
39	S3	$Exp(1, 1/3)$	80	32	0.05	0.05	0.09
40	S3	$Exp(1, 1/3)$	160	16	0.03	0.02	0.05
41	S3	$Exp(1, 1/3)$	160	32	0.04	0.04	0.09
42	S3	$Exp(1, 1/3)$	160	64	0.06	0.05	0.10
43	S3	$Exp(1, 1/3)$	320	32	0.04	0.04	0.07
44	S3	$Exp(1, 1/3)$	320	64	0.05	0.05	0.09
45	S3	$Exp(1, 1/3)$	320	128	0.06	0.06	0.11

Table 3.2: Nonstationary scenarios: sharp change in covariance

Scenario	Model	Covariance Model	M, N	m, n	RR H	RR V	RR Max
1	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	2	0.51	0.02	0.51
2	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	4	0.63	0.05	0.64
3	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	8	1.00	0.17	1.00
4	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	40	4	0.94	0.11	0.94
5	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	40	8	1.00	0.24	1.00
6	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	40	16	1.00	0.27	1.00
7	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	80	8	1.00	0.14	1.00
8	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	80	16	1.00	0.18	1.00
9	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	80	32	1.00	0.18	1.00
10	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	160	16	0.99	0.02	0.99
11	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	160	32	1.00	0.02	1.00
12	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	160	64	1.00	0.03	1.00
13	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	20	2	0.12	0.03	0.15
14	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	20	4	0.08	0.02	0.09
15	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	20	8	0.34	0.09	0.39
16	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	40	4	0.11	0.01	0.11
17	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	40	8	0.54	0.11	0.58
18	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	40	16	0.98	0.15	0.98
19	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	80	8	0.55	0.05	0.55
20	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	80	16	1.00	0.08	1.00
21	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	80	32	1.00	0.10	1.00
22	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	160	16	0.99	0.02	0.99
23	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	160	32	1.00	0.02	1.00
24	NS1	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	160	64	1.00	0.04	1.00
25	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	2	0.31	0.08	0.36
26	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	4	0.25	0.05	0.28
27	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	8	0.67	0.54	0.84
28	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	40	4	0.37	0.06	0.39
29	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	40	8	0.80	0.57	0.90
30	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	40	16	0.99	0.96	1.00
31	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	80	8	0.58	0.28	0.68
32	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	80	16	1.00	0.99	1.00
33	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	80	32	1.00	1.00	1.00
34	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	160	16	0.56	0.40	0.71
35	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	160	32	1.00	1.00	1.00
36	NS2	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	160	64	1.00	1.00	1.00
37	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	20	2	0.12	0.03	0.13
38	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	20	4	0.08	0.01	0.08
39	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	20	8	0.14	0.06	0.19
40	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	40	4	0.09	0.01	0.09
41	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	40	8	0.16	0.06	0.21
42	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	40	16	0.19	0.12	0.28
43	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	80	8	0.10	0.02	0.11
44	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	80	16	0.11	0.05	0.15
45	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	80	32	0.15	0.09	0.22
46	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	160	16	0.02	0.01	0.03
47	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	160	32	0.03	0.02	0.04
48	NS2	$C_1 = Exp(1,1), C_2 = Gauss(1,1)$	160	64	0.10	0.04	0.12

Table 3.3: Nonstationary scenarios: graduate change in covariance

Scenario	Model	Covariance Model	M, N	m, n	RR H	RR V	RR Max
1	NS3	C_{PS}	20	2	0.04	0.05	0.11
2	NS3	C_{PS}	20	4	0.05	0.05	0.09
3	NS3	C_{PS}	20	8	0.14	0.12	0.23
4	NS3	C_{PS}	40	4	0.05	0.05	0.09
5	NS3	C_{PS}	40	8	0.14	0.12	0.24
6	NS3	C_{PS}	40	16	0.50	0.35	0.65
7	NS3	C_{PS}	80	8	0.07	0.07	0.13
8	NS3	C_{PS}	80	16	0.31	0.21	0.44
9	NS3	C_{PS}	80	32	0.97	0.84	1.00
10	NS3	C_{PS}	160	16	0.07	0.05	0.10
11	NS3	C_{PS}	160	32	0.57	0.37	0.68
12	NS3	C_{PS}	160	64	1.00	1.00	1.00

Chapter 4: Spatial Partition

When the test in Chapter 3 suggests the data are from a nonstationary random field, the spectral difference statistics defined in Section 3.1 may be used to locate the potential breaks and partition the nonstationary random field into several stationary sub-fields.

Suppose there is a piecewise stationary zero-mean random field Z_S consists of K non-overlapping second-order stationary sub-fields $Z_{S_1}, \ldots, Z_{S_K}$. The boundaries between $S_1, \ldots, S_K$ defines the partition and the covariance structure is given by (2.10). We assume that K is minimal in this representation in the sense that Z_S cannot be partitioned into fewer stationary sub-fields. In this chapter, we still assume the field Z_S is observed over a regular grid G with spacing Δ. We also assume that $S_1, S_2, \ldots, S_K$ have linear boundaries.

The partition procedure consists of two steps. In the first step, we create an initial partition by an iterative search. In the second step, we reduce the number of regions in the initial partition by combining the connected regions with similar spectral densities. This two steps procedure is motivated by Pope et al. (2021) and image segmentation algorithms based on clustering the super-pixels (see a review in Cong et al. (2018)). The proposed spatial partition procedures are summarised in Algorithm 2 and Algorithm 3.

4.1 Initial Partition

The first step of the algorithm is to find all possible boundaries within the random fields, creating an initial partition of K_0 non-overlapping rectangular sub-fields, such that each the sub-field is second-order stationary. The boundaries between the sub-fields are selected by the spectral difference statistics in Section 3.1. We use a tree-like iterative search procedure where we conduct the horizontal comparison and vertical comparisons alternatively in each of the sub-regions created in the previous iteration, and declare new potential boundaries at coordinates where the value of spectral difference statistics exceeds a data-driven threshold. In order to pick up a single boundary we use a max-suppression procedure commonly used in time-series change point detection literature such as Preuss et al. (2015). This picks up a boundary where the spectral difference statistics is largest first, and set value of the statistics at near-by locations to be 0 in order to prevent over-identifying boundaries. In each iteration, we create a finer partition, and we will stop when we cannot further partition the field based on the rules defined in Algorithm 2.

In algorithm 2, given a rectangular region S_k let M_k and N_k respectively be the numbers of horizontal and vertical coordinates in $S_k \cap G$, $O_k = (o_k^1, o_k^2)$ be the origin (bottom-left corner) of S_k, m (n) be the parameter that controls the horizontal (vertical) side length of the windows for comparison as in equation 3.2.

During an iteration in Algorithm 2, adding J_k elements in $B_2^h = \{b_1^h, \ldots, b_{J_k}^h\}$ means partitioning S_k into J_k+1 sub-rectangles separated by the vertical boundaries defined by B_2^h. Similarly, when J_k elements are added to $B_2^v = \{b_1^v, \ldots, b_{J_k}^v\}$ it corresponds to partitioning S_k into $J_k + 1$ sub-regions separated by the horizontal boundaries defined by B_2^v.

The consistency of this partitioning algorithm has been investigated in section A.3. In particular, Theorem A.3.1 guarantees that for appropriate choice of the threshold, the probability of picking all the coordinates along which there is a boundary converges to 1, and picking any wrong partition converges to 0. Although in practice it will in fact give us the coordinates of the partition along each direction, rather than the actual partition lines. So we are expected to do a over-partitioning using this algorithm, which is later rectified by the merging algorithm.

Algorithm 2: Create the initial partition

Result: $P = \{S_k\}, k = 1, \dots, K_0$

Set the initial partition to be $P = \{S_1\}$, where $S_1 = S$, $K_0 = 0$ and $K_h = 1$;

while $\min_k M_k >= m$, $\min_k N_k >= n$, $K_h > K_0$ **do**

  Set $K_0 = K_h$ and $B_1^h = B_2^h = \emptyset$;

  for $k = 1 \dots, K_h$ **do**

    for $g_1 = o_k^1 + m\Delta, \dots, o_k^1 + (M_k - m)\Delta$ **do**

      Let $g_1 \in B_1^h$ if $(mN_k\Delta^2)^\gamma \sup_{\omega_1,\omega_2} \hat{D}^h_{m,N_k,\Delta,\omega_1,\omega_2}(g_1, o_k^2) > \epsilon^h_{m,N_k,\Delta}(g_1, o_k^2)$;

    end

    while $B_1^h \neq \emptyset$ **do**

      Include $b^h = \arg\max_{g_1 \in B_1^h} \sup_{\omega_1,\omega_2} \hat{D}^h_{m,N_k,\Delta,\omega_1,\omega_2}(g_1, o_k^2)$ into B_2^h and eliminate $\{b^h - m + 1, b^h + m - 1\}$ from B_1^h;

    end

  end

  Set $K_v = \sum_{k=1}^{K_h}(J_k + 1)$ and $B_1^v = B_2^v = \emptyset$# current $P = \{S_1, \dots, S_{K_v}\}$;

  for $k = 1, \dots, K_v$ **do**

    for $g_2 = o_k^2 + n\Delta, \dots, o_k^2 + (M_k - n)\Delta$ **do**

      Let $g_2 \in B_1^v$ if $(M_k n\Delta^2)^\gamma \sup_{\omega_1,\omega_2} \hat{D}^v_{M_k,n,\Delta,\omega_1,\omega_2}(o_k^1, g_2) > \epsilon^v_{M_k,n,\Delta}(o_k^1, g_2)$;

    end

    while $B_1^v \neq \emptyset$ **do**

      Include $b^v = \arg\max_{g_2 \in B_1^v} \sup_{\omega_1,\omega_2} \hat{D}^v_{M_k,n,\Delta,\omega_1,\omega_2}(o_k^1, g_2)$ into B_2^v and eliminate $\{b^v - n + 1, b^v + n - 1\}$ from B_1^v

    end

  end

  Set $K_h = \sum_{k=1}^{K_v}(J_k + 1)$# current $P = \{S_1, \dots, S_{K_h}\}$;

end

4.1.1 Choice of parameters

- The algorithm requires choosing a tuning parameter $\gamma \in (0, 0.5]$, which controls detection sensitivity. A smaller value of γ suppresses partitioning, while a larger value promotes partitioning. $\gamma = 0.4$ is recommended in Preuss et al. (2015) in a time series context. We use $\gamma = 0.5$ in all simulations to be a little more aggressive when picking up potential boundaries, and we rely on the additional merging step to guard against over-partition. In practice, this parameter can be set on a case by case basis based on the data.

- The algorithm also requires specification of the comparison window size, namely m and n. A large value of m, n prevents the generation of small sub-fields, but may miss the breaks that are close to the boarders. We can use the proposed test and the disparity map to help selecting these values. If the test rejects the null hypothesis when $m = m_0, n = n_0$, that not only indicates the second-order characteristics of the field may vary across space, but also tell us the potential boundaries may be picked up when $m = m_0, n = n_0$, hence we could use them as the parameters for partition. We can also plot the disparity map with $m = m_0, n = n_0$ to see if the partition generated by Algorithm 2 makes sense.

- We use a data driven threshold for comparing the spectral difference statistic to pick potential partition location. The horizontal comparison threshold $\epsilon^h_{m,n,\Delta}(g_1, g_2)$ is defined as:

$$\epsilon^h_{m,n,\Delta}(g_1, g_2) = \sqrt{2V_{g_1,g_2,m,n,\Delta} \log\left(\frac{M_k}{m}\right)}, \tag{4.1}$$

where $V_{g_1,g_2,m,n,\Delta} = \frac{1}{2mn}\frac{\Delta^4}{(2\pi)^4 m^2 n^2}\sum_{\lambda_1,\lambda_2}[I_{W^{h1}_{g_1,g_2,m,n,\Delta} \cup W^{h2}_{g_1,g_2,m,n,\Delta}}(\lambda_1, \lambda_2)]^2$. For the vertical comparison threshold $\epsilon^v_{m,n,\Delta}(g_1, g_2)$ we define:

$$\epsilon^v_{m,n,\Delta}(g_1, g_2) = \sqrt{2V_{g_1,g_2,m,n,\Delta} \log\left(\frac{N_k}{n}\right)}, \tag{4.2}$$

where $V_{g_1,g_2,m,n,\Delta} = \frac{1}{2mn}\frac{\Delta^4}{(2\pi)^4 m^2 n^2}\sum_{\lambda_1,\lambda_2}[I_{W^{v1}_{g_1,g_2,m,n,\Delta}\cup W^{v2}_{g_1,g_2,m,n,\Delta}}(\lambda_1,\lambda_2)]^2$. The form of the thresholds is motivated by Preuss et al. (2015) and Fan (1996).

4.2 Merging Regions with Similar Spectral Density

The initial partition gives us rectangular region, some of which may still have same covariance structure. Now these K_0 sub-regions created in the initial partition can be treated as the data for clustering, where each cluster contains some connected sub-regions and the procedure can be thought of as merging the sub-regions one by one in $K_0 - 1$ steps as in an agglomerative hierarchical clustering. In Algorithm 3, the input at step r is $P^r = \{P^r_k, k = 1, \ldots, K_0\}$, where each element P^r_k is a set of some sub-regions from $S_k, k \in 1, \ldots, K_0$. In particular, at step 1, the set of elements to cluster is $P^1 = \{P^1_k = \{S_k\}, k = 1, \ldots, K_0\}$, where $S_k \in P$ is defined in Algorithm 2.

Algorithm 3: Hierarchical clustering

Result: The cluster assignment at step r: $P^r = \{P_k^r\}$, $k = 1, \ldots, K_0$, $r = 1, \ldots, K_0 - 1$

Calculate the set of largest pairwise integrated periodogram difference $PD = \{\delta_{j,k}\}$, where

$$\delta_{j,k} = \frac{1}{MN\Delta^2} \sup_{\omega_1,\omega_2 \in [0,1]} \sum_{\Lambda_{M,N,\Delta,\omega_1,\omega_2}} I_{G_j}(\lambda_1, \lambda_2) - I_{G_k}(\lambda_1, \lambda_2)$$

for $j, k = 1, \ldots, K_0$, where $M = \max M_j, N = \max N_j$ for $j = 1, \ldots, K_0$;

Create a set $E = \{e_{j,k}, j, k = 1, \ldots, K_0\}$. $e_{j,k}$ is 1 if S_j and S_k are connected, and is 0 if not;

Create a set $d = \emptyset$;

for $r = 1, \ldots, K_0 - 1$ **do**

Find a pair of j, k such that $e_{j,k} = 1$, and $\delta_{j,k}$ is the smallest value in PD, let $P_j^{r+1} = P_k^{r+1} = P_j^r \cup P_k^r$, $P_l^{r+1} = P_l^r$ for $l \neq j, k$;

Eliminate $\delta_{j,k}$ and $\delta_{k,j}$ from PD;

Add $d_r = \delta_{j,k}$ to d;

For all P_i^r containing regions connected to some regions in P_j or P_k, set $e_{i',j'}$, $e_{j',i'}$, $e_{i',k'}$, $e_{k',i'}$ to be 1 for all i', j', k' such that $S_{i'} \in P_i^r, S_{j'} \in P_j^r, S_{k'} \in P_k^r$;

end

Create a dendrogram to visualize the clustering results, where each leaf node is a region in the initial partition, and the merge at step r happens when the height axis takes the value d_r;

Create a scree plot by plotting d_r against r;

In this algorithm, the connected regions are merged based on a similarity measurement determined by the average difference in spectral density. This merging can be visualized by a dendogram. The appropriate number of cluster is chosen using the knee of a scree plot.

Another more objective way of choosing the number of clusters is using the gap statistic as defined in Tibshirani et al. (2001). However we need to modify the original definition proposed by Tibshirani et al. (2001) where we treat the sub-regions from the initial partition as observations. Other forms of the gap statistics such as Mohajer et al. (2010) can also be used. Once we have the cluster dendogram we use the following algorithm to choose the number of clusters. The dendogram is then cut accordingly to produce the final partition.

Algorithm 4: Gap statistics

Result: K clusters to retain

Let $U^r = \{U_l^r, l = 1, \ldots, r\}$ be the set of unique elements in P^{K_0-r+1};

for $r = 1, \ldots, K_0$ **do**

Calculate $W_r = \sum_{l=1}^{r} \frac{1}{2|U_l^r|} D_l$, where $D_l = \sum_{j,k \ s.t.\ S_j, S_k \in U_l^r} \delta_{j,k}$;

Calculate the gap statistics:

$$Gap(r) = E_B^* \left(\log(W_r^b)\right) - \log(W_r) = \frac{1}{B} \sum_{b=1,\ldots,B} \log(W_r^b) - \log(W_r),$$

where $E_B^* \left(\log(W_r^b)\right)$ is the average of B copies $\log(W_r^b)$, each of which is computed from a Monte Carlo sample $I_1, \ldots, I_{K_0}$ drawn from the reference distribution;

Calculate $sd_r = \left[\frac{1}{B} \sum_{b=1,\ldots,B} [\log(W_r^b) - E_B^* \left(\log(W_r^b)\right)]^2\right]^{1/2}$ and define $s_r = sd_r \sqrt{1 + \frac{1}{B}}$;

end

Choose the number of clusters K to be

$K =$ smallest r s.t. $GD(r) = Gap(r) - Gap(r+1) + s_{r+1} \geq 0$;

To generate $I_1, \ldots, I_{K_0}$, we follow the following steps. Let X to be a $K_0 \times J$ matrix, where $J = MN$. The entries $X_{k,j}$ has the value $I_{G_k}(\lambda_j)$, where λ_j is the jth element in $\Lambda_{M,N,\Delta,\omega_1,\omega_2}$. Let Y be the normalized X by subtracting the column mean $\bar{X}$ from each column. Then we compute the singular value decomposition $Y = UDV^T$. Let $Y' = YV$, and then draw I' uniformly over the ranges of the columns of Y'. Finally, we compute $I = I'V^T + \bar{X}$, and the sample $I_k, k = 1, \ldots, K_0$ in Algorithm 4 is the kth row of I.

4.3 Simulation

We investigate the performance of spatial partitioin algorithms in Section 5 in the above nonstationary scenarios NS1, NS2, NS3, and also

- (NS4) The two sub-fields are separated by a square-shaped boundary in the middle, and the side length of the boundary is the half of the side length of the whole field. Both sub-fields are zero-mean Gaussian random fields, with different covariance C_1 and C_2.

We set $M = N = 150$, $m = n = 32$ in all the scenarios.

4.3.1 Rand Index

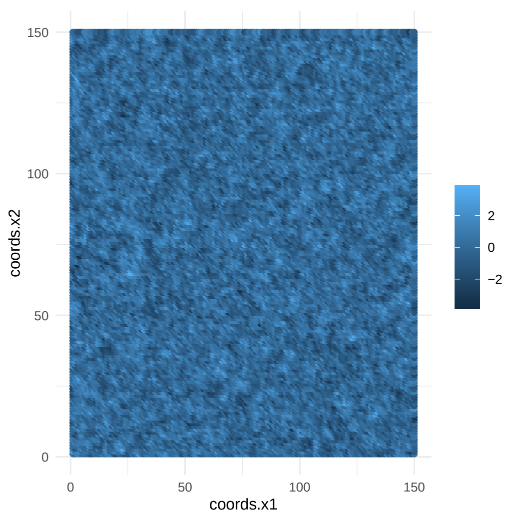
We generate 100 sets of observations from the non-stationary models NS1, NS2 and NS4 to see how well our method performs on average. We assume we know the correct number of sub-regions and cut the dendrogram accordingly. To evaluate the performance, we use the Rand index, which is a measure of the similarity between two cluster assignments. It takes value in $[0, 1]$, and a value greater than 0.65 means moderate agreement and a value of 1 means perfect agreement. In each repetition, we compute the rand index between the true cluster assignment and the cluster assignment given by our algorithm. The averaged rand index in each scenario is presented in Table 4.1. We can see that when the linear boundaries are orthogonal to the searching directions, the method picks up the correct boundaries most of the time. However, in NS2, where the boundary is not orthogonal to the searching directions, the method can only approximate the true boundary.

Table 4.1: Rand Index

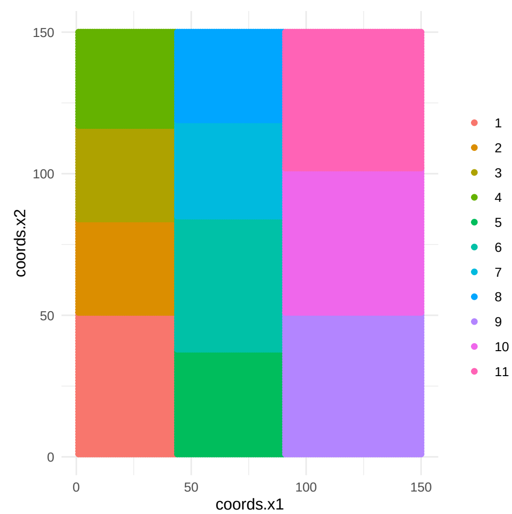
Scenario	M,N	m,n	Averaged Rand Index
NS1	150	8	0.50
NS2	150	8	0.51
NS4	150	8	0.79
NS1	150	16	0.88
NS2	150	16	0.54
NS4	150	16	0.83
NS1	150	32	1.00
NS2	150	32	0.66
NS4	150	32	0.96

4.3.2 Visualization of Results with One Repitition

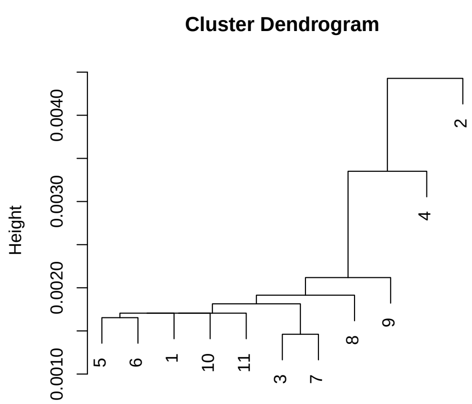
In this subsection, we generate one set of observations from each scenarios and visualize the partition results in Figure 4.1-4.5. The number of sub-regions in the final partitions are picked manually based on checking the dendrograms. We can see that in the case of abrupt change, the boundary selected by the proposed method is closed to the correct one. While in the case of graduate change, our method provides a piecewise stationary approximation to the original field and the location of the boundaries seems to have the clearest change in covariance.



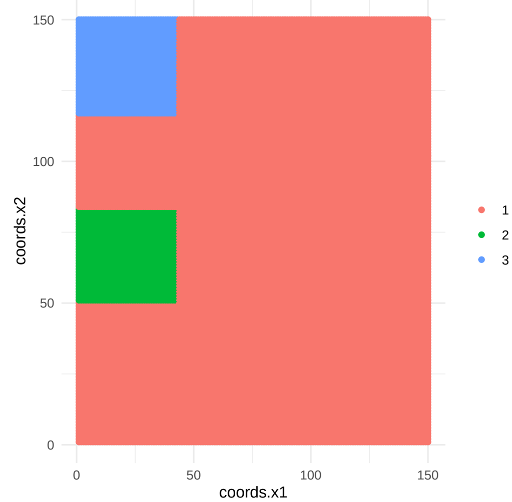
(a) Simulated random field



(b) Initial partition

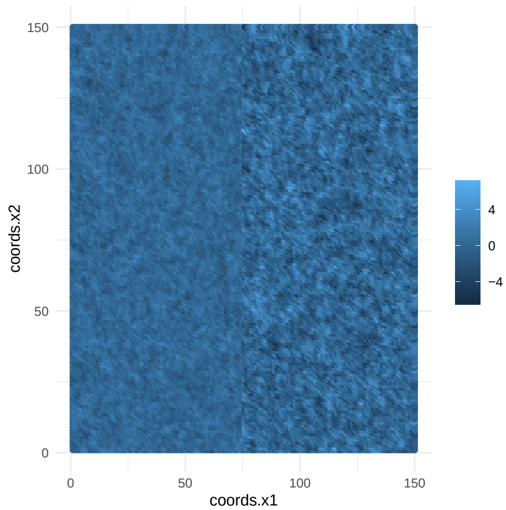


(c) Dendrogram

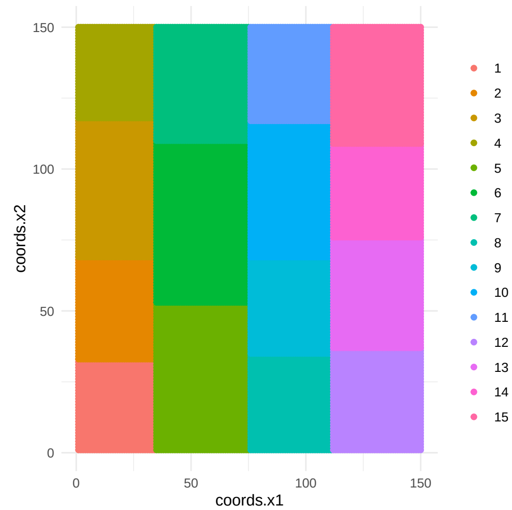


(d) Final partition

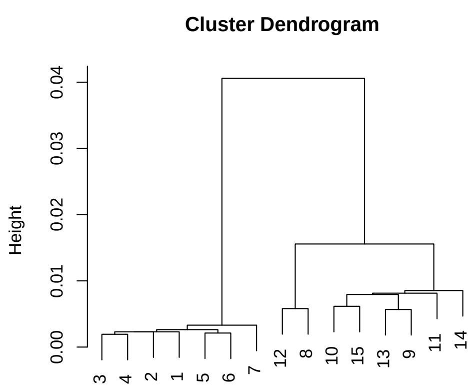
Figure 4.1: S1: Stationary random field with $Exp(1, 1)$ covariance. $M = N = 150$, $m = n = 32$. Critical value $U = 0.28$. Test statistics $\mathcal{T} = 0.24$. The proposed test fails to reject H_0.



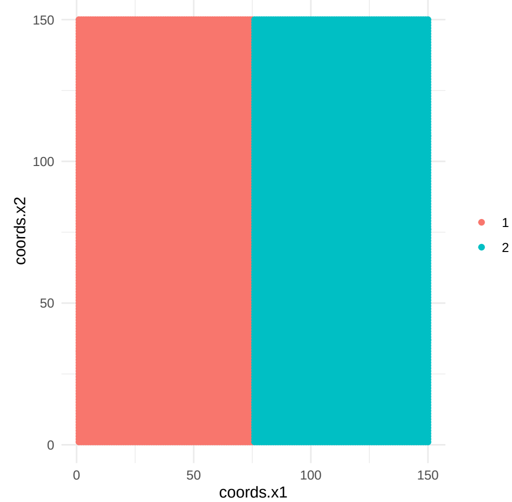
(a) Simulated random field



(b) Initial partition

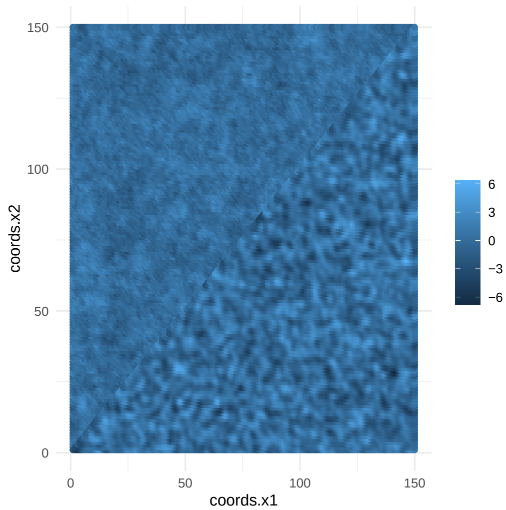


(c) Dendrogram

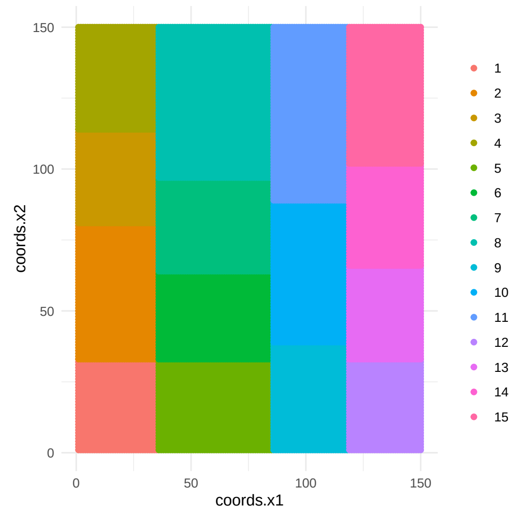


(d) Final partition

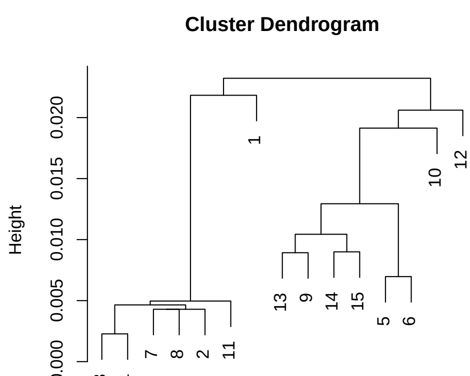
Figure 4.2: NS1: Nonstationary random field with a boundary in the middle. $C_1 = Exp(1,1)$, $C_2 = Exp(3,1)$, $M = N = 150$, $m = n = 32$. Critical value $U = 0.55$. Test statistics $\mathcal{T} = 3.55$. The proposed test rejects H_0.



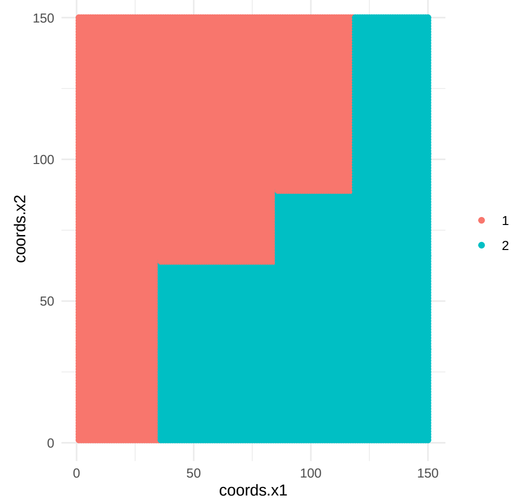
(a) Simulated random field



(b) Initial partition



(c) Dendrogram



(d) Final partition

Figure 4.3: NS2: Nonstationary random field with a boundary in the anti-diagonal direction. $C_1 = Exp(1, 1)$, $C_2 = Exp(3, 1)$, $M = N = 150$, $m = n = 32$. Critical value $U = 0.61$. Test statistics $\mathcal{T} = 1.28$. The proposed test rejects H_0.

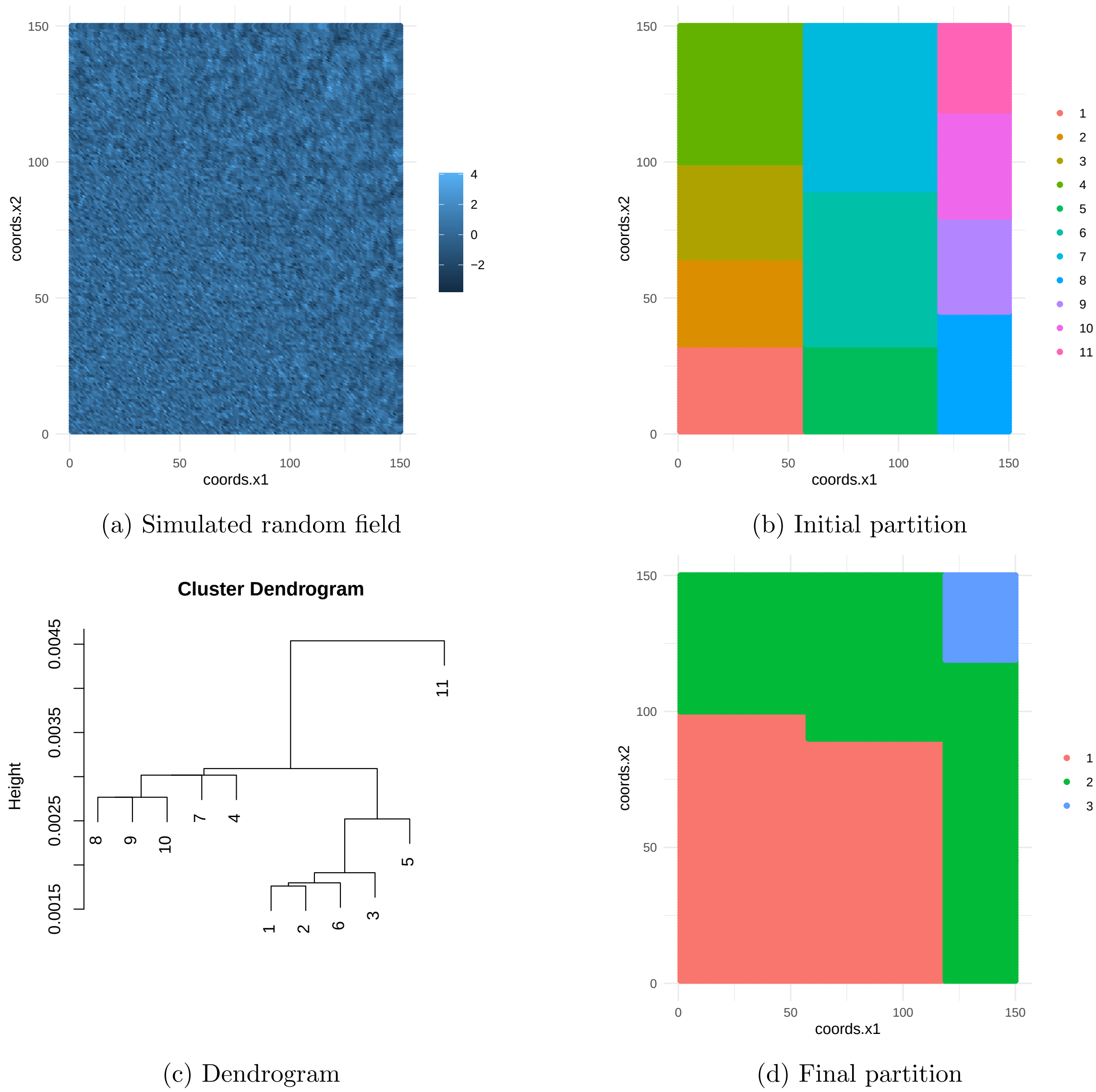
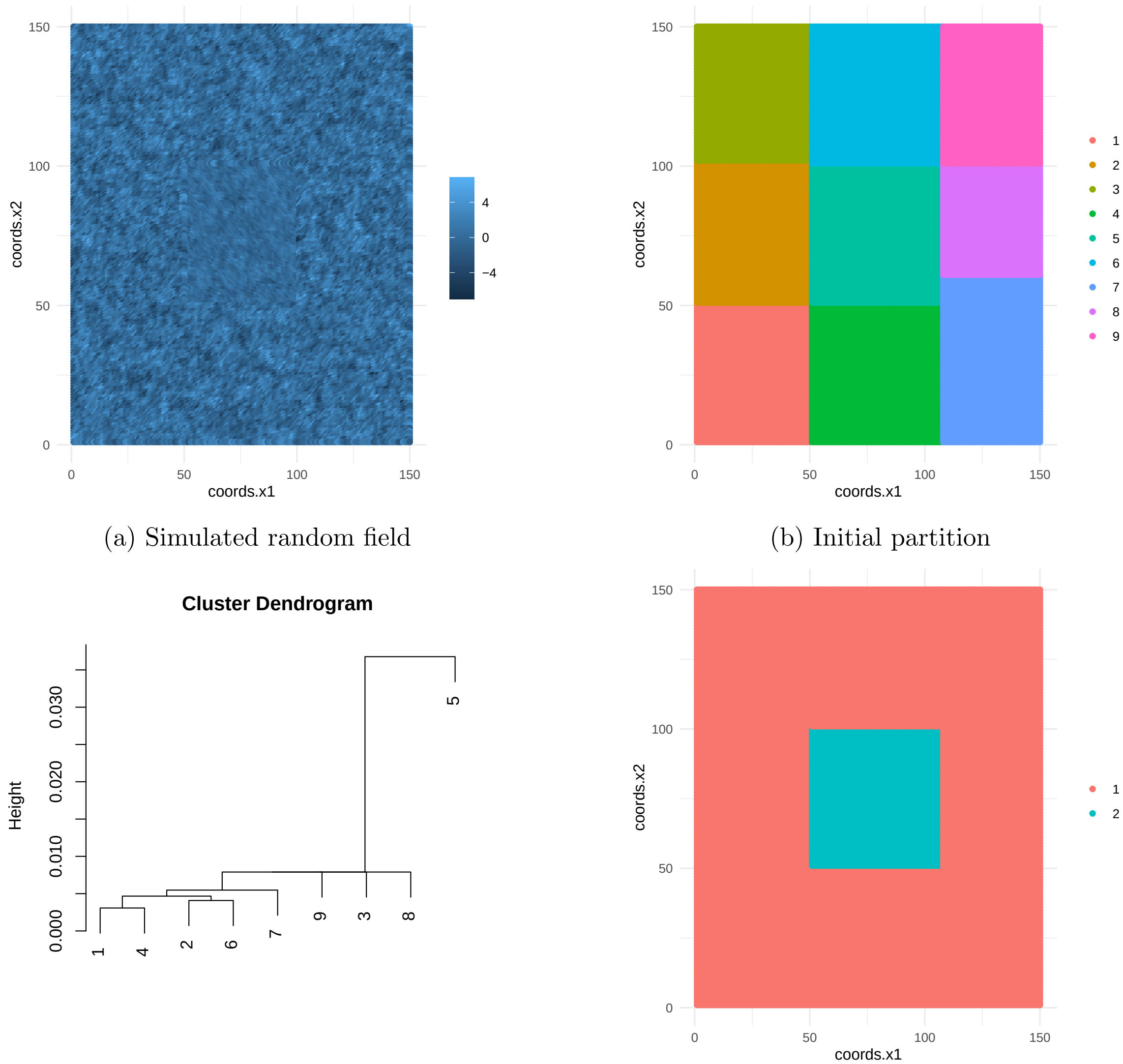


Figure 4.4: NS3: Nonstationary random field with a gradually changing covariance C_{PS}. $\rho = M = N = 150$, $m = n = 32$. Critical value $U = 0.25$. Test statistics $\mathcal{T} = 0.27$. The proposed test rejects H_0.



(a) Simulated random field (b) Initial partition

(c) Dendrogram (d) Final partition

Figure 4.5: NS4: Nonstationary random field with a squared shape boundary. $C_1 = Exp(1,1)$, $C_2 = Exp(3,1)$, $M = N = 150$, $m = n = 32$. $M = N = 150$, $m = n = 32$. Critical value $U = 0.81$. Test statistics $\mathcal{T} = 1.60$. The proposed test rejects H_0.

Chapter 5: Irregular Spaced Data

In the previous chapters, we assume the field Z is observed on a grid. However, regular spaced data are not always available. Following Section 2.5.2, we assume Z is observed at $\mathcal{I} = \{\mathbf{s}_j = (s_j^1, s_j^2) \in [0, L_1] \times [0, L_2], j = 1, 2, \ldots, J\}$. In this chapter we extend the methods developed in the previous chapters to handle such irregular spaced data.

5.1 Test for Stationarity

Recall that for observations on a regular grid, our proposed test in Chapter 3 requires comparing of spectral densities using a moving window method, such that the regions in comparison ($W^{d_1}_{(s_1,s_2),m,n,\Delta}$ and $W^{d_2}_{(s_1,s_2),m,n,\Delta}$) are the regions covered by a moving $m\Delta \times n\Delta$ window placed at one step away from (s_1, s_2) and just at (s_1, s_2). The window moves Δ along direction d each time and it always has an observation at its origin.

For a random field observed over a general set of locations, we analogously define Δ as the step size of the moving window. Let $M = \frac{\lceil L_1 \rceil}{\Delta}, N = \frac{\lceil L_2 \rceil}{\Delta}$, and we still use a half open half close window of size $m\Delta \times n\Delta$ as defined in (3.3) and (3.4). A general version of the periodogram as defined in Section 2.5.2 is given by:

$$I_{s_1,s_2,m,n,\Delta}(\lambda_1, \lambda_2) = \frac{mn\Delta}{(2\pi)^2 J^2} \left| \sum_{j=1}^{J} Z(s_j^1, s_j^2) \exp(-i(s_j^1 \lambda_1 + s_j^2 \lambda_2)) \right|^2, \tag{5.1}$$

where J is the number of observations in the region covered by the window. Notice that when the observations are indeed on a regular grid with spacing Δ, (5.1) is exactly same as (2.6). With this extended notation, the quantities needed in the rest of the testing procedures can be written in the same form as in Chapter 3.

In particular, the horizontal spectral difference statistics for irregular spaced data is defined as:

$$\hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(s_1,s_2) = \hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(g_1,g_2)$$
$$= \frac{1}{mn\Delta^2} \sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}} \left[I_{g_1-m\Delta,g_2,m,n,\Delta}(\lambda_1,\lambda_2) - I_{g_1,g_2,m,n,\Delta}(\lambda_1,\lambda_2) \right], \tag{5.2}$$

where $(g_1, g_2) = (\lceil s_1 \rceil, \lceil s_2 \rceil) \in \{m\Delta, \ldots, (M-m)\Delta\} \times \{0, \ldots, (N-n)\Delta\}$. The vertical spectral difference statistic is similarly defined as:

$$\hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(s_1,s_2) = \hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(g_1,g_2)$$
$$= \frac{1}{mn\Delta^2} \sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}} \left[I_{g_1,g_2-n\Delta,m,n,\Delta}(\lambda_1,\lambda_2) - I_{g_1,g_2,m,n,\Delta}(\lambda_1,\lambda_2) \right], \tag{5.3}$$

where $(g_1, g_2) = (\lceil s_1 \rceil, \lceil s_2 \rceil) \in \{0, \ldots, (M-m)\Delta\} \times \{n\Delta, \ldots, (N-n)\Delta\}$. The directional maximal difference statistics $T^h_{M,N,m,n,\Delta}$ and $T^v_{M,N,m,n,\Delta}$ are:

$$T^h_{M,N,m,n,\Delta} = \sup_{(s_1,s_2),\omega_1,\omega_2} \left| \hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(s_1,s_2) \right|, \tag{5.4}$$

$$T^v_{M,N,m,n,\Delta} = \sup_{(s_1,s_2),\omega_1,\omega_2} \left| \hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(s_1,s_2) \right|. \tag{5.5}$$

And the overall maximal difference test statistics is:

$$\mathcal{T} = \max\left\{ \sqrt{mN}\Delta T^h_{M,N,m,N,\Delta}, \sqrt{Mn}\Delta T^v_{M,N,M,n,\Delta} \right\}. \tag{5.6}$$

To conduct the test, we also need the distribution of $\mathcal{T}$. However the asymptotic distribution of this quantity is dependent on the distribution of the points on which the random field is observed.

In practice, the number of observations J is known, and we need to choose Δ and m, n.

The choice of Δ is difficult, because it controls the window size and the step size. A smaller value of Δ leads to a smaller window, so less observations will be available for estimating the local spectral density and there may exist empty cells. There will also be more comparisons, which may be computationally expensive. A bigger window, however, may miss the local variation and the variation around the boarders. The choice of m, n faces a similar trade-off. In addition, the computation is much slower in this case, because we cannot make use of the fast Fourier transformation to calculate the periodogram.

We investigate the performance of the test by two simulations. In both scenarios, the locations are $i.i.d$ from a two dimensional uniform distribution Uniform$[0, 20) \times [0, 20)$. The step size Δ is chosen to be 1. The testing procedure follows Algorithm 1. Since we have not derived the asymptotic distribution yet, we generate the critical values U_h, U_v, U_{max} by assuming the observations are on a regular grid with spacing Δ.

In S1 (see Table 5.1), the underlying field is stationary, so the rejection rates reflect the size of the test. We observe that as observations get denser (increasing J), or the window gets larger (increasing m, n), the rejection rates reduce. However, comparing the results with the regular spaced data case (S1 in Table 3.1), the type I error seems to be much higher in this case. This is not surprising due to the wrong specification of the null distribution.

The scenario NS1 here is similar to NS1 in Chapter 3. The power of the test is reflected by the rejection rate in Table 5.2, which indicates the test has a high power in this case as the null hypothesis is rejected in all the 1,000 repetitions using level $\alpha = 0.05$. The decreasing type I error and high power indicates a similar rate of convergence as the regular grid case, however the results suggest the asymptotic distribution is probably different than the one derived in Theorem 3.2.1.

Table 5.1: Stationary scenarios

Scenario	Model	Covariance Model	M, N	J	m, n	RR H	RR V	RR Max
1	S1	$Exp(1,1)$	20	400	2	0.97	0.97	0.99
2	S1	$Exp(1,1)$	20	1600	2	0.44	0.42	0.64
3	S1	$Exp(1,1)$	20	6400	2	0.16	0.15	0.28
4	S1	$Exp(1,1)$	20	400	4	0.93	0.93	0.98
5	S1	$Exp(1,1)$	20	1600	4	0.34	0.38	0.54
6	S1	$Exp(1,1)$	20	6400	4	0.13	0.14	0.24
7	S1	$Exp(1,1)$	20	400	8	0.82	0.80	0.95
8	S1	$Exp(1,1)$	20	1600	8	0.29	0.29	0.44
9	S1	$Exp(1,1)$	20	6400	8	0.11	0.12	0.20

Table 5.2: Nonstationary scenarios

Scenario	Model	Covariance Model	M, N	J	m, n	RR H	RR V	RR Max
1	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	400	2	1.00	1.00	1.00
2	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	1600	2	1.00	1.00	1.00
3	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	6400	2	1.00	1.00	1.00
4	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	400	4	1.00	1.00	1.00
5	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	1600	4	1.00	1.00	1.00
6	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	6400	4	1.00	1.00	1.00
7	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	400	8	1.00	1.00	1.00
8	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	1600	8	1.00	1.00	1.00
9	NS1	$C_1 = Exp(1,1), C_2 = Exp(3,1)$	20	6400	8	1.00	1.00	1.00

Chapter 6: Applications

To illustrate our methods, we apply them to two datasets. Both datasets consist of regular-spaced data, and are publicly available online.

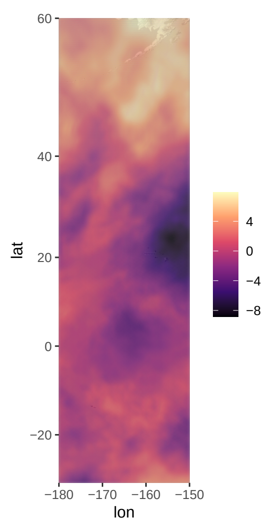
6.1 ERA5 Wind Data

The first dataset is from the ERA5 climate dataset (Hersbach et al. (2018)). In specific, we consider the hourly data of the V-component of the wind (horizontal speed of air moving towards the north, in metres per second) at the pressure level 300 hPa in December 2019 from the geographical region over 150°W-180°W in longitude, 60°N-30°S in latitude. The V-component of the wind, also known as the meridional flow is associated with more amplified troughs and ridges than the zonal flow (air moving towards east), and is indicative of the pole-ward transport of heat and the equator-ward transport of cold air. The more amplified the troughs and ridges are (i.e., the more meridional the flow is), the more extreme the associated weather tends to be. For example, heat waves and droughts typically occur under amplified upper-tropospheric ridges, while cold air outbreaks, clouds, and precipitation are frequently associated with various sectors of strong troughs. Understanding the meridional wind on this region help predict the occurrence and strength of El-Nino over Pacific coast.

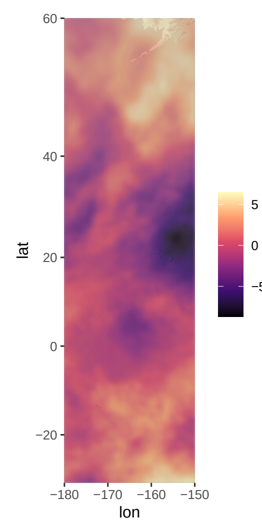
In this analysis we retain the observations on each day at 12:00 PM and take the average, so at each location we only have up to one averaged value. There are 121×361 observations, and the adjacent observations are 0.25 degree away in longitude or latitude. We estimate the mean of the data using two linear models. Model 1 only has an intercept term, so the residuals are the demeaned averaged hourly wind data. In model 2, the longitude and latitude are included as independent variables. The residuals from the models are visualized in Figure 6.1.

The mean structure of the field for the demeaned data seems to be nonstationary. The value is higher in the region close to Alaska, and there is a region around Hawaii where the value is lower. The inclusion of longitude and latitude in model 2 does not remove the trend. The covariance structure also seems to be nonstationary, as the values appear to be more homogeneous in the regions close to Alaska and Hawaii. We use our proposed method to test for stationarity over the entire region, with $m = n = 16$ and $m = 16$, $n = 32$. The null hypothesis is rejected in both cases. When $m = n = 16$, $\mathcal{T} = 10.51$, $U_{max} = 7.01$, p-value is 0.00. When $m = 16$, $n = 32$, $\mathcal{T} = 25.97$, $U_{max} = 7.15$, p-value is 0.00.

We then plot the disparity map of the residual process from the model 1, using $m = n = 8$ and $m = n = 16$. The disparity map indicates a difference between the land and ocean regions. The disparity maps and the initial partition using $m = n = 16$ and $m = 16$, $n = 32$ are presented in Figure 6.2.



(a) Residuals from model 1



(b) Residuals from model 2

Figure 6.1: ERA5 Wind Data

The final partition results are presented in Figure 6.3 and Figure 6.4. The gap statistics and the scree plot suggest there are 8 sub-fields when $m = n = 16$ and 4 sub-fields when $m = 16, n = 32$. In the 4 region partition, two of these regions correspond to areas around Alaska (partition 4) and Hawaii (partition 2). The other two regions (1 and 3) are over the Pacific ocean. While region 1 covers the largest ares, region 3 corresponds to the middle-left of the region showing a distinct behavior in our disparity map.

6.2 UCAR US Precipitation Data

As another example, we use the precipitation data from https://www.image.ucar.edu/Data/US.monthly.met/ as in Fuglstad et al. (2015). More specifically, we use the annual precipitation data measured in millimeters over a regular grid, such that adjacent observations are 0.25 degree away in longitude or latitude. The region we study is defined over 91.75°W-120°W in longitude, and 35°N-48°N in latitude. This area stretches across the mid-west to west part of the USA. The data contains 113×53 observations in total. Fuglstad et al. (2015) showed that a piecewise stationary field (STAT2 in the paper) such that the two stationary sub-fields are separated by a boundary at 100°W could be a good alternative to more flexible but complicated nonstationary models. This boundary is selected because it is roughly the boundary between the mountainous area in the west and the plain area in the east. We are interested in whether our method supports this partition.

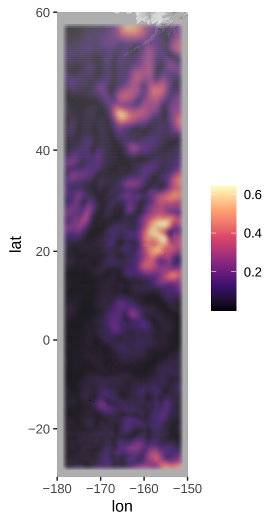
We estimate the mean using two linear models. Model 1 only has an intercept term, so the residuals are the demeaned annual precipitation data. Model 2 includes elevation as an independent variable. The original data and the residuals from the models are visualized in Figure 6.5.

A visual inspection of the plot of the residual process indicates a smooth variation of the random field across the eastern region, and a more abrupt variation over the mountainous area in the west. This strongly suggests the need of a non-stationary covariance. We test the hypothesis of stationarity for both the residual fields using $m = n = 16$ and the null hypothesis is rejected in both cases. For the residuals from model 1, $\mathcal{T} = 145.51$,

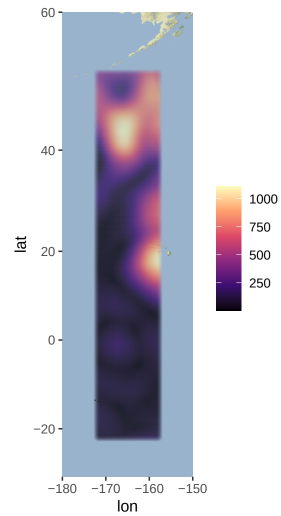
$U_{max} = 19.76$, p-value is 0.00. For the residuals from model 2, $\mathcal{T} = 106.57$, $U_{max} = 17.43$, p-value is 0.00. These results support our observation from the field plot of the residual processes.

Next we plot the disparity map for the residuals from model 1. We use $m = n = 8$ and $m = n = 16$ to construct these plots. These plots show a possible break in the north-west and another in south-east region. The disparity maps and the initial partition of the residual fields from both models are presented in Figure 6.6.

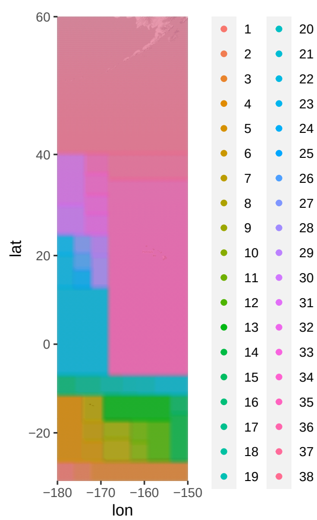
The final partition of the two residual fields when $m = n = 16$ are presented in Figure 6.7 and 6.8. The numbers of sub-regions are chosen to be 4 as suggested by the gap statistics or the scree plot. We note that the our method not only suggests a boundary around 96°W, which supports the choice for STAT2 in Fuglstad et al. (2015), but also suggests two additional stationary sub-fields in the residual fields.



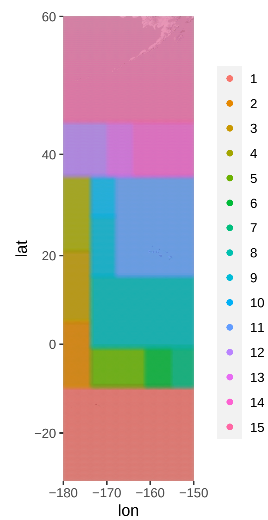
(a) Disparity map when $m = n = 8$



(b) Disparity map when $m = n = 16$

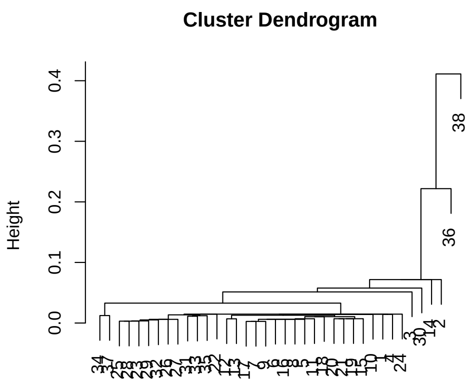


(c) Initial partition of the demeaned field when $m = n = 16$

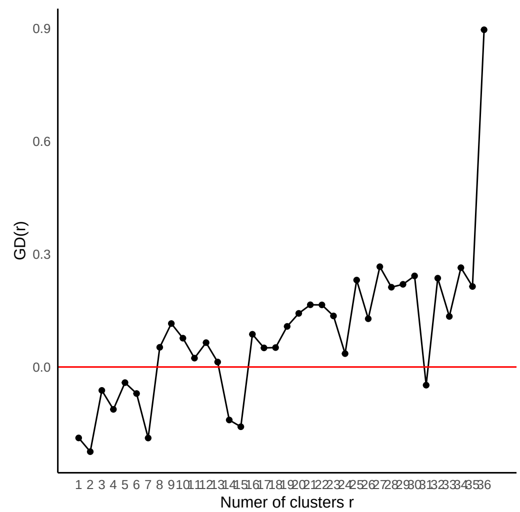


(d) Initial partition of the demeaned field when $m = 16, n = 32$

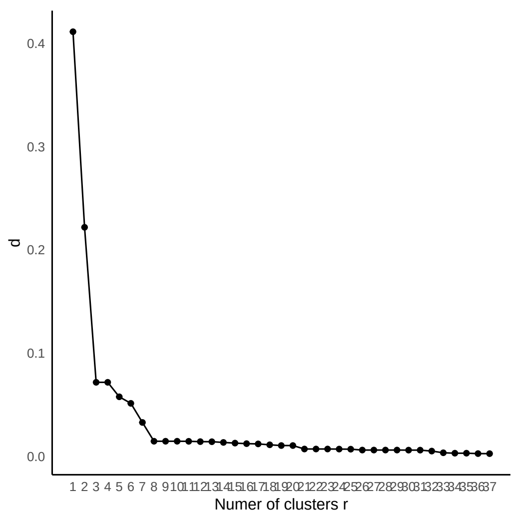
Figure 6.2: Disparity maps and the initial partitions



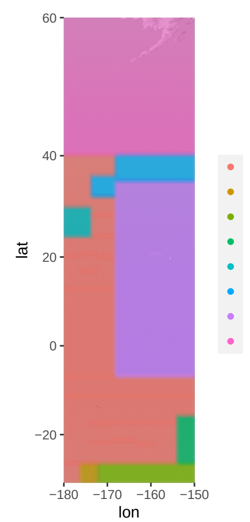
(a) Dendrogram



(b) Gap statistics difference curve

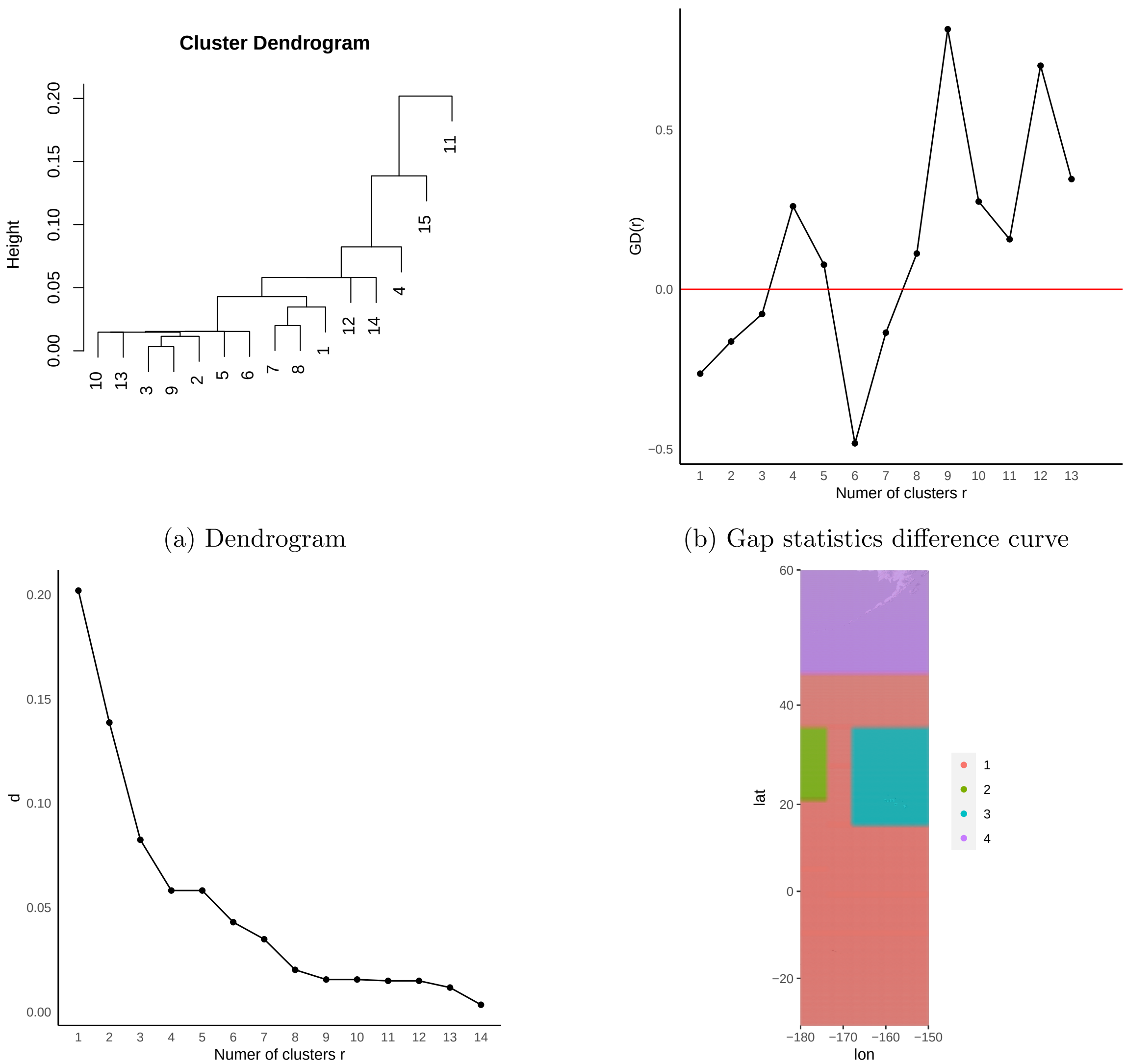


(c) Scree plot



(d) Final partition

Figure 6.3: Partition results for the demeaned data when $m = n = 16$



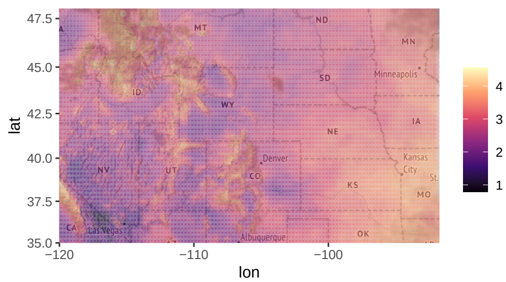
(a) Dendrogram

(b) Gap statistics difference curve

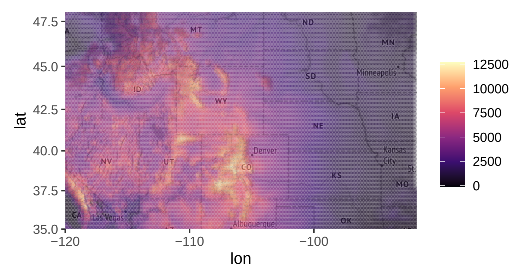
(c) Scree plot

(d) Final partition

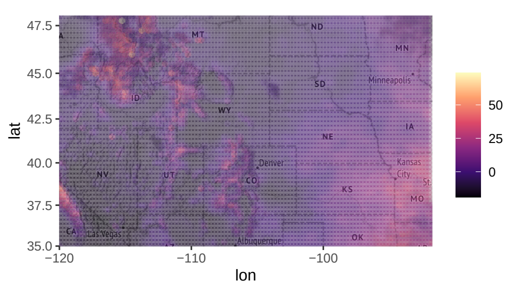
Figure 6.4: Partition results for the demeaned data when $m = 16, n = 32$



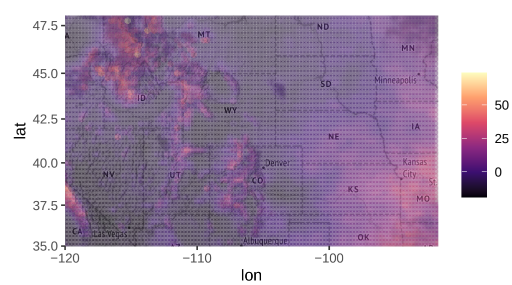
(a) Precipitation



(b) Elevation

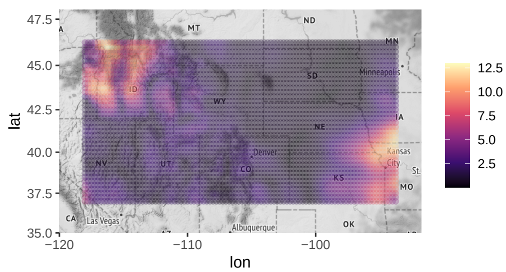


(c) Residuals from model 1

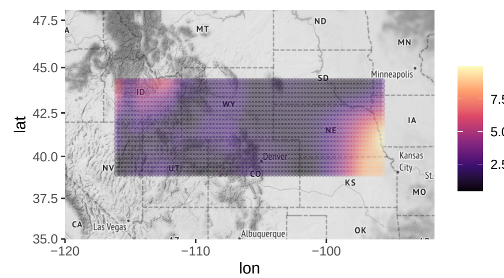


(d) Residuals from model 2

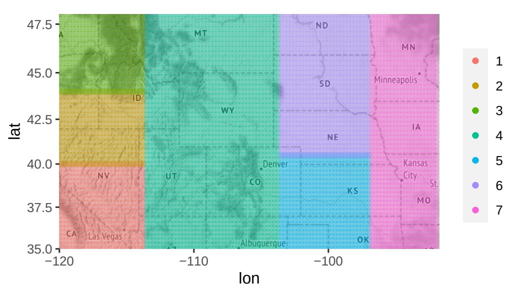
Figure 6.5: UCAR US precipitation data



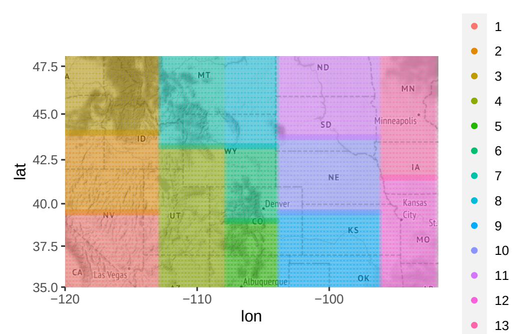
(a) Disparity map when $m = n = 8$



(b) Disparity map when $m = n = 16$

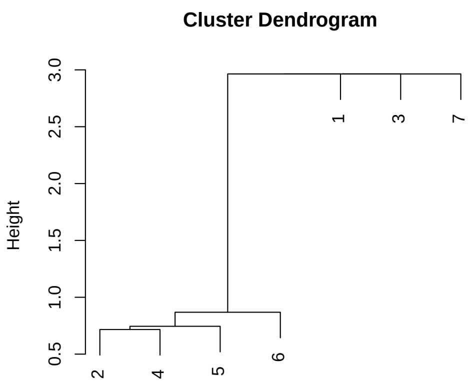


(c) Initial partition of the demeaned field when $m = n = 16$

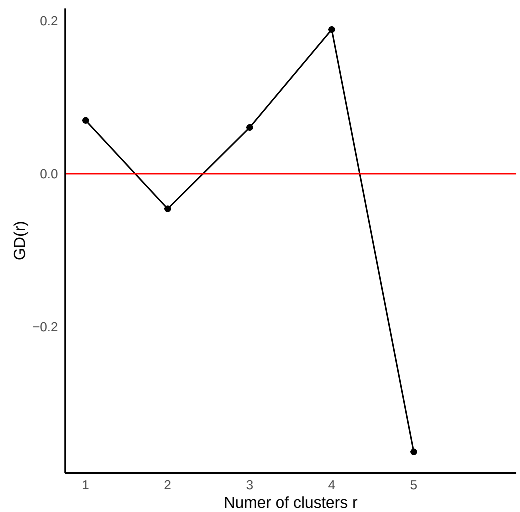


(d) Initial partition of the residual field from model 2 when $m = n = 16$

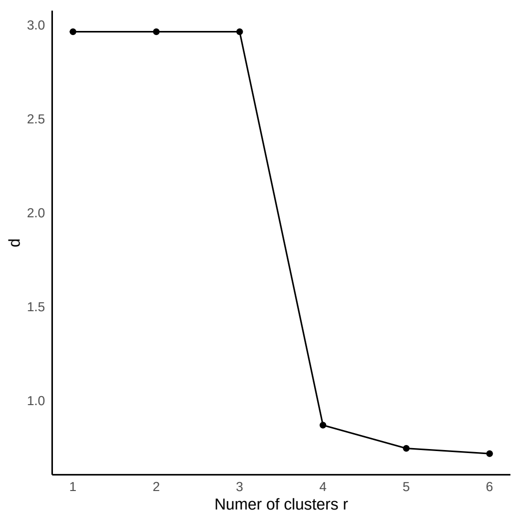
Figure 6.6: Disparity maps and the initial partitions



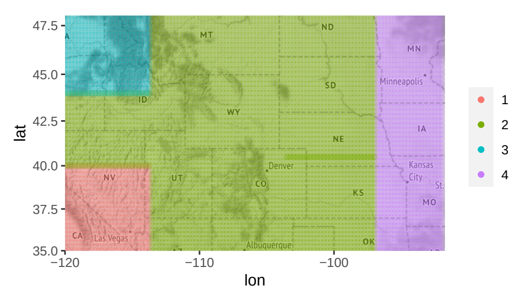
(a) Dendrogram



(b) Gap statistics difference curve

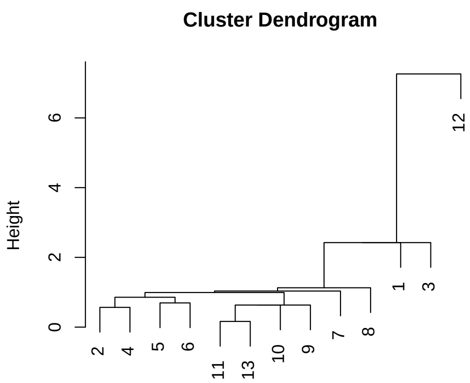


(c) Scree plot

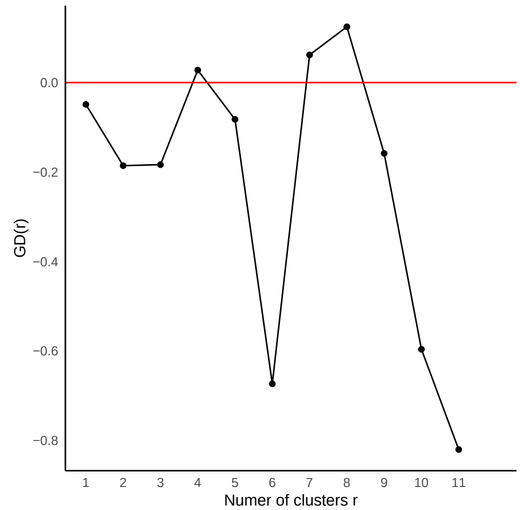


(d) Final partition

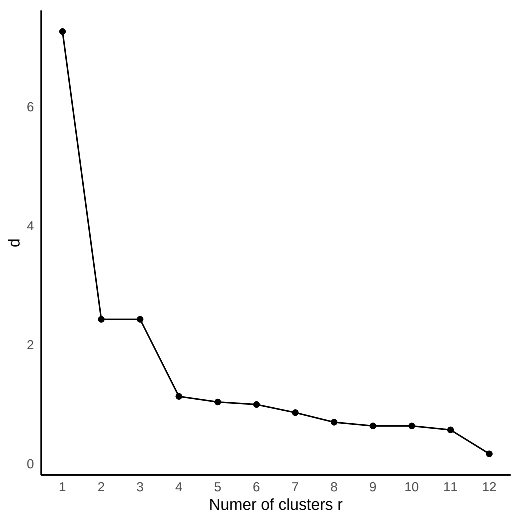
Figure 6.7: Partition results for the demeaned data when $m = n = 16$



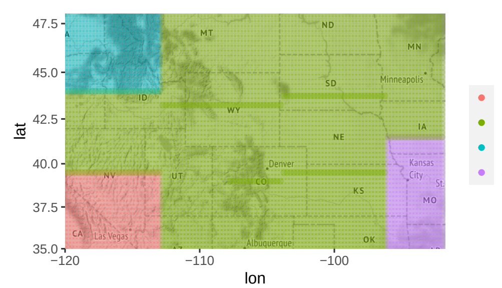
(a) Dendrogram



(b) Gap statistics difference curve



(c) Scree plot



(d) Final partition

Figure 6.8: Partition results for the residuals from model 2 when $m = n = 16$

Chapter 7: Conclusion and Future Work

In this work, we have proposed a method for testing stationarity and understanding the nature of non-stationarity present in spatial data observed over a grid. We further provide a method to find a piece-wise stationary approximation using linear boundaries. Our proposed partition methodology is the first in the literature that does not require a prior selection. We adopt some structural break ideas from time series literature, but adopt it in a 2-dimensional spatial field, and provide a computationally efficient implementation of the proposed methodology.

As an extension of this work, we plan to use this disparity map and spectral difference process to construct more general partition using Voronoi tessellation method. This requires deciding the number of centers and a distance measure within every location. We can use the disparity map to choose the center locations and use a modified version of this spectral difference measure to assess the distance. This will provide us with a non-linear boundary for the homogeneous regions.

We also explored the same questions from spatial data observed on a general set of irregularly spaced point. Our simulations indicate the usefulness of the spectral difference statistic and a strong possibility of extending this methodology in this set-up possibly with a different asymptotic distribution. We plan to derive rigorous characterization of the spectral density process to properly extend the test for stationarity and partitioning methodology.

Appendix A: Appendix

A.1 Properties of The Spectral Difference Statistics

A.1.1 Expectation

Under H_0, the expectation of the spectral difference statistics at location (g_1, g_2) is

$$E\left(\hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(g_1,g_2)\right)$$

$$=\frac{1}{mn\Delta^2}E\Big(\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}} I_{g_1-m\Delta,g_2,m,n,\Delta}(\lambda_1,\lambda_2)-I_{g_1,g_2,m,n\Delta}(\lambda_1,\lambda_2)\Big)$$

$$=\frac{1}{mn\Delta^2}\frac{\Delta^2}{(2\pi)^2mn}E\Big(\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}}\Big|\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}Z(g_1+\Delta(p-m),g_2+\Delta q)exp(-i\Delta(p\lambda_1+q\lambda_2))\Big|^2$$

$$-\Big|\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}Z(g_1+\Delta p,g_2+\Delta q)exp(-i\Delta(p\lambda_1+q\lambda_2))\Big|^2\Big)$$

$$=\frac{1}{mn\Delta^2}\frac{\Delta^2}{(2\pi)^2mn}\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}}E\Big(\Big|\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}Z(g_1+\Delta(p-m),g_2+\Delta q)exp(-i\Delta(p\lambda_1+q\lambda_2))\Big|^2$$

$$-\Big|\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}Z(g_1+\Delta p,g_2+\Delta q)exp(-i\Delta(p\lambda_1+q\lambda_2))\Big|^2\Big)$$

$$=\frac{1}{mn\Delta^2}\frac{\Delta^2}{(2\pi)^2mn}\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}}E\Big(A-B\Big),$$

where

$$A=\Big|\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}Z(g_1+\Delta(p-m),g_2+\Delta q)exp(-i\Delta(p\lambda_1+q\lambda_2))\Big|^2$$

$$=\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1}Z(g_1+\Delta(p-m),g_2+\Delta q)Z(g_1+\Delta(p'-m),g_2+\Delta q')$$

$$exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2]),$$

and

$$B = \Big| \sum_{p=0}^{m-1}\sum_{q=0}^{n-1} Z(g_1+\Delta p, g_2+\Delta q) exp(-i\Delta(p\lambda_1+q\lambda_2))\Big|^2$$

$$= \sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} Z(g_1+\Delta p, g_2+\Delta q) Z(g_1+\Delta p', g_2+\Delta q')$$

$$exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2]).$$

Then we have

$$E\Big(A\Big)$$

$$=E\Big(\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} Z(g_1+\Delta(p-m), g_2+\Delta q) Z(g_1+\Delta(p'-m), g_2+\Delta q')$$

$$exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big)$$

$$= \sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])$$

$$E\Big(Z(g_1+\Delta(p-m), g_2+\Delta q) Z(g_1+\Delta(p'-m), g_2+\Delta q')\Big),$$

where

$$E\Big(Z(g_1+\Delta(p-m), g_2+\Delta q) Z(g_1+\Delta(p'-m), g_2+\Delta q')\Big)$$

$$= \begin{cases} Var(Z(g_1,g_2)), & \text{if } p=p', q=q', \\ \\ C(h_1,h_2), & \text{if } p-p'=h_1, q-q'=h_2. \end{cases}$$

We also have

$$E\Big(B\Big)$$

$$=E\Big(\Big|\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}Z(g_1+\Delta p,g_2+\Delta q)exp(-i\Delta(p\lambda_1+q\lambda_2))\Big|^2\Big)$$

$$=E\Big(\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1}Z(g_1+\Delta p,g_2+\Delta q)Z(g_1+\Delta p',g_2+\Delta q')$$

$$exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big)$$

$$=\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1}exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])$$

$$E\Big(Z(g_1+\Delta p,g_2+\Delta q)Z(g_1+\Delta p',g_2+\Delta q')\Big),$$

where

$$E\Big(Z(g_1+\Delta p,g_2+\Delta q)Z(g_1+\Delta p',g_2+\Delta q')\Big)$$

$$=\begin{cases}Var(Z(g_1,g_2)), & \text{if } p=p',q=q',\\ \\ C(h_1,h_2), & \text{if } p-p'=h_1,q-q'=h_2.\end{cases}$$

Therefore

$$E\left(\hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(g_1,g_2)\right)=0. \tag{A.1}$$

Similarly we have

$$E\left(\hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(g_1,g_2)\right)=0. \tag{A.2}$$

A.1.2 Asymptotic Covariance

$$Cov\Big(\hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(g_1,g_2),\hat{D}^h_{m,n,\Delta,\omega_1',\omega_2'}(g_1',g_2')\Big)$$

$$=\frac{1}{m^2n^2\Delta^4}Cov\Big(\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}}I_{g_1-m\Delta,g_2,m,n,\Delta}(\lambda_1,\lambda_2)-I_{g_1,g_2,m,n,\Delta}(\lambda_1,\lambda_2),$$

$$\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1',\omega_2'}} I_{g_1'-m\Delta,g_2',m,n,\Delta}(\lambda_1,\lambda_2) - I_{g_1',g_2',m,n,\Delta}(\lambda_1,\lambda_2)\Big)$$

$$=\frac{1}{m^2n^2\Delta^4}\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}}\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1',\omega_2'}} Cov\Big(I_{g_1-m\Delta,g_2,m,n,\Delta}(\lambda_1,\lambda_2), I_{g_1'-m\Delta,g_2',m,n,\Delta}(\lambda_1',\lambda_2')\Big)$$

$$- Cov\Big(I_{g_1-m\Delta,g_2,m,n,\Delta}(\lambda_1,\lambda_2), I_{g_1',g_2',m,n,\Delta}(\lambda_1',\lambda_2')\Big)$$

$$- Cov\Big(I_{g_1,g_2,m,n,\Delta}(\lambda_1,\lambda_2), I_{g_1'-m\Delta,g_2',m,n,\Delta}(\lambda_1',\lambda_2')\Big)$$

$$+ Cov\Big(I_{g_1,g_2,m,n,\Delta}(\lambda_1,\lambda_2), I_{g_1',g_2',m,n,\Delta}(\lambda_1',\lambda_2')\Big)$$

$$=\frac{1}{m^2n^2\Delta^4}\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1,\omega_2}}\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\omega_1',\omega_2'}} A - B - C + D$$

$$=*,$$

Using the expression of the spatially varying periodogram in (3.1) and using linearity of covariance, along with the formula

$$Cov(AB,CD) = Cov(A,C)Cov(B,D) + Cov(A,D)Cov(B,C)$$

from Theorem 2.3.2 from Brillinger, 1981 we have

$$A = Cov\Big(I_{g_1-m\Delta,g_2,m,n,\Delta}(\lambda_1,\lambda_2), I_{g_1'-m\Delta,g_2',m,n,\Delta}(\lambda_1',\lambda_2')\Big)$$

$$=\frac{\Delta^4}{(2\pi)^4m^2n^2}\Big[\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'))$$

$$\exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2]) \times \sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'))$$

$$\exp(i\Delta[(p+p')\lambda_1+(q+q')\lambda_2]) + \sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'))$$

$$\exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\times\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1}C(g_1-g_1'+\Delta(p-p'),g_2-g_2'+\Delta(q-q'))$$

$$\exp(i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big]+cum_4$$

$$=\frac{\Delta^4}{(2\pi)^4m^2n^2}\Big[\Big|\sum_{p,q,p',q'}C(g_1-g_1'+\Delta(p-p'),g_2-g_2'+\Delta(q-q'))exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2])\Big|^2$$

$$+\Big|\sum_{p,q,p',q'}C(g_1-g_1'+\Delta(p-p'),g_2-g_2'+\Delta(q-q'))exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big|^2\Big]+cum_4,$$

$$B=Cov\Big(I_{g_1-m\Delta,g_2,m,n,\Delta}(\lambda_1,\lambda_2),I_{g_1',g_2',m,n,\Delta}(\lambda_1',\lambda_2')\Big)$$

$$=\frac{\Delta^4}{(2\pi)^4m^2n^2}\Big[\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1}C(g_1-g_1'+\Delta(p-p'-m),g_2-g_2'+\Delta(q-q'))$$

$$\exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2])\times\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1}C(g_1-g_1'+\Delta(p-p'-m),g_2-g_2'+\Delta(q-q'))$$

$$\exp(i\Delta[(p+p')\lambda_1+(q+q')\lambda_2])+\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1}C(g_1-g_1'+\Delta(p-p'-m),g_2-g_2'+\Delta(q-q'))$$

$$\exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\times\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1}C(g_1-g_1'+\Delta(p-p'-m),g_2-g_2'+\Delta(q-q'))$$

$$\exp(i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big]+cum_4$$

$$=\frac{\Delta^4}{(2\pi)^4m^2n^2}\Big[\Big|\sum_{p,q,p',q'}C(g_1-g_1'+\Delta(p-p'-m),g_2-g_2'+\Delta(q-q'))$$

$$exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2])\Big|^2$$

$$+\Big|\sum_{p,q,p',q'}C(g_1-g_1'+\Delta(p-p'-m),g_2-g_2'+\Delta(q-q'))exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big|^2\Big]+cum_4,$$

$$C=Cov\Big(I_{g_1,g_2,m,n,\Delta}(\lambda_1,\lambda_l^2),I_{g_1'-m\Delta,g_2',m,n,\Delta}(\lambda_1',\lambda_2')\Big)$$

$$=\frac{\Delta^4}{(2\pi)^4 m^2 n^2}\Big[\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} C(g_1-g_1'+\Delta(p-p'+m), g_2-g_2'+\Delta(q-q'))$$

$$\exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2]) \times \sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} C(g_1-g_1'+\Delta(p-p'+m), g_2-g_2'+\Delta(q-q'))$$

$$\exp(i\Delta[(p+p')\lambda_1+(q+q')\lambda_2]) + \sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} C(g_1-g_1'+\Delta(p-p'+m), g_2-g_2'+\Delta(q-q'))$$

$$\exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2]) \times \sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\sum_{p'=0}^{m-1}\sum_{q'=0}^{n-1} C(g_1-g_1'+\Delta(p-p'+m), g_2-g_2'+\Delta(q-q'))$$

$$\exp(i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big] + cum_4$$

$$=\frac{\Delta^4}{(2\pi)^4 m^2 n^2}\Big[\Big|\sum_{p,q,p',q'} C(g_1-g_1'+\Delta(p-p'+m), g_2-g_2'+\Delta(q-q'))exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2])\Big|$$

$$+\Big|\sum_{p,q,p',q'} C(g_1-g_1'+\Delta(p-p'+m), g_2-g_2'+\Delta(q-q'))exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big|^2\Big] + cum_4,$$

and $D = A$. $C(h_1, h_2) := Cov(Z(s_1, s_2), Z(s_1+h_1, s_2+h_2))$ Note that we assume $\lambda_1 = \lambda_1'$ and $\lambda_2 = \lambda_2'$ because the periodogram at different frequencies are asymptotically uncorrelated. The fourth order cumulants cum_4 are asymptotically negligible. Then for large m, n, we have

$$* = \frac{1}{m^2 n^2 \Delta^4}\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\min(\omega_1,\omega_1'),\min(\omega_2,\omega_2')}} A - B - C + D$$

$$=\frac{1}{(2\pi)^4 m^4 n^4}\sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\min(\omega_1,\omega_1'),\min(\omega_2,\omega_2')}}\Big[$$

$$2\Big|\sum_{p,q,p',q'} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'))exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2])\Big|^2$$

$$+2\Big|\sum_{p,q,p',q'} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'))exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big|^2$$

$$-\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p - p' - m), g_2 - g_2' + \Delta(q - q'))exp(-i\Delta[(p + p')\lambda_1 + (q + q')\lambda_2])\Big|^2$$

$$-\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p - p' - m), g_2 - g_2' + \Delta(q - q'))exp(-i\Delta[(p - p')\lambda_1 + (q - q')\lambda_2])\Big|^2$$

$$-\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p - p' + m), g_2 - g_2' + \Delta(q - q'))exp(-i\Delta[(p + p')\lambda_1 + (q + q')\lambda_2])\Big|^2$$

$$-\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p - p' + m), g_2 - g_2' + \Delta(q - q'))exp(-i\Delta[(p - p')\lambda_1 + (q - q')\lambda_2])\Big|^2\Big] \tag{A.3}$$

Note that each $|.|^2$ terms can be further decomposed. For example

$$\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p - p'), g_2 - g_2' + \Delta(q - q'))exp(-i\Delta[(p + p')\lambda_1 + (q + q')\lambda_2])\Big|^2$$

$$= \sum_{p,q,p',q'} C^2(g_1 - g_1' + \Delta(p - p'), g_2 - g_2' + \Delta(q - q')) + \sum_{(p,q,p',q') \neq (r,s,r's')}$$

$$C(g_1 - g_1' + \Delta(p - p'), g_2 - g_2' + \Delta(q - q'))C(g_1 - g_1' + \Delta(r - r'), g_2 - g_2' + \Delta(s - s'))$$

$$\exp(-i\Delta(p + p' - r' - r)\lambda_1)\exp(-i\Delta(q + q' - s' - s)\lambda_2)$$

In the second sum we must have either $p + p' \neq r + r'$ or $q + q' \neq s + s'$. Consider the set of terms where $p + p' \neq r + r'$ and $s + s' \neq q + q'$. For those terms we have

$$\sum_{p,q,h_1,h_2,r,s,h_3,h_4} C(g_1 - g_1' + \Delta h_1, g_2 - g_2' + \Delta h_2)C(g_1 - g_1' + \Delta h_3, g_2 - g_2' + \Delta h_4)$$

$$\times \exp(-i\Delta(2p - h_1 - 2r + h_3)\lambda_1)\exp(-i\Delta(2q - h_2 - 2s + h_4)\lambda_2)$$

Note that with our choice of λ_1 and λ_2 we have

$$\sum_{p,r} \exp(-i\Delta(2p - 2r)\lambda_1) = \Big|\sum_k \exp(-i2\Delta k\lambda_1)\Big|^2 = 0.$$

Thus this term does not have any contribution. Similarly, the terms where only one of $p+p' \neq r+r'$ and $s+s' \neq q+q'$ holds will also be zero.

All other terms can be simplified similarly. Further using our weak dependence condition we can argue that the last 4 terms will be asymptotically negligible as $m, n \to \infty$. Thus we get the final covariance expression

$$
\begin{aligned}
C^{h,h}_{\omega_1,\omega_2,\omega_1',\omega_2'}\big((g_1,g_2),(g_1',g_2')\big) =& 8\min(\omega_1,\omega_1'), \min(\omega_2,\omega_2') \\
& \times \int_0^\infty \int_0^\infty C^2(g_1-g_2'+h_1, g_2-g_2'+h_2)dh_1dh_2 \qquad \text{(A.4)}
\end{aligned}
$$

Similarly we have,

$$
\begin{aligned}
& Cov\Big(\hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(g_1,g_2), \hat{D}^v_{m,n,\Delta,\omega_1',\omega_2'}(g_1',g_2')\Big) \\
=& \frac{\Delta^4}{(2\pi)^4 m^4 n^4} \sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\min(\omega_1,\omega_1'),\min(\omega_2,\omega_2')}} \Big[\\
& 2\Big|\sum_{p,q,p',q'} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'))exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2])\Big|^2 \\
+& 2\Big|\sum_{p,q,p',q'} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'))exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big|^2 \\
-& \Big|\sum_{p,q,p',q'} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'-n))exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2])\Big|^2 \\
-& \Big|\sum_{p,q,p',q'} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'-n))exp(-i\Delta[(p-p')\lambda_1+(q-q')\lambda_2])\Big|^2 \\
-& \Big|\sum_{p,q,p',q'} C(g_1-g_1'+\Delta(p-p'), g_2-g_2'+\Delta(q-q'+n))exp(-i\Delta[(p+p')\lambda_1+(q+q')\lambda_2])\Big|^2
\end{aligned}
$$

$$-\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p-p'), g_2 - g_2' + \Delta(q-q'+n))exp(-i\Delta[(p-p')\lambda_1 + (q-q')\lambda_2])\Big|^2\Big]$$

$$\tag{A.5}$$

and

$$Cov\Big(\hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(g_1,g_2), \hat{D}^v_{m,n,\Delta,\omega_1',\omega_2'}(g_1',g_2')\Big)$$

$$=\frac{\Delta^4}{(2\pi)^4 m^4 n^4} \sum_{(\lambda_1,\lambda_2)\in\Lambda_{m,n,\Delta,\min(\omega_1,\omega_1'),\min(\omega_2,\omega_2')}} \Big[$$

$$2\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p-p'), g_2 - g_2' + \Delta(q-q'))exp(-i\Delta[(p+p')\lambda_1 + (q+q')\lambda_2])\Big|^2$$

$$+2\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p-p'), g_2 - g_2' + \Delta(q-q'))exp(-i\Delta[(p-p')\lambda_1 + (q-q')\lambda_2])\Big|^2$$

$$-\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p-p'+m), g_2 - g_2' + \Delta(q-q'-n))exp(-i\Delta[(p+p')\lambda_1 + (q+q')\lambda_2])\Big|^2$$

$$-\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p-p'-m), g_2 - g_2' + \Delta(q-q'-n))exp(-i\Delta[(p-p')\lambda_1 + (q-q')\lambda_2])\Big|^2$$

$$-\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p-p'), g_2 - g_2' + \Delta(q-q'+n))exp(-i\Delta[(p+p'-m)\lambda_1 + (q+q')\lambda_2])\Big|^2$$

$$-\Big|\sum_{p,q,p',q'} C(g_1 - g_1' + \Delta(p-p'+m), g_2 - g_2' + \Delta(q-q'+n))exp(-i\Delta[(p-p')\lambda_1 + (q-q')\lambda_2])\Big|^2\Big]$$

$$\tag{A.6}$$

Thus we have for $h_1, h_2 \in \{h, v\}$, $\omega_1, \omega_1', \omega_2, \omega_2' \in [0,1]$ and $(s_1, s_2), (s_1', s_2') \in S$, we have

$$C^{d_1,d_2}_{\omega_1,\omega_2,\omega_1',\omega_2'}\big((s_1,s_2),(s_1',s_2')\big) = 8\min(\omega_1,\omega_1'), \min(\omega_2,\omega_2')$$

$$\times \int_0^\infty \int_0^\infty C^2(s_1 - s_2' + h_1, s_2 - s_2' + h_2)dh_1 dh_2 \tag{A.7}$$

Remark A.1.1. *We note that*

- *In finite sample the two difference components are positively correlated, and as* $m, n \to \infty$ *this correlation converges to 1. This is only valid, when the underlying random field is stationary.*
- *The correlation structure is separable in* ω_1,ω_2 *and* (s_1, s_2).
- *The correlation between the statistics at different locations decreases as the locations gets further apart, i.e., as* $g_1 - g_1'$ *or* $g_2 - g_2'$ *increases.*
- *The correlation has a marginal Brownian Motion like structure in both* ω_1 *and* ω_2.

A.1.3 Higher Order Cumulants

First note that

$$\tilde{Z}_{g_1,g_2,m,n,\Delta}(\lambda_1,\lambda_2) = \frac{\Delta}{2\pi\sqrt{mn}} \sum_{p=0}^{m-1}\sum_{q=0}^{n-1} Z(g_1+\Delta p, g_2+\Delta q)\exp(-i\Delta(p\lambda_1+q\lambda_2)).$$

Then the joint cumulant is

$$cum(\tilde{Z}_{g_1,g_2,m,n,\Delta}(\lambda_1^1,\lambda_1^2),\ldots,\tilde{Z}_{g_1,g_2,m,n,\Delta}(\lambda_l^1,\lambda_l^2))$$

$$= \left(\frac{\Delta}{2\pi\sqrt{mn}}\right)^l cum\left(\sum_{p_1=0}^{m-1}\sum_{q_1=0}^{n-1} Z(g_1^1+\Delta p_1, g_1^2+\Delta q_1)\exp(-i\Delta(\lambda_1^1 p_1+\lambda_1^2 q_1)),\ldots,\right.$$

$$\left.\sum_{p_l=0}^{m-1}\sum_{q_l=0}^{n-1} Z(g_l^1+\Delta p_l, g_l^2+\Delta q_l)\exp(-i\Delta(\lambda_l^1 p_l+\lambda_l^2 q_l))\right)$$

$$= \left(\frac{\Delta}{2\pi\sqrt{mn}}\right)^l \sum_{p_1=0}^{m-1}\sum_{q_1=0}^{n-1}\cdots\sum_{p_l=0}^{m-1}\sum_{q_l=0}^{n-1} cum(Z(g_1^1+\Delta p_1, g_1^2+\Delta q_1)\exp(-i\Delta(\lambda_1^1 p_1+\lambda_1^2 q_1)),\ldots,$$

$$Z(g_l^1+\Delta p_l, g_l^2+\Delta q_l)\exp(-i\Delta(\lambda_l^1 p_l+\lambda_l^2 q_l)))$$

$$= \left(\frac{\Delta}{2\pi\sqrt{mn}}\right)^l \sum_{p_1=0}^{m-1}\sum_{q_1=0}^{n-1}\cdots\sum_{p_l=0}^{m-1}\sum_{q_l=0}^{n-1} cum(Z(g_1+\Delta p_1, g_2+\Delta q_1),\ldots,Z(g_1+\Delta p_l, g_2+\Delta q_l))\times$$

$$\exp(-i\Delta[\lambda_1^1 p_1+\cdots+\lambda_l^1 p_l+\lambda_1^2 q_1+\cdots+\lambda_l^2 q_l])$$

$$let\ p_m = p_l + u_m, q_m = q_l + v_m\ for\ m = 1,\ldots,l-1$$

$$= \left(\frac{\Delta}{2\pi\sqrt{mn}}\right)^l \sum_{u_1=-n+1}^{n-1}\sum_{v_1=-n+1}^{n-1}\cdots\sum_{u_{l-1}=-n+1}^{n-1}\sum_{v_{l-1}=-n+1}^{n-1}\sum_{p_l}\sum_{q_l}$$

$$cum(Z(g_1+\Delta(p_l+u_1), g_2+\Delta(q_l+v_1)),\ldots,$$

$$Z(g_1+\Delta(p_l+u_{l-1}), g_2+\Delta(q_l+v_{l-1})), Z(g_1+\Delta p_l, g_2+\Delta q_l))\times$$

$$\exp(-i\Delta[\lambda_1^1 u_1 + \cdots + \lambda_{l-1}^1 u_{l-1} + \lambda_1^2 v_1 + \cdots + \lambda_{l-1}^2 v_{l-1}]) \exp\left(-i\Delta\left[p_l \sum_{m=1}^{l} \lambda_m^1 + q_l \sum_{m=1}^{l} \lambda_m^2\right]\right)$$

$$= \left(\frac{\Delta}{2\pi\sqrt{mn}}\right)^l \sum_{u_1=-n+1}^{n-1} \sum_{v_1=-n+1}^{n-1} \cdots \sum_{u_{l-1}=-n+1}^{n-1} \sum_{v_{l-1}=-n+1}^{n-1}$$

$$cum(Z(g_1 + \Delta(p_l + u_1), g_2 + \Delta(q_l + v_1)), \ldots, Z(g_1 + \Delta(p_l + u_{l-1}), g_2 + \Delta(q_l + v_{l-1})),$$

$$Z(g_1 + \Delta p_l, g_2 + \Delta q_l)) \exp(-i\Delta[\lambda_1^1 u_1 + \cdots + \lambda_{l-1}^1 u_{l-1} + \lambda_1^2 v_1 + \cdots + \lambda_l^2 v_{l-1}]) \times$$

$$\sum_{p_l} \sum_{q_l} \exp\left(-i\Delta\left[p_l \sum_{m=1}^{l} \lambda_m^1 + q_l \sum_{m=1}^{l} \lambda_m^2\right]\right)$$

$$= \frac{\Delta}{2\pi\sqrt{mn}} f_{g_1,g_2}(\lambda_1^1, \lambda_1^2, \ldots, \lambda_{l-1}^1, \lambda_{l-1}^2) \sum_{p_l} \sum_{q_l} \exp\left(-i\Delta\left[p_l \sum_{m=1}^{l} \lambda_m^1 + q_l \sum_{m=1}^{l} \lambda_m^2\right]\right)$$

The last expression can be simplified as

$$\frac{\Delta}{2\pi\sqrt{mn}} f_{g_1,g_2}(\lambda_1^1, \lambda_1^2, \ldots, \lambda_{l-1}^1, \lambda_{l-1}^2) \boldsymbol{\Delta}\left(\sum_{m=1}^{l} \lambda_m^1\right) \boldsymbol{\Delta}\left(\sum_{m=1}^{l} \lambda_m^2\right) \tag{A.8}$$

where $\boldsymbol{\Delta}(\omega) = \sum_{k=1}^{n} \exp(-i\omega\Delta k)$. Note that $\boldsymbol{\Delta}(\omega)$ is 0 unless $\omega = 0 \mod 2\pi$. If ω is of the form $2\pi k$ then $\boldsymbol{\Delta}(\omega) = n$.

We then want to show for $d_1, \ldots, d_l \in \{h, v\}$

$$cum\left(\sqrt{mn}\Delta \hat{D}^{d_1}_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{g}^1), \ldots, \sqrt{mn}\Delta \hat{D}^{d_l}_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}(\mathbf{g}^l\right) = o(1).$$

Without loss of generality assume $d_1 = d_2 = \cdots = d_l = h$. We write

$$cum\left(\sqrt{mn}\Delta \hat{D}^{h}_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{g}^1), \ldots, \sqrt{mn}\Delta \hat{D}^{h}_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}(\mathbf{g}^l)\right)$$

$$= \left(\frac{1}{mn\Delta^2}\right)^{l/2} cum(\sum_{(\lambda_1^{(1)},\lambda_2^{(1)}) \in \Lambda_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}} I_{g_1^{(1)}-m\Delta, g_2^{(1)},m,n,\Delta}(\lambda_1^{(1)}, \lambda_2^{(1)}) - I_{g_1^{(1)},g_2^{(1)},m,n,\Delta}(\lambda_1^{(1)}, \lambda_2^{(1)}),$$

$$\ldots, \sum_{(\lambda_1^{(l)},\lambda_2^{(l)})\in\Lambda_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}} I_{g_1^{(l)}-m\Delta,g_2^{(l)},m,n,\Delta}(\lambda_1^{(l)},\lambda_2^{(l)}) - I_{g_1^{(l)},g_2^{(l)},m,n,\Delta}(\lambda_1^{(l)},\lambda_2^{(l)})\Big)$$

$$= \left(\frac{1}{mn\Delta^2}\right)^{l/2} \sum_{(\lambda_1^{(1)},\lambda_2^{(1)})\in\Lambda_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}} \cdots \sum_{(\lambda_1^{(l)},\lambda_2^{(l)})\in\Lambda_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}}$$

$$cum\left(I_{g_1^{(1)}-\Delta m,g_2^{(1)},m,n,\Delta}(\lambda_1^{(1)},\lambda_2^{(1)}) - I_{g_1^{(1)},g_2^{(1)},m,n,\Delta}(\lambda_1^{(1)},\lambda_2^{(1)}),\ldots, I_{g_1^{(l)},g_1^{(l)}}(\lambda_1^{(l)},\lambda_2^{(l)}) - I_{w_{s_l^1,s_l^2,n}}(\lambda_1^{(l)},\lambda_2^{(l)})\right)$$

$$= \left(\frac{1}{mn\Delta^2}\right)^{l/2} \sum_{(\lambda_1^{(1)},\lambda_2^{(1)})\in\Lambda_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}} \cdots \sum_{(\lambda_1^{(l)},\lambda_2^{(l)})\in\Lambda_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}} \sum_{i=1}^{2^l} cum(B_1^i,\ldots,B_l^i)$$

where B_j^i is either $I_{g_1^{(j)}-\Delta m,g_2^{(j)},m,n,\Delta}(\lambda_1^{(j)},\lambda_2^{(j)})$ or $-I_{g_1^{(j)},g_2^{(j)},m,n,\Delta}(\lambda_1^{(j)},\lambda_2^{(j)})$.

Since 2^l is a finite number, we only need to show

$$\left(\frac{1}{mn\Delta^2}\right)^{l/2} \sum_{(\lambda_1^{(1)},\lambda_2^{(1)})\in\Lambda_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}} \cdots \sum_{(\lambda_1^{(l)},\lambda_2^{(l)})\in\Lambda_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}} cum(B_1,\ldots,B_l) = o(1),$$

where B_j is either $I_{g_1^{(j)}-\Delta m,g_2^{(j)},m,n,\Delta}(\lambda_1^{(j)},\lambda_2^{(j)})$ or $-I_{g_1^{(j)},g_2^{(j)},m,n,\Delta}(\lambda_1^{(j)},\lambda_2^{(j)})$.

Without loss of generality, we will show

$$\left(\frac{1}{mn\Delta^2}\right)^{l/2} \sum_{(\lambda_1^{(1)},\lambda_2^{(1)})\in\Lambda_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}} \cdots \sum_{(\lambda_1^{(l)},\lambda_2^{(l)})\in\Lambda_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}}$$

$$cum(I_{g_1^{(1)}-\Delta m,g_2^{(1)}}(\lambda_1^{(1)},\lambda_2^{(1)}),\ldots, I_{g_1^{(l)}-\Delta m,g_2^{(l)}}(\lambda_1^{(l)},\lambda_2^{(l)})) = o(1).$$

Recall that

$$I_{g_1^{(j)}-\Delta m,g_2^{(j)},m,n,\Delta}(\lambda_1^{(j)},\lambda_2^{(j)}) = \tilde{Z}_{g_1^{(j)}-\Delta m,g_2^{(j)},m,n,\Delta}(\lambda_1^{(j)},\lambda_2^{(j)})\tilde{Z}_{g_1^{(j)}-\Delta m,g_2^{(j)},m,n,\Delta}(-\lambda_1^{(j)},-\lambda_2^{(j)}).$$

Using Theorem 2.3.2 in Brillinger (1981), we have

$$cum(I_{g_1^{(1)}-\Delta m,g_2^{(1)}}(\lambda_1^{(1)},\lambda_2^{(1)}),\ldots, I_{g_1^{(l)}-\Delta m,g_2^{(l)}}(\lambda_1^{(l)},\lambda_2^{(l)}))$$

$$= \sum_{\nu} cum(Y_{\alpha\beta} : \alpha\beta \in \nu_1) \cdots cum(Y_{\alpha\beta} : \alpha\beta \in \nu_p),$$

where $Y_{\alpha 1} = \tilde{Z}_{g_1^{(j)}-\Delta m, g_2^{(j)}, m,n,\Delta}(\lambda_1^{(j)}, \lambda_2^{(j)})$, $Y_{\alpha 2} = \tilde{Z}_{g_1^{(j)}-\Delta m, g_2^{(j)}, m,n,\Delta}(-\lambda_1^{(j)}, -\lambda_2^{(j)})$, and the sum is over all indecomposable partitions $v = v_1 \cup v_2 \cup \cdots \cup v_p$ of

$$\begin{array}{cc} (1,1) & (1,2) \\ (2,1) & (2,2) \\ \vdots & \vdots \\ (l,1) & (l,2) \end{array}$$

Then

$$\left(\frac{1}{mn\Delta^2}\right)^{l/2} \sum_{(\lambda_1^{(1)},\lambda_2^{(1)}) \in \Lambda_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}} \cdots \sum_{(\lambda_1^{(l)},\lambda_2^{(l)}) \in \Lambda_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}}$$

$$cum(I_{g_1^{(1)}-\Delta m, g_2^{(1)}}(\lambda_1^{(1)}, \lambda_2^{(1)}), \ldots, I_{g_1^{(l)}-\Delta m, g_2^{(l)}}(\lambda_1^{(l)}, \lambda_2^{(l)}))$$

$$= \left(\frac{1}{mn\Delta^2}\right)^{l/2} \sum_{(\lambda_1^{(1)},\lambda_2^{(1)}) \in \Lambda_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}} \cdots \sum_{(\lambda_1^{(l)},\lambda_2^{(l)}) \in \Lambda_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}} \sum_{\nu}$$

$$cum(Y_{\alpha\beta} : \alpha\beta \in \nu_1) \cdots cum(Y_{\alpha\beta} : \alpha\beta \in \nu_p)$$

As the underlying random field is Gaussian, all the cumulants of order 3 or higher is zero. So the only term that have a positive contribution is where all $\nu_1, \nu_2, \ldots, \nu_p$ contains two elements, which are mutual conjugate (i.e., (λ_1, λ_2) has to appear with $(-\lambda_1, -\lambda_2)$. Given we are only considering indecomposable partition, this can only happen if $\lambda_1^{(1)} = \lambda_1^{(2)} = \cdots = \lambda_1^{(l)}$ and $\lambda_2^{(1)} = \lambda_2^{(2)} = \cdots = \lambda_2^{(l)}$, thus we can ignore the later $l-1$ summations over $\lambda_j^{(k)}$'s. Using (A.8), each such partition will contribute $f_{g_1,g_2}(\lambda_1^{(1)}, \lambda_2^{(1)})(mn)$. Thus last expression boils down to

$$(2l)!(mn)^{1-l/2} f^l_{g_1,g_2}(\lambda_1^{(1)}, \lambda_2^{(1)}).$$

This gives us

$$cum\left(\sqrt{mn}\Delta \hat{D}^{d_1}_{m,n,\Delta,\omega_1^{(1)},\omega_2^{(1)}}(\mathbf{g}^1), \ldots, \sqrt{mn}\Delta \hat{D}^{d_l}_{m,n,\Delta,\omega_1^{(l)},\omega_2^{(l)}}(\mathbf{g}^l\right) = (2l)!2^l(mn)^{1-l/2}C^l, \tag{A.9}$$

for some constant C and this proves our assertion.

A.1.4 Proof of (3.10)

To show (3.10), it is enough to show (3.16) and (3.17). The proof of (3.16) is elementary. To show (3.17) note that by Theorem 2.4 in Dahlhaus, 1988 it is enough to show that

$$\begin{aligned}
&\left|cum_l\left(\sqrt{mn}\Delta\left(A_{m,n,\Delta}(\omega_1^{(1)},\omega_2^{(1)},s_1^{(1)},s_2^{(1)}), A_{m,n,\Delta}(\omega_1^{(2)},\omega_2^{(2)},s_1^{(2)},s_2^{(2)})\right)\right)\right| \\
&\le (2l)!C^l d^l_\beta((\omega_1^{(1)},\omega_2^{(1)},s_1^{(1)},s_2^{(1)}),(\omega_1^{(2)},\omega_2^{(2)},s_1^{(2)},s_2^{(2)}))
\end{aligned}$$

Arguments similar to the cumulant calculations can be reproduced to show that

$$\begin{aligned}
&\left|cum_l\left(\sqrt{mn}\Delta\left(A_{m,n,\Delta}(\omega_1^{(1)},\omega_2^{(1)},s_1^{(1)},s_2^{(1)}), A_{m,n,\Delta}(\omega_1^{(2)},\omega_2^{(2)},s_1^{(2)},s_2^{(2)})\right)\right)\right| \\
&\le (2l)!2^lC^l(log(mn))^{l-1}(mn)^{1-l/2} \\
&\le (2l)!2^lC^l(mn)^{-l(1/2-1/3-\epsilon}
\end{aligned}$$

for any $\epsilon > 0$. Choose $\epsilon = 1/6 - \beta/2$. Now the proof is done observing that $\frac{1}{mn} \le C\|(\omega_1^{(1)},\omega_2^{(1)},s_1^{(1)},s_2^{(1)}) - (\omega_1^{(2)},\omega_2^{(2)},s_1^{(2)},s_2^{(2)})\|$ as long as $(\omega_1^{(j)},\omega_2^{(j)},s_1^{(j)},s_2^{(j)})$ for $j = 1, 2$ are distict points in $\mathcal{P}$.

A.2 Behavior of the Spectral Difference Process under Non-stationarity

When the underlying field is not-stationary we assume a piece-wise stationary structure as in (2.10) with

$$\int_{[0,\infty)^2} \|h\| C_{ij}(\mathbf{h}) d\mathbf{h} < \infty, \quad i, j = 1, 2, \ldots, K. \tag{A.10}$$

Under this assumption a calculation similar to subsection A.1.1 gives us

$$\mathbf{E}\left(\hat{D}^h_{m,n,\Delta,\omega_1,\omega_2}\right) = D^h_{m,n,\Delta,\omega_1,\omega_2}$$

$$\mathbf{E}\left(\hat{D}^v_{m,n,\Delta,\omega_1,\omega_2}\right) = D^v_{m,n,\Delta,\omega_1,\omega_2}$$

To this end we state the following result showing consistency of the spectral difference process.

Lemma A.2.1. *Under the assumptions of Theorem 3.4.1 we have*

$$\sup_{(\omega_1,\omega_2)\in[0,1]^2, \mathbf{s}\in S} \sqrt{mn}\Delta \left\| \begin{array}{c} \hat{D}^h_{m,n,\Delta,\omega_1,\omega_2} - D^h_{m,n,\Delta,\omega_1,\omega_2} \\ \hat{D}^v_{m,n,\Delta,\omega_1,\omega_2} - D^v_{m,n,\Delta,\omega_1,\omega_2} \end{array} \right\| = O_p(1)$$

The proof of this result is similar to the proof of Theorem 3.2.1 with some additional notation.

A.3 Consistency of the Partition Selection

Let G be the grid on which we observe the field, i.e., $G = \{0, \Delta, \ldots, M\Delta\} \times \{0, \Delta, \ldots, N\Delta\}$. Furthermore define $X_h := \{x_1, x_2, \ldots, x_H\}$ be the set of all X-coordinates of horizontal boundaries and $Y_v := \{y_1, y_2, \ldots, y_V\}$ be the set of all Y-coordinates of vertical boundaries.

Theorem A.3.1. *Assume that the assumptions of Theorem 3.2.1 holds. Furthermore, for all $\omega_1, \omega_2 \in (0, 1)$ and $\gamma \in (0, 0.5]$ let $\epsilon_{m,n,\omega_1,\omega_2}$ be a sequence such that $\epsilon_{m,n,\omega_1,\omega_2} = o(N^\gamma)$ and $\liminf_{m,n\to\infty} \inf_{\omega_1,\omega_2\in[0,1]} \epsilon_{m,n,\omega_1,\omega_2} = C > 0$ for some constant C. Then the detection rule describes in Algorithm 2 is accurate in the following sense:*

(i) Let $\tilde{I}_h = G - X_h \times \{0, \Delta, \ldots, N\Delta\}$. Then the probability

$$\mathbb{P}\left(\bigcup_{\mathbf{s}\in\tilde{I}_h} (mn)^\gamma \sup_{(\omega_1,\omega_2)} |\widehat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s})| > \epsilon_{m,n,\omega_1,\omega_2} \right) \to 0$$

as $m, n, M, N \to \infty$.

(ii) Let $\tilde{I}_v = G - \{0, \Delta, \ldots, M\Delta\} \times Y_h$. Then the probability

$$\mathbb{P}\left(\bigcup_{\mathbf{s}\in\tilde{I}_v} (mn)^\gamma \sup_{(\omega_1,\omega_2)} |\widehat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s})| > \epsilon_{m,n,\omega_1,\omega_2} \right) \to 0$$

as $m, n, M, N \to \infty$

(iii) The probability that the procedure detects all coordinates breaks converges to one, that is,

$$\mathbb{P}\left(\bigcap_{X_h\times\{0,\Delta,\ldots,N\Delta\}} (mn)^\gamma \sup_{(\omega_1,\omega_2)} |\widehat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s})| > \epsilon_{m,n,\omega_1,\omega_2} \right) \to 1$$

and

$$\mathbb{P}\left(\bigcap_{\{0,\Delta,\ldots,M\Delta\}\times Y_v} (mn)^\gamma \sup_{(\omega_1,\omega_2)} |\widehat{D}^v_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s})| > \epsilon_{m,n,\omega_1,\omega_2} \right) \to 1$$

as $m, n, M, N \to \infty$.

Proof. For proof of (i) note that $\sup_{(\omega_1,\omega_2)\in[0,1]^2} |D^h_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s})| = 0$ on $\tilde{I}_h$, where $D^h_{m,n,\Delta,\omega_1,\omega_2}$ is as defined in (3.5). By consistency of the estimator we have

$$\sup_{(\omega_1,\omega_2)} \sup_{\mathbf{s}\in S} (mn)^\gamma |\widehat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s}) - D^h_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s})| = o(1).$$

Thus we have

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{\mathbf{s}\in\tilde{I}_h} (mn)^\gamma \sup_{(\omega_1,\omega_2)} |\widehat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s})| > \epsilon_{m,n,\omega_1,\omega_2}\right)\\
\leq &\mathbb{P}\left((mn)^\gamma \sup_{(\omega_1,\omega_2)} \sup_{\mathbf{s}\in S} |\widehat{D}^h_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s}) - D^h_{m,n,\Delta,\omega_1,\omega_2}(\mathbf{s})| > C/2\right) \to 0.
\end{aligned}$$

The proof of (ii) is analogous. Part (iii) follows from A.2.1.

□

Bibliography

Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association*, *93*(444), 1488–1501. https://doi.org/10.1080/01621459.1998.10473808

Anselin, L. (2001). Spatial econometrics. *A companion to theoretical econometrics*.

Bandyopadhyay, S., Jentsch, C., & Rao, S. (2016). A spectral domain test for stationarity of spatio-temporal data. *Journal of Time Series Analysis*, *38*. https://doi.org/10.1111/jtsa.12222

Bandyopadhyay, S., & Rao, S. S. (2017). A test for stationarity for irregularly spaced spatial data. *Journal of the Royal Statistical Society Series B*, *79*(1), 95–123. https://ideas.repec.org/a/bla/jorssb/v79y2017i1p95-123.html

Bose, S., & Steinhardt, A. (1996). Invariant tests for spatial stationarity using covariance structure. *IEEE Transactions on Signal Processing*, *44*(6), 1523–1533. https://doi.org/10.1109/78.506617

Brillinger, D. R. (1981). *Time series data analysis and theory*. Society for Industrial; Applied Mathematics SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104.

Cong, L., Ding, S., Wang, L., Zhang, A., & Jia, W. (2018). Image segmentation algorithm based on superpixel clustering. *IET Image Processing*, *12*(11), 2030–2035.

Dahlhaus, R. (1988). Empirical spectral processes and their applications to time series analysis. *Stochastic Processes and their Applications*, *30*(1), 69–83.

Ephraty, A., Tabrikian, J., & Messer, H. (1996). A test for spatial stationarity and applications. *Proceedings of 8th Workshop on Statistical Signal and Array Processing*, 412–415. https://doi.org/10.1109/SSAP.1996.534903

Fan, J. (1996). Test of significance based on wavelet thresholding and neyman's truncation. *Journal of the American Statistical Association*, *91*(434), 674–688.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, *89*(1), 197–210. https://doi.org/10.1093/biomet/89.1.197

Fuentes, M. (2005). A formal test for nonstationarity of spatial stochastic processes. *Journal of Multivariate Analysis*, *96*(1), 30–54. https://doi.org/10.1016/j.jmva.2004.09.003

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, *102*, 321–331. https://doi.org/10.1198/016214506000000852

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2015). Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics*, *14*, 505–531. https://doi.org/https://doi.org/10.1016/j.spasta.2015.10.001

Gramacy, R. B., & Lee, H. K. H. (2008). Bayesian treed gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, *103*(483), 1119–1130.

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, *69*(1), 95–105.

Guyon, X. (1995). *Random fields on a network : Modeling, statistics, and applications*. Springer-Verlag.

Hersbach, H., Bell, B., Berrisford, P., Biavati, A., G. andHoranyi, Munoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thepaut, J.-N. (2018). Era5 hourly data on pressure levels from 1979 to present. copernicus climate change service (c3s). *Climate Data Store (CDS)*.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*, *5*(2), 173–190.

Holland, D. M., Cox, W. M., Scheffe, R., Cimorelli, A. J., Nychka, D., & Hopke, P. K. (2003). Spatial prediction of air quality data. *EM-PITTSBURGH-AIR AND WASTE MANAGEMENT ASSOCIATION-*, 31–35.

Jun, M., & Genton, M. G. (2012). A test for stationarity of spatio-temporal random fields on planar and spherical domains. *Statistica Sinica*. https://doi.org/10.5705/ss.2010.251

Kim, H.-M., Mallick, B. K., & Holmes, C. C. (2005). Analyzing nonstationary spatial data using piecewise gaussian processes. *Journal of the American Statistical Association*, *100*(470), 653–668.

Kurisu, D. (2022). Nonparametric regression for locally stationary random fields under stochastic sampling design. *Bernoulli*, *28*(2), 1250–1275.

Magnussen, S. (1993). Bias in genetic variance estimates due to spatial autocorrelation. *Theoretical and Applied Genetics*, *86*(2), 349–355.

Masotti, M., Zhang, L., Leng, E., Metzger, G. J., & Koopmeiners, J. S. (2021). A novel bayesian functional spatial partitioning method with application to prostate cancer lesion detection using mri. *Biometrics*.

Matsuda, Y., & Yajima, Y. (2009). Fourier analysis of irregularly spaced data on rd. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(1), 191–217.

Mets, K. D., Armenteras, D., & Davalos, L. M. (2017). Spatial autocorrelation reduces model precision and predictive power in deforestation analyses. *Ecosphere*, *8*(5), e01824. https://doi.org/https://doi.org/10.1002/ecs2.1824

Mohajer, M., Englmeier, K., & Schmid, V. (2010). *A comparison of gap statistic definitions with and with-out logarithm function* (tech. rep.). Technical Report.

Morris, L. (2021). *Spatio-temporal modelling for non-stationary point referenced data* (Doctoral dissertation). https://doi.org/10.26686/wgtn.15147396.v1

Nychka, D., Wikle, C., & Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, *2*(4), 315–331.

Paciorek, C., & Schervish, M. (2003). Nonstationary covariance functions for gaussian process regression. *Advances in neural information processing systems*, *16*.

Paciorek, C. J., & Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, *17*(5), 483–506. https://doi.org/https://doi.org/10.1002/env.785

Pope, C., Gosling, J., Barber, S., Johnson, J., Yamaguchi, T., Feingold, G., & Blackwell, P. (2021). Gaussian process modeling of heterogeneity and discontinuities using voronoi tessellations.

Preuss, P., Puchstein, R., & Dette, H. (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association*, *110*(510), 654–668.

Rao, S. S. (2018). Statistical inference for spatial statistics defined in the Fourier domain. *The Annals of Statistics*, *46*(2), 469–499. https://doi.org/10.1214/17-AOS1556

Ripley, B. D. (2005). *Spatial statistics* (1st ed.). Wiley.

Rollinson, C. R., Finley, A. O., Alexander, M. R., Banerjee, S., Dixon Hamil, K.-A., Koenig, L. E., Locke, D. H., DeMarche, M. L., Tingley, M. W., Wheeler, K., Youngflesh, C., & Zipkin, E. F. (2021). Working across space and time: Nonstationarity in ecological research and application. *Frontiers in Ecology and the Environment*, *19*(1), 66–72. https://doi.org/https://doi.org/10.1002/fee.2298

Rosenblatt, M. (2012). *Stationary sequences and random fields*. Springer Science & Business Media.

Sampson, P. D., & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, *87*(417), 108–119.

Stein, M. (1995). Fixed-domain asymptotics for spatial periodograms. *Journal of the American Statistical Association*, *90*(432), 1277–1288.

Stein, M. (2015). *Interpolation of spatial data: Some theory for kriging*. https://doi.org/10.1007/978-1-4612-1494-6

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423.

Vaart, A. W., & Wellner, J. A. (1996). Weak convergence. *Weak convergence and empirical processes* (pp. 16–28). Springer.

Whittle, P. (1954). ON STATIONARY PROCESSES IN THE PLANE. *Biometrika*, *41*(3-4), 434–449. https://doi.org/10.1093/biomet/41.3-4.434

Curriculum Vitae

Academic Experience

George Mason University 2018.8-2022.8
Ph.D. in Statistical Science
Binghamton University, The State University of New York 2016.8-2018.5
Ph.D. in Mathematical Science (drop out)
Nankai University 2012.8-2016.5
Bachelor of Economics

Industry Experience

Eli Lilly and Company 2021.6-2021.8
Research Scientist Intern
Biogen 2020.6-2020.8
Biostatistician Intern
Tianhong Fund Management 2016.1-2016.5
Business Development Intern

Publications

Supply-Side Predictors of Fatal Drug Overdose in the Washington/Baltimore HIDTA Region: 2016-2020, Evan Lowder, Weiyu Zhou, Lora Peppard, Rebecca Bates, Thomas Carr, *International Journal of Drug Policy* (in submission)
E-cigarette use during pregnancy and its association with adverse birth outcomes, Hong Xue, Shuo-yu Lin, Liang Wang, Weiyu Zhou, Panagiota Kitsantas, Xiaozhong Wen, *Preventive Medicine* (in submission)
The Association Between Tobacco Retailer Outlet Density and Prevalence of Cigarette Smoking in Virginia, Hong Xue, Shuo-yu Lin, Liang Wang, Weiyu Zhou, Panagiota Kitsantas, Xiaozhong Wen, *Nicotine and Tobacco Research* (2022)
Impact of Generic Competition on Statin Utilization and Expenditures: An Analysis of 2002-2018 Medical Expenditure Panel Survey Data, Hong Xue, Shuo-yu Lin, Liang Wang, Weiyu Zhou, Panagiota Kitsantas, Xiaozhong Wen, *JAMA Network Open* (2021)