CONTRIBUTIONS TO SPECTRAL SPATIAL STATISTICS

Rosa María Crujeiras Casais
Departamento de Estatística e Investigación Operativa
Universidade de Santiago de Compostela

PhD Dissertation

Todos os coxos
teñen unha colección
de caxatos
nos, como somos aviadores,
xuntamos volvoretas.

Ronseltz

Don Wenceslao González-Manteiga, catedrático do Departamento de Estatística e Investigación Operativa da Universidade de Santiago de Compostela e Don Rubén Fernández Casal, profesor axudante do Departamento de Matemáticas da Universidade da Coruña, informan que a memoria titulada:

CONTRIBUTIONS TO SPECTRAL SPATIAL STATISTICS

foi realizada baixo a súa dirección por Dona Rosa María Crujeiras Casais, estimando que a interesada se atopa en condicións de optar ao grao de Doutor en Ciencias Matemáticas, polo que solicitan que sexa admitida a trámite para a súa lectura e defensa pública.

En Santiago de Compostela, a 16 de Outubro de 2006.

Os directores:

Prof. Dr. Wenceslao González Manteiga

Prof. Dr. Rubén Fernández Casal

A doutoranda: Rosa María Crujeiras Casais

Contents

Preface

Spatial statistics is one of the most important methodologies for a large diversity of disciplines, such as ecology, hidrology, environmental sciences, etc. In all these fields, experiments involve spatial-referred data. When collecting data, that is, when measuring a certain variable at different locations, specialist on these areas have the intuition that places close to one another tend to have similar values, whereas ones that are farther apart differ more. Then, it seems obvious that such samples can not be treated as independent. General references on the spatial statistics topic, which collect many practical examples, are Cressie (1993), Chilès and Delfiner (1999) (for the geostatistics context) or Schabenberger and Gotway (2005).

The design of spatial models for representing the variability of a spatial process is one of the fundamental issues in spatial statistics. The variability in the model can be due to two different sources. On one hand, the "small-scale" variability (i.e. dependence) and, on the other hand, the "large-scale" variability (i.e. trend). The large-scale variability has been traditionally modelled by linear regression, nonparametric regression methods (e.g. Lefohn and Shadwick (1991)) or generalized additive models (e.g. Holland *et al.* (2000)), in the spatial and spatio-temporal contexts.

Our interest is focused on the dependence structure of the process. Modelling the dependence structure is a crucial task in spatial statistics, and particularly, in the geostatistical context.

Geostatistics refers to continuous spatial processes, for instance, phosphorus content on the soil or ozone concentration in the atmosphere. Measurements of such quantities may be taken at any location. But measurements are not taken at *all* locations, and prediction is one of the main objectives of a geostatistical analysis. In this context, prediction differs from classical estimation because it relies on spatial models: geostatistical prediction involves the dependence structure of the process. For that reason, much effort has been devoted to describe the behaviour of the dependence structure, above all, under stationary assumptions, both from parametric and nonparametric approaches. Nonetheless, there's still a shortage on techniques for assessing the goodness-of-fit of such estimators.

This is the main purpose of this dissertation: to propose goodness-of-fit testing techniques that allow for checking the validity of a certain model for the dependence structure of a stationary spatial process.

We set our research context in the spectral domain, so the dependence structure is modelled by the spectral density. In our way towards the construction of a goodness-of-fit test for spatial dependence models, we had to confront other related issues, which form now part of this manuscript. First, we explore the estimation problem in the frequency domain, just to highlight some features on the estimation of the spectral density. Then, when trying to apply our goodness-of-fit test in practice, we design a method for simulating spatial process which takes advantages of the spectral domain features. We finish our work proposing a test for comparing spatial spectral densities.

This manuscript is organized as follows:

Chapter 1. Spatial statistics and spectral methods. In this chapter we make a brief overview of the different situations where the scientist confront the treatment of spatially dependent data (e.g. Cressie (1993), Chilès and Delfiner (1999)). In this spatial context, we focus on the geostatistical case, and particularly, on the problem of the dependence structure modelling and kriging interpolation. Besides, we also include a section on spectral representation of random fields (see Grenander (1981) or Yaglom and Yaglom (1987)), which will serve as the basis for the posterior theoretical developments. We also introduce the datasets that we will consider for illustration purposes along the manuscript.

Chapter 2. Spectral techniques for modeling spatial dependence. We introduce in this chapter the concept of spatial periodogram, as a nonparametric estimator for the spatial spectral density. A brief overview on spatial spectral density estimation (from parametric and non-parametric approaches) is also provided. Some considerations on two modified-periodogram estimators of the spatial spectral density are given.

Chapter 3. Simulation of spatial dependence structures. In this chapter, we revise the Fourier Integral Method for simulating stationary random fields (see Pardo-Igúzquiza and Chica-Olmo (1993) and Chilès and Delfiner (1999)) and provide a modification of this method, which shows a better performance. We also include some simulation results for discrete and continuous spatial process.

Chapter 4. Goodness-of-fit tests for the spatial spectral density. The main objective of our work is to propose goodness-of-fit testing techniques for the dependence structure of a spatial random process, in our case, represented through the spatial spectral density. Two goodness-of-fit tests are provided, the first one based on the periodogram (in a analogous way to the test proposed in Paparoditis (2000) for the one-dimensional spectral density) and the other test based on the log-periodogram (similar to Fan and Zhang (2004), for the time series case). The performance of these tests is illustrated by a simulation study and real data application. In the appendix of this chapter, we include the theoretical details.

Chapter 5. Comparison of dependence structures. This last chapter is devoted to a testing technique for comparing two or more spatial spectral densities. Equivalently, we provide a test for checking whether the dependence structure of a collection of sets of observations exhibit the same pattern of dependence. The test is based on an L_2 distance, similar to the test in Vilar-Fernández and González-Manteiga (2004), for comparing regression curves. We also provide some simulation results and application to real data. The theoretical developments are included in the appendix.

Finally, the last part of the manuscript is devoted to the discussion on some future research lines and open problems. We also enclose a summary of this dissertation thesis in Galician language.

I would like to thank my advisors, Prof. Wenceslao González-Manteiga and Prof. Rubén Fernández-Casal for their work and support during these years.

This work has been supported by the Ministry of Education and Science and FEDER, projects BFM2002-03213, MTM2005-0020 and grant BES2003-0581. Also Xunta de Galicia Project PGIDIT06PXIB207

Chapter 1

Spatial statistics and spectral methods.

Contents

1.1 Spatial data and spatial models.

When talking about spatial data, we refer to those random variables whose observations are associated to locations in the d-dimensional Euclidean space $\mathbb{R}^d$ ($d = 2$ for the plane, or $d = 3$ for space-time, for instance). Within this context we may find different relations between the observed data and the reference location. Denote by $\{Z(\mathbf{s}), \mathbf{s} \in D\}$, with $D \subset \mathbb{R}^d$, a spatial random process. We may distinguish three different situations (e.g. Cressie (1993), Sections 1.1 and 1.2):

- **Geostatistics.** The non-stochastic observation region $D \subset \mathbb{R}^d$ contains a d-dimensional rectangle with positive volume and $Z(\mathbf{s})$ is a random vector varying continuously over D. That is, between any two sample locations $\mathbf{s}_i$ and $\mathbf{s}_j$, $\mathbf{s}_i \neq \mathbf{s}_j$, an infinite number of samples could be placed. Data with continuous variation are referred to as geostatistical data. Examples of this kind of spatial data have been commented in the Preface (phosphorus content on the soil, ozone concentration in the atmosphere, etc.).

- **Lattice data.** The non-stochastic observation region D is a fixed collection of countable many points (regularly or irregularly spaced) on $\mathbb{R}^d$ and $Z(\mathbf{s})$ is a random vector at $\mathbf{s} \in D$. In this case, the domain D is non-random and countable. Some examples of lattice data are those collected by ZIP code, for instance, and they are quite common in epidemiology studies. Spatial epidemiology, which concerns the analysis of the spatial distribution of the incidence of a disease, has become a major research topic (e.g. Lawson (2006)). In many cases, sites (spatial locations) represent areal regions. Offently, we must assign a spatial coordinate to each site, as for instance, the centroid of the region. See Cressie (1993), Part II.

- **Point processes.** $Z(\mathbf{s})$ is a random vector at $\mathbf{s} \in D$, but the locations $\mathbf{s}$ are randomly distributed over $D \subset \mathbb{R}^d$. A classical example of such a process is that given by a measure taken on trees in a forest. See Cressie (1993) (Part III), Diggle (2003) or Stoyan *et al.* (1995), for example.

In some cases, the difference may not be clear and geostatistical techniques have been applied for analyzing point processes or lattice data. In any case, when confronting a practical spatial data case, it is usual to begin with a exploratory geostatistical analysis. However, the derived conclussions must be taken carefully under consideration.

In these three situations, there exist different features and objectives. For instance, in the geostatistical context, prediction is an important goal. However, it is not a key point for lattice data analysis, since in this setting, information is usually exhaustive. Nevertheless, modeling

dependence structure is a common goal both in geostatistical, lattice or point process context. In fact, for solving the prediction problem in the geostatistical context, it is crucial to obtain an adequate model for the dependence structure of the data, as we will see later.

We will provide a brief overview on these three topics, just to remark the characteristic features in each case. We will also give some practical examples both for the lattice data and point process cases. Finally, we will focus our attention on the geostatistical context.

1.1.1 Lattice data.

Lattice data (also called regional data or areal data) can be considered the coarse of the three types of spatial data introduced above. Lattice data can be obtained by integration (accumulation) from geostatistical or point-process data. The spatial locations associated with a lattice data process are called *sites*, and they are denoted by $\mathbf{s}_i$, as in the geostatistical case. Then, the observations from a lattice data process are usually denoted by $Z(\mathbf{s}_i)$, but this notation may be misleading. It is not suprising to find lattice data studied in the same way as data from a continuous (geostatistical) process.

When analyzing a lattice structure, a key point is the description of spatial connectivity. Connectivity is expressed in terms of distances between representative points. This quantity takes value zero if the two points are not connected and takes value one if there exists a spatial connection between sites. Therefore, a measure of spatial correlation is needed, in order to define connectness between sites.

The spatial autocorrelation (or correlation) refers to the correlation between $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)$, that is, the correlation between the same variable observed in two different locations. In order to measure this spatial autocorrelation, there exist three classical statistics: the Cross-Product statistic, Moran's I and Geary's c. These three statistics are particular cases of the Mantel's test, introduced by Mantel in 1967.

Denote by $Z_i = Z(\mathbf{s}_i)$ and let Y_{ij} be a measure of the similarity between the response in the locations i and j (for instance, the squared difference between the observations). Let W_{ij} be a measure of the spatial proximity between locations i and j, for instance, $W_{ij} = 1$ if the regions corresponding to sites i and j are neighbours and $W_{ij} = 0$ otherwise. The Cross-Product statistic is defined as:

$$C = \sum_i \sum_j W_{ij} Y_{ij}. \tag{1.1}$$

For $Y_{ij} = (Z_i - Z_j)^2$, low values of C indicate positive spatial correlation (observations taken in locations which are near tend to be similar) and high values of C indicate negative correlation. In order to judge whether an observed value of C is low or high, there exist different approaches based on Monte Carlo simulations, randomization tests or asymptotic Normal approximation, which allow for calibrating the distribution of C.

In the fifties, Moran proposed a statistic which has been widely used in many practical context, and which is stil studied and discussed (see Li *et al.* (2005)). Moran's I statistic is given by:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j W_{ij}(Z_i - \bar{Z})(Z_j - \bar{Z})}{\sum_i (Z_i - \bar{Z})^2}, \tag{1.2}$$

where $\bar{Z} = \sum_i Z_i/n$, n is the number of observation and $S_0 = \sum_{i \neq j} W_{ij}$. Under the null hypothesis that the observed data are an independent and identically distributed sample from a Normal distribution, the distribution of Moran's I is known.

Moran's I is a measure of global autocorrelation, but it has been interpreted as a dependence parameter when adjusting a dependence model to lattice data, such as CAR or SAR models (see Banerjee *et al.* (2004), pp.69-88). In Li *et al.* (2005), the authors shows that Moran's I is only a good estimator of the strength of the spatial dependence parameter when there is little spatial dependence in the data.

The other well-known statistic for measuring spatial correlation in lattice data was introduced by Geary in 1954:

$$c = \frac{n-1}{S_0} \frac{\sum_i \sum_j W_{ij}(Z_i - Z_j)^2}{\sum_i (Z_i - \bar{Z})^2}. \tag{1.3}$$

This statistic does never take negative values, and under the null hypothesis of independence, its mean value is one. Besides, low values (near to zero) indicate positive spatial association. It has also asymptotic Normal distribution.

The performance of this two tests for checking the presence of spatial autocorrelation is evaluated in the same way as for the Cross-Product statistic, that is, by Monte Carlo or randomization tests.

In order to model the dependence structure in a lattice data context, much work has been focused on the Markov Random Field context (see the seminal paper Besag (1974)). In time series analysis, a Markov condition is usually considered: the observation value depends only on

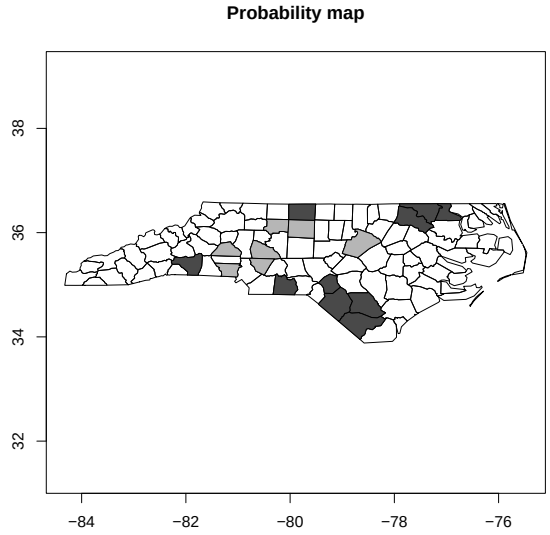

Figure 1.1: Choynowski probability map for North Carolina SID data. Grey: counties with low probabilities. Black: counties with high probabilities.

its inmediate preceding values.

For lattice data, a Markov condition implies that the value of an observation in a spatial site depends only on its neighbour values. This fact confronts the geostatistical point of view, where the dependence between observations is usually a function of the distance (or the difference vector) between the locations where the observations are taken. In the lattice data context, the neighbourhood idea may not be given in terms of Euclidean distances. In some cases, regions which are geographical neighbours may not be considered as lattice neighbours, for instance, because of the existence of geographical barriers.

A well known example in this context is the North Carolina Sudden Infant Death (SID) dataset. Number of sudden infant deaths in all counties in North Carolina were measured from 1974 to 1984. This dataset has been analyzed by different authors (see, for instance, Cressie (1993) Sections 4.4 and 6.2).

In Figure 1.1 we show a choropleth map (Choynowski probability map, Cressie (1993), p.392) for North Carolina SID data. We can see clusters with unusually high values in the north-west and south regions in North Carolina.

	Moran's I		Geary's c	
	$p-$value	I	$p-$value	c
MC	0.01199	0.1436	0.95305	0.8523
RMC	0.007475	0.1436	0.96175	0.8523
RN	0.009691	0.1436	0.98281	0.8523

Table 1.1: Tests for Moran's I and Geary's c. MC: Monte Carlo test. RMC: randomization test, with Monte Carlo approximation of the variance. RN: randomization test, with Normal approximation of the variance.

Both Moran's I and Geary's c provide a statistical measure of global autocorrelation. In Table 1.1 we can see that the value of Moran's I is high (and the value of Geary's c is low), so it is an indicator of positive autocorrelation.

Another atractive feature of Moran's I statistic is that it allows for the construction of a spatial correlogram, due to its similarity with the autocorrelation coefficient from time series. In Figure 1.2, we show the spatial correlogram for this dataset. Lattice neighbours are defined as geographical neighbours (counties that share a border). We can see that there is a significative correlation until second-order neighbours.

This brief illustrative analysis has been done with R software, using `maptools` and `spdep` packages.

1.1.2 Point processes.

A spatial point pattern is a set of locations within a designated region and presumed to have been generated by some stochastic mechanism. Data in this form arise in different context, but a well-known example is that of locations of trees in a forest (see, for instance, Pentinnen *et al.* (1992)). Such data-set is called a spatial point pattern, and locations are refered as *events*, in order to be distinguished from arbitrary points in the region.

In order to illustrate the basic techniques in point process analysis, we will introduce the example of the Laurisilva Forest (see Tawaga (1997)), in the Canary Islands. In this particular example, the DBH (diameter at breast height) has been measure at each event. We work then with a marked spatial point process, since there is a random variable (mark) associated with each

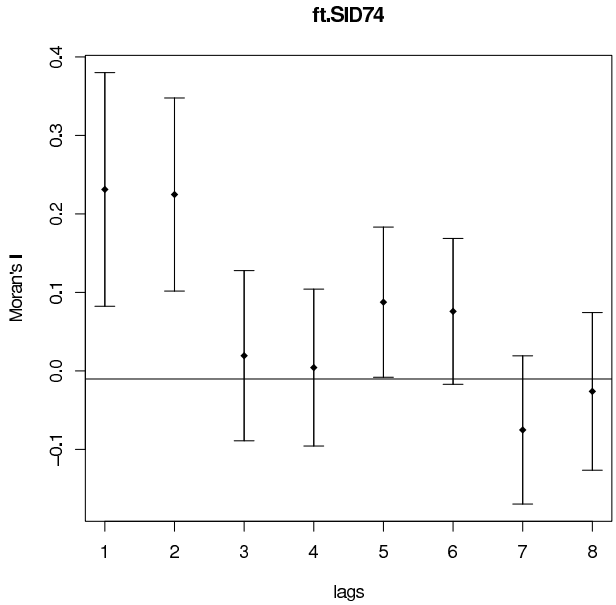

Figure 1.2: Spatial correlogram for North Carolina SID data, based on Moran's I.

spatial location. Data collected in this way can not be treated as independent random variables. Before looking deeper into our example, we introduce some basic concepts.

One of the crucial hypothesis to test when describing a spatial pattern is complete spatial randomness (CSR). For a spatial point pattern, CSR asserts that:

(i) the number of events in any planar region D with area $|D|$ follows a Poisson distribution with mean $\lambda|D|$,

(ii) given n events $\mathbf{x}_i$ in a region D, the $\mathbf{x}_i$ are an independent random sample from a uniform distribution on D.

The constant λ in (i) is the intensity of the process, or mean number of events per unit area. According to (i), CSR therefore implies that the intensity does not vary over the plane and according to (ii), CSR also implies that there are no interactions between the events. Other features that may present a spatial point process are inhibition and cluster behaviour (see Schabenberger and Gotway (2005), Chapter 3).

Rejection of CSR is a previous step for any serious attempt to model an observed pattern as CSR operates as a dividing hypothesis between regular and aggregated (cluster) patterns. Several tests may be used to test the CSR hypothesis, and they can be classified in two groups (e.g.

Schabenberger and Gotway (2005), Section 3.3):

(i) Nearest neighbour distances tests: if d_i denotes the distance from the ith event to the nearest other event in the observation region, then the Empirical Distribution Function (EDF) is given by:

$$\hat{G}(r) = \frac{1}{n}\#(d_i \leq r), \tag{1.4}$$

where # denotes the cardinal of the set.

(i) Point to nearest event distances tests: using distances e_i from each of m sample points in the observation region to the nearest of the n events. The EDF is given by:

$$\hat{F}(r) = \frac{1}{m}\#(e_i \leq r). \tag{1.5}$$

Monte Carlo tests can be implemented for both functions, in order to check for CSR. Also in order to test CSR, other testing techniques are based on the *pair correlation function*. If we denote by $P(r)$ the probability that two circles contain a point of the point process, we can write:

$$P(r) = \lambda^2 g(r)dxdy. \tag{1.6}$$

The function g is called pair correlation function and it is a function of the interpoint distance r. For a CSR-process, this function takes constant value equal to one. Besides, inhibition and clustering behaviour can be detected by this function g. If the pair correlation function takes values larger than 1, this fact means that the interpoint distances around r are more frequent than in a complete random point process. Conversely, when g takes values smaller than one, it indicates that the distances around r are less frequent (inhibition process). When dealing with a clustered process, the pair correlation function will decay to 1, and the distance r_0 from which the pair correlation function exhibits a strictly decreasing behaviour, will give us an idea of the cluster radio. On the other hand, when our data come from an inhibition process, the function will be (almost) increasing. Observing the behaviour of the function with respect to r could provide us with an approximate idea of the inhibition distance.

The function g is related to the cumulative second-order characteristic of the process, given by the $K-$Ripley's function:

$$K(r) = \int_0^r g(u)2\pi u du, \quad r > 0. \tag{1.7}$$

This function has a more intuitive interpretation since $\lambda K(r)$ can be seen as the mean number of further points within a distance no more than r from a randomly chosen point. For a CSR process $K(r) = \pi r^2$. A transformed version of K is given by $L(r) = \sqrt{K(r)/\pi}$. Both K and L are used

for statistical testing whereas g is used for exploratory analysis (Diggle (2003)).

As it has already been mentioned, at each location of the process (that is, at each tree) we have measured the DBH. To measure the dependence between the marks of two events located a distance r apart, we use the mark correlation function (Pentinnen *et al.* (1992)). The mark correlation function can be defined (not in a rigurously way, but quite intuitively) as

$$\rho_f(r) = \frac{Ef(M_1, M_2)}{Ef(M, M')}, \tag{1.8}$$

where M_1 and M_2 are marks at points r units apart and M, M' come from independent realizations of the marginal distribution of the marks. The function f involved in this definition may take different forms, depending on the nature of the marks (for instance, in the case of continuous real-valued processes, $f(x, y) = xy$).

Back to our example, description and interpretation of spatial patterns of trees has been a major focus in forest research. The interaction between trees in a forest makes feasible their consideration as a set of spatially dependent random variables from an underlying stochastic process. We have applied different point processes techniques to describe the behaviour of the trees in a set of laurel forests in the Canary Islands.

Focusing our attention on the spatial pattern of the locations of the trees, we have tested whether the species assess a random distribution, a regular pattern or a cluster behaviour. Besides, we have also performed an analysis on the marks of the trees: the DBH (diameter at breast height). The results we obtain on the statistical analysis confirm the ecological models on competing vegetation and successional status.

Our research has been developed in order to describe the behaviour of macaronesian laurel forest. Macaronesian laurel forest, a subtype of the evergreen lucidophyll oak-laurel forests, is a relic forestand its study is important in helping to understand the composition and ecology of Tertiary Mediterranean flora (at the edges of the Tethys Sea in the Late Miocene epoch). This type of forest is now restricted to northern parts of the Canary Islands (the Laurisilva Forest), Madeira and the Azores.

An experimental plot was located in the Agua García Mountains of Tenerife (UTM x= 362464; y= 3148692) at 820-830 m altitude; with a slope of 8 to 12 facing NNE. Six tree species were considered: *Myrica faya*, *Laurus azorica*, *Erica arborea*, *Persea indica*, *Ilex canariensis*, and *Ilex perado*.

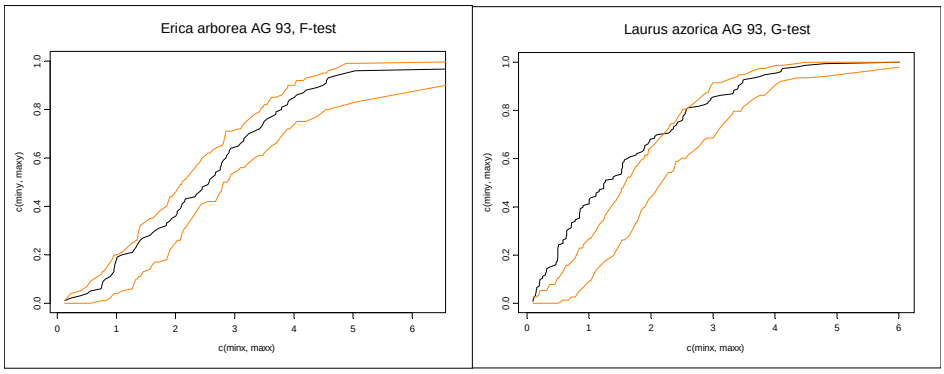

Figure 1.3: F test for *Erica Arborea* and G test for *Laurus Azorica* in Agua García Mountains, for measurements taken in 1993, with 95% confidence bands.

For the pattern of locations of trees measured in 1993, applying the previous tests, we conclude a CSR for the whole forest. When considering just different tree species, we observed a regular pattern for *Erica Arborea* and a cluster pattern for *Laurus Azorica*, with cluster radio smaller than 2 m (see Figure 1.3).

Applications of the mark correlation function have also been conducted. We have seen that *Laurus Azorica*, which exhibits a cluster pattern, shows inhibition between marks (large values of the DBH surrounded by small values), which corresponds to values smaller than one. In the case of *Erica Arborea*, we have seen that there is no strong correlation between marks. At the end of Section 1.2, we will see the realtion between Point Process and Geostatistics.

1.2 A brief introduction to geostatistics.

1.2.1 Stationary spatial processes.

Let $\{Z(\mathbf{s}), \quad \mathbf{s} \in D\}$ be a random process, where $D \subset \mathbb{R}^d$ has positive d-dimensional volume. The description of the process can be done through its finite-dimensional distributions:

$$F_{\mathbf{s}_1,\ldots,\mathbf{s}_m}(z_1,\ldots,z_m) = \mathbb{P}\{Z(\mathbf{s}_1) \leq z_1,\ldots,Z(\mathbf{s}_m) \leq z_m\}, \quad m \geq 1, \tag{1.9}$$

which must satisfy the following conditions of symmetry (F remains invariant if $\mathbf{s}_j$ and z_j are subjected to the same permutation) and consistency:

$$F_{\mathbf{s}_1,\ldots,\mathbf{s}_m}(z_1,\ldots,z_m) = F_{\mathbf{s}_{i1},\ldots,\mathbf{s}_{im}}(z_{i1},\ldots,z_{im}),$$
$$F_{\mathbf{s}_1,\ldots,\mathbf{s}_m,\mathbf{s}_{m+1},\ldots,\mathbf{s}_{m+k}}(z_1,\ldots,z_m,\infty,\ldots,\infty) = F_{\mathbf{s}_1,\ldots,\mathbf{s}_m}(z_1,\ldots,z_m).$$

In order to make inference possible, one usually has to make some assumption on the process. The most widely used are those concerning stationarity and isotropy.

Definition. *Z is said to be strictly stationary if, $\forall \mathbf{s}+\mathbf{u} \in D$ and $\forall m \geq 1$:*

$$F_{\mathbf{s}_1+\mathbf{u},\ldots,\mathbf{s}_m+\mathbf{u}}(z_1,\ldots,z_m) = F_{\mathbf{s}_1,\ldots,\mathbf{s}_m}(z_1,\ldots,z_m). \tag{1.10}$$

Strict stationarity states that the joint distribution of the process, considering a set of observation points, remains the same when this set is translated into any direction $\mathbf{u} \in \mathbb{R}^d$. Indeed, this is a very strong stationarity condition. Weaker stationarity conditions are the following:

Definition. *Z is said to be weakly stationary if:*

$$E(Z(\mathbf{s})) = \mu(\mathbf{s}), \quad \forall \mathbf{s} \in D, \tag{1.11}$$
$$Cov(Z(\mathbf{s}+\mathbf{u}), Z(\mathbf{s})) = C(\mathbf{u}), \quad \forall \mathbf{s}, \mathbf{s}+\mathbf{u} \in D. \tag{1.12}$$

In (1.11), $\mu(\mathbf{s})$ denotes the trend function, which collects the large-scale variability of the process and $C(\mathbf{u})$ denotes the covariance between two observations taken at locations with difference vector $\mathbf{u}$. Weak stationarity means that the first two moments of the process are invariant under translations. This property is also known as second-order stationarity (or the process is said to be homogeneous).

Definition. *Z is said to be intrinsic stationary if:*

$$E(Z(\mathbf{s}+\mathbf{u}) - Z(\mathbf{s})) = 0, \quad \forall \mathbf{s}, \mathbf{s}+\mathbf{u} \in D, \tag{1.13}$$
$$Var(Z(\mathbf{s}+\mathbf{u}) - Z(\mathbf{s})) = 2\gamma(\mathbf{u}), \quad \forall \mathbf{s}, \mathbf{s}+\mathbf{u} \in D. \tag{1.14}$$

The intrinsic stationarity condition implies that for every $\mathbf{u}$, the increment $(Z(\mathbf{s}+\mathbf{u}) - Z(\mathbf{s}))$ is weakly stationary (a more general intrinsic hypothesis assumes that generalized increments are second-order stationary; see, for instance, Chilès and Delfiner (1999), pp.245-251).

The function C in (1.12) is called the *covariogram* (covariance function) and 2γ in (1.14) is the *variogram* (γ is called the *semivariogram* but we may use this notation indistinctly).

Weak and strict stationarity are equivalent for Gaussian processes. In general, the class of intrinsic stationary processes is wider than the class of second-order stationary processes. Let Z be a second-order stationary process with covariogram C. Then:

$$\begin{aligned} &Var(Z(\mathbf{s}+\mathbf{u}) - Z(\mathbf{s})) = \\ &Var(Z(\mathbf{s}+\mathbf{u})) + Var(Z(\mathbf{s})) - 2Cov(Z(\mathbf{s}+\mathbf{u}), Z(\mathbf{s})) = \\ &2(C(\mathbf{0}) - C(\mathbf{u})), \end{aligned}$$

that is

$$2\gamma(\mathbf{u}) = 2(C(\mathbf{0}) - C(\mathbf{u})). \tag{1.15}$$

Thus, a second-order stationary process is intrinsic stationary. This relation does not hold in general on the other direction. For instance, an isotropic $d-$dimensional Brownian motion is intrinsic stationary, but not second-order stationary (Cressie (1993)). When the variogram is a bounded function, we could find an equivalent second-order stationary spatial process whose variogram can be represented as in (1.15) (see Matheron (1973)).

If the covariogram or the variogram can be written as:

$$C(\mathbf{u}) = C^0(\|\mathbf{u}\|) \text{ or } \gamma(\mathbf{u}) = \gamma^0(\|\mathbf{u}\|)$$

the process is *isotropic* (weak isotropy). That is, the dependence structure does not depend on the direction, only on the distance between the locations. The isotropy assumption makes computations easier, since it reduces a d-dimensional problem to one dimension.

A stronger assumption is strict isotropy. A spatial process Z is said to be strictly isotropic (see Stein (1995), p.17) if the finite dimensional joint distributions are invariant under all rigid motions. That is, for all orthogonal $d \times d$ matrix H,

$$\mathbb{P}\left(Z(H\mathbf{s}_1 + \mathbf{s}) \leq z_1, \ldots, Z(H\mathbf{s}_n + \mathbf{s}) \leq z_n\right) = \mathbb{P}\left(Z(\mathbf{s}_1) \leq z_1, \ldots, Z(\mathbf{s}_n) \leq z_n\right).$$

Random fields that become isotropic after a linear transformation of the coordinates are said to exhibit geometric anisotropy. Recently, a test for (weak) isotropy has been proposed in Guan *et al.* (2006).

Both the variogram and the covariogram are used to model the dependence structure in geostatistical processes and satisfy some conditions, which must be taken into account when constructing

estimators for these functions.

- Properties of the covariogram.

1. $C(\mathbf{0}) = Var(Z(\mathbf{s})) \geq 0$.
2. $C(\mathbf{u}) = C(-\mathbf{u})$.
3. $|C(\mathbf{u})| \leq C(\mathbf{0})$.
4. Semidefinite positive. For $n \geq 1$:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0, \quad \forall\{\mathbf{s}_i\}_{i=1}^{n} \subset D, \quad \forall\{a_i\}_{i=1}^{n} \subset \mathbb{R}.$$

Besides, given any semidefinite positive function C, one can construct a Gaussian stationary random process with covariance function given by C (by Bochner's theorem, as we will see later). Thus, the class of semidefinite positive functions in $\mathbb{R}^d$ and the class of covariance functions in $\mathbb{R}^d$ are identical.

Some other properties also satisfied by the covariance functions are the so called *stability properties*. Let C, C_1 and C_2 be covariograms in $\mathbb{R}^d$. In order to obtain valid covariograms in $\mathbb{R}^d$, the following properties hold:

1. aC is a covariogram in $\mathbb{R}^d$, $\forall a \geq 0$.
2. $C_1 + C_2$ is a covariogram in $\mathbb{R}^d$.
3. $C = C_1 \cdot C_2$ is a covariogram in $\mathbb{R}^d$.

Besides, if $\{C_n\}_{n\in\mathbb{N}}$ is a sequence of covariograms in $\mathbb{R}^d$, then

$$C(\mathbf{u}) = \lim_{n\to\infty} C_n(\mathbf{u})$$

is a covariogram, provided that this limit exists for all $\mathbf{u}$.

These properties are derived from the semidefinite condition of the covariogram. The property concerning the product of covariograms is directly obtained by establishing the covariance of the product of two independent random processes with covariance functions C_1 and C_2. Under isotropy, if C is a covariogram in $\mathbb{R}^d$, then it is also a covariogram in $\mathbb{R}^p$, $\forall p \leq d$.

- Properties of the variogram.

The variogram shows how the dissimilarity between $Z(\mathbf{s}+\mathbf{u})$ and $Z(\mathbf{s})$ evolves with the difference vector $\mathbf{u}$. It satisfies the following properties:

1. $\gamma(\mathbf{0}) = 0$.
2. $\gamma(\mathbf{u}) = \gamma(-\mathbf{u})$.
3. $\gamma(\mathbf{u}) \geq 0$.
4. Conditionally semidefinite negative. For $n \geq 1$ and for all $\{a_i\}_{i=1}^n \subset \mathbb{R}$ such that $\sum_{i=1}^n a_i = 0$:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0, \quad \forall \{\mathbf{s}_i\}_{i=1}^n \subset D.$$

Let γ, γ_1 and γ_2 be semivariograms in $\mathbb{R}^d$:

1. $a\gamma$ is also a semivariogram in $\mathbb{R}^d$, $\forall a \geq 0$.
2. $\gamma_1 + \gamma_2$ is also a semivariogram in $\mathbb{R}^d$.

Besides, under isotropy, if γ is a semivariogram in $\mathbb{R}^d$, then it is also a semivariogram in $\mathbb{R}^p$, $\forall p \leq d$.

Some other characteristics related to the variogram function will be introduced. If the variogram is a bounded function, then

$$\lim_{\|\mathbf{u}\|\to\infty} \gamma(\mathbf{u}) = \sigma^2 \tag{1.16}$$

and σ^2 is called the *sill*. If Z is a second-order stationary process and $\lim_{\|\mathbf{u}\|\to\infty} C(\mathbf{u}) = 0$, then $\sigma^2 = C(\mathbf{0})$. If σ^2 is the sill of the semivariogram, the *range* in the direction $\mathbf{e}_0 = \mathbf{u}_0/\|\mathbf{u}_0\| \in \mathbb{R}^d$ is defined by:

$$r_0 = \min\{r : \gamma(r(1+\varepsilon)\mathbf{e}_0) = \sigma^2, \forall \varepsilon > 0\}. \tag{1.17}$$

The range in direction $\mathbf{e}_0$ can be seen as the distance for *independence*: observations of the process Z taken at locations separated by r_0 units in direction $\mathbf{e}_0$ are uncorrelated.

The value of the variogram at the origin is zero, but it may present a discontinuity at this point. This microscale variation is called the *nugget effect*. The nugget, c_0, is defined by:

$$\lim_{\mathbf{u}\to\mathbf{0}} \gamma(\mathbf{u}) = c_0 > 0. \tag{1.18}$$

Provided that σ^2 exists, the difference $(\sigma^2 - c_0)$ is the *partial sill*. Both sill and range are features of the variogram related to the behaviour of this function for large distances. Another property of the variogram is the following:

$$\lim_{\|\mathbf{u}\|\to\infty} \frac{\gamma(\mathbf{u})}{\|\mathbf{u}\|^2} = 0. \tag{1.19}$$

If we try to relate the covariance function of a process and the smoothness of its realizations, we will not find an easy answer. Nevertheless, there is a relation between the covariogram and the mean square properties of a spatial process.

Definition. *A random process Z is mean square continuous at a point $\mathbf{s}$ if*

$$\lim_{\|\mathbf{u}\|\to 0} E(Z(\mathbf{s}+\mathbf{u}) - Z(\mathbf{s}))^2 = 0.$$

This property is also called L^2-continuity.

Therefore, a second order stationary process is mean square continuous if and only if its covariogram C is continuous at the origin. In terms of the variogram, any intrinsic stationary process is L^2-continous everywhere if and only if its variogram in continous at the origin. Mathematically speaking, none L^2-continuous processes can present a discontinuity at the origin: L^2-continuous processes do not present nugget effect.

Further discussion on mean square continuity and differentiability of a random process can be found in Stein (1995), Section 2.4.

Just to complete this overview on the different scenes in spatial statistics, we must remark that point process analysis and geostatistics are closely related. Consider a realization of a (point) process Z on a set of spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, so we have a random field (for the random locations) and a marked point process. In general, a marked point process can not be viewed as a geostatistical process, since there may exist interactions between the locations and the marks. This extension is possible in the random field model introduce in (Walden and Stoyan (1996)). In that context, under stationarity and isotropy assumptions, Mateu and Ribeiro (1999) study the second order characteristics of a marked point process and obtain the following relation between the mark correlation function and the mark variogram:

$$\rho_f(r)\mu^2 = E^2(Z(0)) - \gamma(r). \tag{1.20}$$

1.2.2 Kriging.

Consider that the spatial process Z admits the following decomposition:

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \varepsilon(\mathbf{s}), \tag{1.21}$$

where $\mu(\mathbf{s})$ is the trend of the process, a deterministic function, and ε is a zero-mean process, with known variogram or covariogram. If we observe Z at a set of locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ and we want to predict $Z(\mathbf{s}_0)$ we could consider the best linear unbiased predictor (BLUP), that minimizes the prediction Mean Square Error (see Stein (1995), pp. 2-3 and pp.7-9). Best linear unbiased predictor for $Z(\mathbf{s}_0)$ is called kriging predictor in the geostatistical context. If the mean function μ is assumed to be a known, the best linear unbiased prediction is called simple kriging. If the unknown trend function is assumed to be constant, the best linear unbiased prediction is called ordinary kriging. In a more general setting where μ is unknown but it can be written as a linear combination of a collection of known functions,

$$\mu(\mathbf{s}) = \sum_{j=0}^{p} f_j(\mathbf{s})\beta_j, \tag{1.22}$$

where $(\beta_0, \beta_1, \ldots, \beta_p) \in \mathbb{R}^{p+1}$ is a vector of unknown coefficients, the best linear unibased prediction is called universal kriging. Ordinary kriging can be seen as a particular case of universal kriging.

These prediction methods assume that the covariogram or the variogram are known. For illustration purposes, we will describe the universal kriging method. In this case, if $f_0 = 1$, we can assume that the variogram is known. Otherwise, kriging equations can only be written in terms of the covariogram.

We will introduce the equations for universal kriging prediction, just to highlight the importance of the dependence structure in the construction of such predictors of the process. Consider representation (1.21) for the spatial process Z, where $\mu(\mathbf{s})$ is given by (1.22) and ε is a zero-mean spatial process with known variogram (usually stationary), given by:

$$2\gamma(\mathbf{s}_1, \mathbf{s}_2) = Var(\varepsilon(\mathbf{s}_1) - \varepsilon(\mathbf{s}_2)). \tag{1.23}$$

Assume that $f_0 = 1$ (if $f_0 = 1$ and $p = 0$ we are in an ordinary kriging setting). In matrix notation, we can write:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1.24}$$

where $\boldsymbol{\varepsilon} = (\varepsilon(\mathbf{s}_1), \ldots, \varepsilon(\mathbf{s}_n))^T$, and $\mathbf{X}$ is an $(n \times (p+1))$ matrix where each entry $X_{ij} = f_{j-1}(\mathbf{s}_i)$. Besides,

$$Z(\mathbf{s}_0) = X(\mathbf{s}_0)^T\boldsymbol{\beta} + \varepsilon(\mathbf{s}_0) \tag{1.25}$$

where $X(\mathbf{s}_0) = (f_0(\mathbf{s}_0), \ldots, f_p(\mathbf{s}_0))$. The universal kriging predictor has the following expression:

$$p(Z, \mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i) + \lambda_0, \tag{1.26}$$

and since

$$E\left(\sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i) + \lambda_0\right) = \boldsymbol{\lambda}^T \mathbf{X}\boldsymbol{\beta} + \lambda_0 \tag{1.27}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^T$, a necessary and sufficient condition for $p(Z, \mathbf{s}_0)$ in (1.26) to be uniformly unbiased, that is to say:

$$E(p(\mathbf{Z}, \mathbf{s}_0)) = E(Z(\mathbf{s}_0)) = \mathbf{x}^T\boldsymbol{\beta}, \quad \forall \boldsymbol{\beta} \in \mathbb{R}^{p+1}$$

is that $\lambda_0 = 0$ and

$$\boldsymbol{\lambda}^T \mathbf{X} = \mathbf{x}^T. \tag{1.28}$$

For the particular case of ordinary kriging, since $f_0 = 1$, the restriction on the weights is:

$$\sum_{i=1}^{n} \lambda_i = 1. \tag{1.29}$$

Therefore, the universal kriging predictor is given by:

$$p(\mathbf{Z}, \mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i), \tag{1.30}$$

such that (1.28) is satisfied and the Mean Square Prediction Error is minimized. Then, we must minimize:

$$E\left(Z(\mathbf{s}_0) - \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i)\right)^2 - 2\sum_{j=0}^{p} m_j \left(\sum_{i=1}^{n} \lambda_i f_j(\mathbf{s}_i) - f_j(\mathbf{s}_0)\right) \tag{1.31}$$

with respect to $\{\lambda_i : i = 1, ..., n\}$ and $\{m_j : j = 0, ..., p\}$, which are Lagrange multipliers chosen in order such that (1.28) is satisfied.

Since the predictor is unbiased and the weights satisfy (1.29), we have:

$$\begin{aligned}\left(\sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i) - Z(\mathbf{s}_0)\right)^2 &= \left(\sum_{i=1}^{n} \lambda_i \varepsilon(\mathbf{s}_i) - \varepsilon(\mathbf{s}_0)\right)^2 \\ = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j \left(\varepsilon(\mathbf{s}_i) - \varepsilon(\mathbf{s}_j)\right)^2 &+ \sum_{i=1}^{n} \lambda_i \left(\varepsilon(\mathbf{s}_i) - \varepsilon(\mathbf{s}_0)\right)^2,\end{aligned}$$

and we can write (1.31) as:

$$-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(\mathbf{s}_i,\mathbf{s}_j)+2\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{s}_i,\mathbf{s}_0)-2\sum_{j=0}^{p}m_j\left(\sum_{i=1}^{n}\lambda_i f_j(\mathbf{s}_i)-f_j(\mathbf{s}_0)\right). \tag{1.32}$$

Deriving with respect to $\{\lambda_i : i = 1, ..., n\}$ and $\{m_j : j = 0, ..., p\}$, and taking the equality to zero, we get the $n + p + 1$ kriging equations. In matrix form:

$$\mathbf{\Gamma}_U\boldsymbol{\lambda}_U=\boldsymbol{\gamma}_U, \tag{1.33}$$

with

$$\mathbf{\Gamma}_U=\begin{pmatrix}\mathbf{\Gamma} & \mathbf{X}\\ \mathbf{X}^T & 0\end{pmatrix},\quad \boldsymbol{\lambda}_U=(\boldsymbol{\lambda},\mathbf{m})^T \quad \text{and } \boldsymbol{\gamma}_U=(\gamma,\mathbf{x})^T.$$

where $\boldsymbol{\gamma}=(\gamma(\mathbf{s}_1,\mathbf{s}_0),\ldots,\gamma(\mathbf{s}_n,\mathbf{s}_0))^T$, $\mathbf{m}=(m_0,\ldots,m_p)^T$ and $\mathbf{\Gamma}$ is an $n\times n$ matrix with entries $\mathbf{\Gamma}_{ij}=\gamma(\mathbf{s}_i,\mathbf{s}_j)$. Besides, the minimum Mean Square Prediction Error:

$$\sigma^2_{KU}(\mathbf{s}_0)=2\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{s}_0,\mathbf{s}_i)-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(\mathbf{s}_i,\mathbf{s}_j)$$

can be obtained, just taking into account (1.33), as:

$$\sigma^2_{KU}(\mathbf{s}_0)=\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{s}_0,\mathbf{s}_i)+\sum_{j=0}^{p}m_j f_j(\mathbf{s}_0)=\boldsymbol{\lambda}_U^T\boldsymbol{\gamma}_U.$$

For the particular case of the ordinary kriging (recall, $p = 0$), the expression for the kriging variance is given by:

$$\sigma^2_{KO}(\mathbf{s}_0)=\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{s}_0,\mathbf{s}_i)+m_0.$$

Kriging methods provide an estimation of the prediction variances, which are not obtained from other spatial prediction methods, such as inverse distance (see Cressie (1993), Section 5.9). One of the appealing features of these estimators is that they can be used in the construction of prediction intervals.

In case the covariogram C exists, the kriging equations can be written in terms of C (in facto, if no f_j is equal to 1, then the equations can be only written in terms of the covariogram). The equations can be obtained following a similar process. Some considerations on the kriging interpolation can be seen in Chilès and Delfiner (1999), Sections 3.3.2 and 3.4.2. Further discussion on the effect of the variogram estimation in the kriging equations can be found in Fernández-Casal (2003), Section 3.4.2.

1.2.3 Modeling spatial dependence.

The variogram or the covariogram have a crucial role in spatial prediction, as we can see from (1.33). In practice, for a given realization of a process, the covariogram or the variogram are unknown. Therefore, we must obtain valid covariogram or variogram estimators in order to compute the kriging equations.

The traditional approach for modeling the spatial dependence is to adjust a valid variogram or covariogram model to a pilot estimator.

Pilot estimators.
Assume that the spatial process Z is intrinsic stationary, observed on a set of points $\{\mathbf{s}_i\}_{i=1}^n \subset D \subset \mathbb{R}^d$. A natural estimator for the variogram (known as *classical* or *empirical*) was proposed by Matheron in 1962, based on the estimation method-of-moments:

$$2\hat{\gamma}(\mathbf{u}) = \frac{1}{|N(\mathbf{u})|} \sum_{N(\mathbf{u})} \left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2, \tag{1.34}$$

where

$$N(\mathbf{u}) = \{(i,j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{u}; \quad i,j = 1,\ldots,n\}$$

denotes the set of pairs of points separated by $\mathbf{u}$, and $|N(\mathbf{u})|$ is the cardinal of this set. The main drawback of this estimator is that it is not defined for $\mathbf{u}$ such that $N(\mathbf{u}) = \oslash$. Besides, $N(\mathbf{u}) \neq N(-\mathbf{u})$ whereas $\gamma(\mathbf{u}) = \gamma(-\mathbf{u})$. For large sets of data, obtaining this estimator may be computationally expensive. In Fuentes (2001), a modification of (1.34) using subsampling is proposed, for regularly spaced observations.

In the case that observations are taken at irregularly spaced locations, a smoothed version of the estimator (1.34) is built by defining a tolerance region (usually disjoint) for each $\mathbf{u}$, namely $Tol(\mathbf{u})$. The smoothed empirical estimator is then given by:

$$2\hat{\gamma}(\mathbf{u}) = \frac{1}{|\tilde{N}(\mathbf{u})|} \sum_{\tilde{N}(\mathbf{u})} \left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2, \tag{1.35}$$

where

$$\tilde{N}(\mathbf{u}) = \{(i,j) : \mathbf{s}_i - \mathbf{s}_j \in Tol(\mathbf{u})\}.$$

Another problem of (1.34) is its lack of robustness. If Z is a Gaussian process, then the square of the difference follows approximately a shifted χ^2 distribution; it is highly asymmetric. A first approach to solve this drawback is replacing the sum of squared differences by a lower power.

Cressie and Hawkins (1980) present in their work a more robust estimator for the variogram:

$$2\hat{\gamma}(\mathbf{u}) = \frac{1}{0.457 + \frac{0.494}{|N(\mathbf{u})|}} \left(\frac{1}{|N(\mathbf{u})|} \sum_{N(\mathbf{u})} |Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{1/2} \right)^4 . \tag{1.36}$$

Another alternative is to consider a robust statistic on the differences $\{|Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{1/2}\}$, such as the median. In that case, the denominator in (1.36) must be adapted for bias correction. On this scope, Armstrong and Delfiner (1980) propose an estimator based on the quantiles of the squared-differences of the process:

$$\hat{\gamma}(\mathbf{u}) = Q_p \left(\frac{1}{2}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2 \right), \quad \text{where } \mathbf{s}_i - \mathbf{s}_j \sim Tol(\mathbf{u}) \tag{1.37}$$

and Q_p denotes the p-quantile. If the process Z is Gaussian, then the distribution of the differences in (1.37) are shifted χ^2, with scaling factor $\gamma(\mathbf{u})$.

In Genton (1998) the robustness of the estimator (1.36) is discussed and a highly robust estimator based on M-estimation procedures is proposed (see Shao (2003), Section 5.2.2). Both (1.34) and (1.36) can be seen as particular cases of the general estimator proposed in Genton (1998), but none of them are bias-robust (infinitesimal modifications in observations may provoque unexpectedly large changes in the estimator).

Nonparametric estimators of the variogram have been also introduced by García-Soidán *et al.* (2003) and García-Soidán *et al.* (2004). These estimators are based on nonparametric regression techniques, just viewing the variogram cloud as a dispersion plot. The Nadaraya-Watson estimator for the variogram is given by:

$$2\hat{\gamma}(\mathbf{u}) = \frac{\sum_{i \neq j} W_{ij} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2}{\sum_{i \neq j} W_{ij}}, \tag{1.38}$$

where the weigths W_{ij}, in the isotropic case, are defined as:

$$W_{ij} = K \left(\frac{\|\mathbf{u}\| - (\|\mathbf{s}_i - \mathbf{s}_j\|)}{h} \right),$$

with K a kernel function and h a bandwidth parameter. These estimators are asymptotically unbiased and efficient. Despite their flexibility, these estimators present an important drawback since, in general they are not valid variograms (they do not satisfy the conditional negative definite property). Besides, as it happens in the regression context, this estimator may present edge-effect problems, which can be mitigated by the used of a boundary-kernel function K.

The local linear estimator in García-Soidán *et al.* (2003) for $2\hat{\gamma}(\mathbf{u})$ is obtained, in the isotropic case, from (1.38) but considering weights:

$$W_{ij} = K\left(\frac{\|\mathbf{u}\| - \|\mathbf{s}_i - \mathbf{s}_j\|}{h}\right) \cdot \sum_{i\neq j}\sum_{k\neq l} K\left(\frac{\|\mathbf{u}\| - \|\mathbf{s}_k - \mathbf{s}_l\|}{h}\right)(\|\mathbf{s}_k - \mathbf{s}_l\| - \|\mathbf{s}\|)(\|\mathbf{s}_k - \mathbf{s}_l\| - \|\mathbf{s}_i - \mathbf{s}_j\|).$$

This estimator avoids the edge-effects, but as the Nadaraya-Watson estimator, it does not provide a valid variogram model.

In some cases, it is not easy to determine whether a nonparametric estimator of the variogram is valid or not, and this condition is necessary in order to compute the kriging equations. If these estimators do not satisfy the conditionally definite negative condition, kriging variance estimators may be negative or even the kriging system matrix may be singular.

The power of these estimators, such as the Nadaraya-Watson or the local-linear one, lies in the fact that they can be used as pilot estimators when trying to adjust a valid parametric model. Before going into these adjustment methods, we will introduce some valid parametric families of variograms.

Parametric models.
Most of the models proposed below are valid (conditionally definite negative) isotropic variograms in $\mathbb{R}^d$, for $d \geq 1$. For simplicity, we will consider isotropic models.

- Linear model.

$$\gamma_\theta(\mathbf{u}) = \begin{cases} 0 & \text{if} \quad \mathbf{u} = \mathbf{0}, \\ c_0 + c_1\|\mathbf{u}\| & \text{if} \quad \mathbf{u} \neq \mathbf{0}. \end{cases} \tag{1.39}$$

This model depends on two parameters: $\theta = (c_0, c_1)^T$. The nugget effect is denoted by c_0 throughout the text and $c_1 \geq 0$ is a scale parameter. It is straightforward to see that this model can not correspond to a second-order stationary process. For instance, it is easy to see that the Brownian motion has linear semivariogram. In practice, one could also consider a truncated linear model.

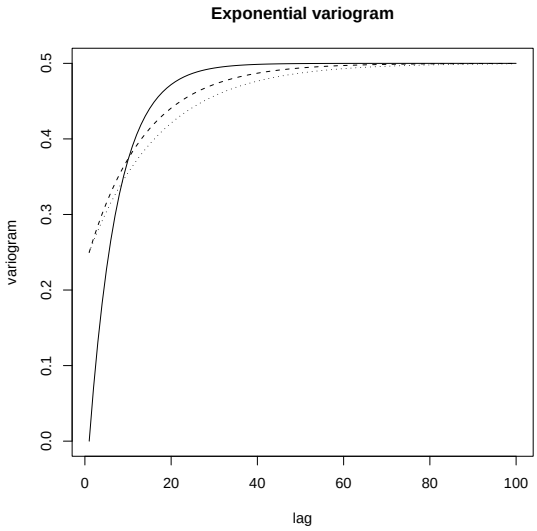

Figure 1.4: Exponential variograms. Solid line: $c_0 = 0.0$, $c_1 = 0.5$, $a = 20$. Dashed line: $c_0 = 0.25$, $c_1 = 0.25$, $a = 30$. Dotted line: $c_0 = 0.25$, $c_1 = 0.25$, $a = 50$.

- Exponential model.

$$\gamma_\theta(\mathbf{u}) = \begin{cases} 0 & \text{if} \quad \mathbf{u} = \mathbf{0}, \\ c_0 + c_1 \left(1 - e^{-\frac{3\|\mathbf{u}\|}{a}}\right) & \text{if} \quad \mathbf{u} \neq \mathbf{0}. \end{cases} \tag{1.40}$$

The exponential variogram is a bounded model, but it does not reach the sill, $c_0 + c_1$. Parameter a denotes the practical range, which is defined as the lag where the 95% of the sill is reached.

- Rational model.

$$\gamma_\theta(\mathbf{u}) = \begin{cases} 0 & \text{if} \quad \mathbf{u} = \mathbf{0}, \\ c_0 + c_1 \frac{\|\mathbf{u}\|^2}{a^2/19 + \|\mathbf{u}\|^2} & \text{if} \quad \mathbf{u} \neq \mathbf{0}. \end{cases} \tag{1.41}$$

The sill is given by $c_0 + c_1$ and the pratical range is a.

- Power model.

$$\gamma_\theta(\mathbf{u}) = \begin{cases} 0 & \text{if} \quad \mathbf{u} = \mathbf{0}, \\ c_0 + c_1 \|\mathbf{u}\|^{c_2} & \text{if} \quad \mathbf{u} \neq \mathbf{0}, \end{cases} \tag{1.42}$$

with $0 \leq c_2 < 2$, $c_1 \geq 0$. This model is not bounded and it encloses the linear variogram as a particular case.

- Spherical model (valid for $\mathbb{R}^d$, $d = 1, 2, 3$).

$$\gamma_\theta(\mathbf{u}) = \begin{cases} 0 & \text{if} \quad \|\mathbf{u}\| = 0, \\ c_0 + c_1 \left(\frac{3}{2} \frac{\|\mathbf{u}\|}{a} - \frac{1}{2} \left(\frac{\|\mathbf{u}\|}{a} \right)^3 \right) & \text{if} \quad 0 < \|\mathbf{u}\| \leq a, \\ c_0 + c_1 & \text{if} \quad \|\mathbf{u}\| > a. \end{cases} \tag{1.43}$$

The spherical model shows a behaviour similar to the exponential variogram. It is bounded and it reaches the sill, $c_0 + c_1$. Nugget and range are given by c_0 and a, respectively.

- Wave model (valid for $\mathbb{R}^d$, $d = 1, 2, 3$).

$$\gamma_\theta(\mathbf{u}) = \begin{cases} 0 & \text{if} \quad \mathbf{u} = \mathbf{0}, \\ c_0 + c_1 \left(1 - \frac{a}{\|\mathbf{u}\| \sin\left(\frac{\|\mathbf{u}\|}{a} \right)} \right) & \text{if} \quad \mathbf{u} \neq \mathbf{0}. \end{cases} \tag{1.44}$$

For large values of $\|\mathbf{u}\|$, the wave models oscillates around the sill, $c_0 + c_1$.

- Matérn model.

$$\gamma_\theta(\mathbf{u}) = \begin{cases} 0 & \text{if} \quad \mathbf{u} = \mathbf{0}, \\ c_0 + c_1 \left(1 - \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\|\mathbf{u}\|}{a} \right)^\nu \mathcal{K}_\nu \left(\frac{\|\mathbf{u}\|}{a} \right) \right) & \text{if} \quad \mathbf{u} \neq \mathbf{0}, \end{cases} \tag{1.45}$$

where $\mathcal{K}_\nu$ is a modified Bessel function of the second kind of order ν (see Abramowitz and Stegun (1965), pp.358-389). Special attention is paid to this family of variograms at the end of the chapter.

Fitting a valid model.

Variogram estimators can not be used directly in the kriging equations, since if they do not satisfy the conditional semidefinite negative property one may obtain negative estimations for the mean square prediction error, as we have already noticed. Usually, one tries to adapt a parametric model in order to describe the dependence in a dataset. For that purpose, different adjustment techniques have been proposed, such as Maximun Likelihood, Restricted Maximum Likelihood and Least Squares (LS) methods. An overview on these techniques can be seen in Cressie (1993). As an example, we will introduce the LS methods, since they involve a pilot estimator of the dependence structure (for instance, any of the nonparametric estimators introduced above) and a

parametric model.

Denote by $2\gamma(\mathbf{h}, \theta_0)$ the theoretical variogram and denote by $\hat{\gamma}_i = \hat{\gamma}(\mathbf{h}_i)$ the semivariogram estimations obtained by any of the nonparametric methods introduced along this section.

The LS estimator of $\hat{\theta}_0$ (see Cressie (1993), pp.96-97) is obtained by minimizing

$$(\hat{\gamma} - \gamma(\theta))^T \mathbf{V}(\theta) (\hat{\gamma} - \gamma(\theta)),$$

where $\hat{\gamma} = (\hat{\gamma}(\mathbf{h}_1), \ldots, \hat{\gamma}(\mathbf{h}_k))^T$, $\gamma(\theta) = (\gamma(\mathbf{h}_1, \theta), \ldots, \gamma(\mathbf{h}_k, \theta))^T$ and $\mathbf{V}(\theta)$ is a $k \times k$ semidefinite positive matrix, which may depend on θ on any of the following ways:

- Ordinary Least Squares (OLS): $\mathbf{V}(\theta) = I_k$, where I_k denotes the $k \times k$ identity matrix.
- Weighted Least Squares (WLS): $\mathbf{V}(\theta) = diag(w_1(\theta), \ldots, w_k(\theta))$, with $w_i(\theta) \geq 0$, for $i = 1, \ldots, k$. These weights are usually proportional to the inverse of $Var(\hat{\gamma}(\mathbf{h}_i))$.
- Generalized Least Squares (GLS): $\mathbf{V}(\theta) = \Sigma_{\hat{\gamma}}(\theta)^{-1}$, where $\Sigma_{\hat{\gamma}}(\theta)$ is the covariance matrix of $\hat{\gamma}$, considering $\gamma(\theta)$ as the true model for the dependence.

This approach has been criticized since the weighting matrix depends also on the target parameter. An alternative would be to consider the minimization of:

$$(\hat{\gamma} - \gamma(\theta))^T \mathbf{V}(\theta_0) (\hat{\gamma} - \gamma(\theta)).$$

But this method can not be directly applied in practice, since it depends on the unknown parameter θ_0 through the variance-covariance matrix $\mathbf{V}$. This problem ca be solved by using an iterative method.

Nonparametric methods.
A different nonparametric alternative is found in the spectral representation of the covariogram (see Section 1.3.3 devoted to the Hankel transform). Based on this spectral representation of the isotropic covariogram, Shapiro and Botha (1991) proposed a method for adjusting a valid variogram.

The representation of the variogram of an isotropic second-order stationary process can be written as:

$$\gamma(u) = \begin{cases} c_0 - C(u) & u \neq 0, \\ 0 & u = 0, \end{cases} \tag{1.46}$$

where $c_0 > 0$ and C is a covariogram, with the following spectral representation:

$$C(u) = \int_0^\infty \kappa_d(\omega u) dG(\omega), \tag{1.47}$$

where

$$\kappa_d(x) = \left(\frac{2}{x}\right)^{\frac{d-2}{2}} \Gamma\left(\frac{d}{2}\right) J_{\frac{d-2}{2}}(x),$$

with J_l a Bessel function of order l (see Abramowitz and Stegun (1965), pp.358-389). G is a positive function, bounded and non-decreasing on $[0, \infty)$. Therefore, the problem reduces to finding a positive constante c_0 and a positive, bounded and non-decreasing function G such that the variogram in (1.46) describes the dependence structure of the dataset. However, this problem may present some incovenients when trying to solve it numerically. Shapiro and Botha (1991) consider a discretization of the function G, considering that dG is an atomic measure with finite jumps z_j at points x_j, $j = 1, \ldots, J$:

$$G(x) = \sum_{x_j \leq x} z_j,$$

and, for simplicity, the authors considered equally-spaced points $x_j = z\phi$, where $\phi > 0$ is a fixed constant.

The variograms obtained by this method have the following representation:

$$\gamma(u) = \sum_{j=1}^{J} \kappa_d(x_j u) z_j, \tag{1.48}$$

subject to

$$c_0 - \sum_{j=1}^{J} z_j \geq 0. \tag{1.49}$$

The restriction (1.49) corresponds to the following property of semidefinite positive functions: the absolute value of any semidefinite positive function is bounded by its value at the origin. Therefore, in (1.46), we have that:

$$c_0 - C(0) = c_0 - \int_0^\infty dG(\omega) d\omega \geq 0.$$

By a Weighted Least Squares procedure, one could determine the vector of unknown parameters $(z_1, \ldots, z_J, c_0)^T$, by quadratric programming. A Generalized Least Squares method could be also applied.

Other approaches have been also given by Barry and Ver Hoef (1996) (black-box models) and by Lele (1995), based on splines.

1.3 Spectral representation of random fields.

At first sight, spectral representation of random fields may appear artificious since mathematics in the frequency domain involve complex-valued random variables. Besides, the interpretation of the spectrum requires a further effort since the intuition we may have when dealing with the data may be lost when working with their signal. Nevertheless, spectral methods are gaining an increasing importance in spatial data analysis. The main advantage of an spectral domain perspective is that mathematical theory is often simpler in the frequency domain. Besides, the spectral density function and the covariance function of a stationary stochastic process are closely related since they form a Fourier Transform pair. So, studying the second-order properties of a random field via the covariance function or the spectral density can be viewed as equivalent, as we will see in subsequent sections.

However, we should note that the spectral density and the covariance are two different but complementary representations of the second-order properties of a stochastic process. The covariance function emphasizes spatial dependence as a function of coordinate separation. The spectral density fucntion emphasizes the association of components of variability with frequencies.

The spectral density function can be estimated from data via the periodogram, that will be introduced later. This feature does not provide any particular challenges beyond computing the sample covariances, at least if data are observed on a grid. Nevertheless, summary statistics calculated from data in the spatial domain are usually correlated. This correlation may arise from the fact that the same data point $Z(\mathbf{s}_i)$ is repeatedly used in multiple summaries and/or from the spatial autocorrelation. In the spectral domain, the ordinates of the periodogram, the data-based estimate of the spectral density function, are -at least asymptotically- independent and have simple distributional properties. This enables the construction of test statistic with standard properties.

Since covariance function and spectral density form a Fourier transform pair, it is worth it making a brief review of the concepts of Fourier analysis. Fourier analysis is the study of how functions defined on a continuum can be represented and analyzed in terms of periodic functions like sines and cosines. We will recall the notions of Fourier Transform (FT), Discrete Fourier Transform (DFT) and Fast Fourier Transform (FFT), in order to understand better their relationship and to clarify technical details which may arise in theoretical developments involving this functions. We will introduce first the Fourier Transform of L^1-functions. Then, we will also define the Fourier Transform of an absolutely summable sequence. Once we have defined the Fourier Transform, we will introduce a natural approximation: the Discrete Fourier Transform.

We will also give a brief overview on the Fast Fourier Transform algorithm, which will be used in practice. Finally, we will present the spectral representation of stationary processes, giving the relationship between the covariance function and the spectral density of a process.

1.3.1 Fourier transform in $L^1(\mathbb{R}^d)$.

More theoretical details on this topic can be seen in (Dautray and Lions (1985)). Denote by $\mathbf{x}^T\boldsymbol{\omega} =< \mathbf{x}, \boldsymbol{\omega} >= \sum_{k=1}^{d} x_k\omega_k$, the usual scalar product in $\mathbb{R}^d$ and let $L^1(\mathbb{R}^d)$ be the set of absolutely integrable functions on $\mathbb{R}^d$.

Definition. *For $g \in L^1(\mathbb{R}^d)$ define the Fourier Transform of g by:*

$$\hat{g}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} g(\mathbf{x})e^{-i\mathbf{x}^T\boldsymbol{\omega}}, \quad \boldsymbol{\omega} \in \mathbb{R}^d. \tag{1.50}$$

It is easy to see that the Fourier transform exchanges translation of amplitude and multiplication by an exponential and the mapping $g \to \hat{g}$ is linear. Besides, taking into account that

$$|\hat{g}(\boldsymbol{\omega})| \leq \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} |g(\mathbf{x})|d\mathbf{x}, \quad \forall\boldsymbol{\omega} \in \mathbb{R}^d,$$

it is clear that $\hat{g}$ is a bounded, continuous function on $\mathbb{R}^d$ and $\|\hat{g}\|_\infty \leq \|g\|_1$, where

$$\|\hat{g}\|_\infty = \sup_{\boldsymbol{\omega}} |\hat{g}(\boldsymbol{\omega})|, \quad \text{and } \|\hat{g}\|_1 = \int |\hat{g}(\boldsymbol{\omega})|d\boldsymbol{\omega}.$$

In addition, $|\hat{g}(\boldsymbol{\omega})| \to 0$ in $\mathbb{C}$, as $|\boldsymbol{\omega}| \to \infty$, by the Riemann-Lebesgue Theorem. It may be interesting to guarantee that the Fourier Transform exchanges differentiability and decay at infinity; then, it is necessary to introduce a new space of functions of rapid decay at infinity and belonging to $\mathcal{C}^\infty$ (the space of infinitely differentiable functions). Let's introduce the following notation:

$$\mathbf{x}^\alpha = \prod_{i=1}^{d} x_i^{\alpha_i}, \quad \text{and} \quad D^\beta g = \frac{\partial^{\beta_1+\ldots+\beta_d} g}{\partial x_1^{\beta_1}\ldots\partial x_d^{\beta_d}}.$$

Definition. *The Schwarz space of functions is defined as:*

$$\mathcal{S}(\mathbb{R}^d) = \{g \in \mathcal{C}^\infty; \forall\alpha, \beta \in \mathbb{N}^d, \mathbf{x}^\alpha D^\beta g \to 0, \text{ as } |\mathbf{x}| \to \infty\}.$$

The Schwarz space $S(\mathbb{R}^d)$ is not a normed space, but it is dense in $L^p(\mathbb{R}^d)$, $\quad \forall 1 \leq p < \infty$. In our case, the most important characteristic of this space is that, for $g \in \mathcal{S}$, we can define its Fourier Transform $\hat{g}$, which is also an element of $\mathcal{S}$. The corresponding inversion formula is:

$$g(\mathbf{x}) = \int_{\mathbb{R}^d} \hat{g}(\boldsymbol{\omega})e^{i\mathbf{x}^T\boldsymbol{\omega}}, \quad \mathbf{x} \in \mathbb{R}^d. \tag{1.51}$$

This is the Inverse Fourier Transform (IFT). The Fourier Transform is an isomorphism of $\mathcal{S}$ onto itself, with inverse given by (1.51). Expression (1.50) and (1.51) are said to form a Fourier Transform pair.

Many authors begin by discussing the Fourier transform on $L^1(\mathbb{R}^d)$ (or even directly on $L^2(\mathbb{R}^d)$). Dautray and Lions (1985) give two reasons for considering the Schwartz space when defining the Fourier Transform. First, the Fourier transform is a one-to-one map of Schwartz space onto itself. This makes it easy to talk about the inverse Fourier transform, which turns out to be the inverse map. That is, on Schwartz space, it is possible to deal with the transform and the inverse transform. Consider now $g, f \in \mathcal{S}(\mathbb{R}^d)$. Then, the following statements hold:

- Convolution formula.

$$\widehat{(g * f)} = \hat{g}\hat{f}, \quad \text{where } (g * f)(\mathbf{x}) = \int_{\mathbb{R}^d} g(\mathbf{x} - \mathbf{u}) f(\mathbf{u}) d\mathbf{u},$$

denotes the convolution operator.

- Derivatives.

$$\widehat{D^\alpha g} = (i\boldsymbol{\omega})^\alpha \hat{g}, \quad \text{and} \quad D^\beta \hat{g} = \widehat{(-i\mathbf{x})^\beta g}, \quad \forall \alpha, \beta \in \mathbb{N}^d.$$

Providing $\mathcal{S}(\mathbb{R}^d)$ with the topology induced by that of $L^2(\mathbb{R}^d)$, the following statements hold:

- The Plancherel Theorem.

$$< g, f >= (2\pi)^d < \hat{g}, \hat{f} > \quad \text{and} \quad \|g\| = (2\pi)^2 \|\hat{g}\|^2. \tag{1.52}$$

where

$$< g, f >= \int_{\mathbb{R}^d} g(\mathbf{x}) \overline{f(\mathbf{x})} d\mathbf{x} \quad \text{and} \quad \|g\| = \sqrt{< g, g >},$$

and $\|\cdot\|$ denotes the L^2-norm. Though all these statements are also true for the Fourier Transform on $L^2(\mathbb{R}^d)$, it is not possible to define the Fourier Transform on $L^2(\mathbb{R}^d)$ directly, by the integral formula, since $L^2(\mathbb{R}^d)$ functions may not be in $L^1(\mathbb{R}^d)$; a limiting procedure must be used.

In $L^2(\mathbb{R}^d)$, the integral $\int_{\mathbb{R}^d} g(\mathbf{x})e^{i\mathbf{x}^T\boldsymbol{\omega}}d\mathbf{x}$ does not have a meaning in so far as the Lebesgue integral $\int_{\mathbb{R}^d} |g(\mathbf{x})|d\mathbf{x}$ diverges for $g \in L^2$ but $g \notin L^1$. Anyway, we use the same notation.

Before going on, we should point out that the definition of Fourier Transform that we have adopted here is not universally accepted. Two other frequently used definitions are:

$$\tilde{g}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} g(\mathbf{x})e^{-i\mathbf{x}^T\boldsymbol{\omega}}, \quad g(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \tilde{g}(\boldsymbol{\omega})e^{i\mathbf{x}^T\boldsymbol{\omega}}d\boldsymbol{\omega},$$

and

$$\check{g}(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} g(\mathbf{x})e^{-i2\pi\mathbf{x}^T\boldsymbol{\omega}}, \quad g(\mathbf{x}) = \int_{\mathbb{R}^d} \check{g}(\boldsymbol{\omega})e^{i2\pi\mathbf{x}^T\boldsymbol{\omega}}.$$

There may be also differences in the exponential function. The minus sign may be omitted in $\hat{g}$, $\tilde{g}$ and $\check{g}$, but the it appears in the exponent of the inversion formula. These three versions of the Fourier Transform have their advantages and disadvantages. $\tilde{g}$ has the advantage of getting rid of the factor $(2\pi)^d$ in Plancherel's theorem, but it introduces it in the convolution formula, $\tilde{(f * g)} = (2\pi)^{d/2}\tilde{f}\tilde{g}$. $\check{g}$ obviates this factor in both Plancherel's and convolution formulae, but it introduces it in the formula for the derivatives $\check{g}'(\boldsymbol{\omega}) = (2\pi)^d i\boldsymbol{\omega}\check{g}(\boldsymbol{\omega})$. In any case, the Fourier Transform preserves exactly the scalar product and the L^2-norm of the elements of $\mathcal{S}(\mathbb{R}^d)$.

Our interest may not be always focused on $L^1(\mathbb{R}^d)$ functions, but on sequences of real numbers. Consider and absolutely summable sequence

$$\{\rho(\mathbf{r}); \mathbf{r} \in \mathbb{Z}^d\}, \quad \sum_{\mathbf{r}\in\mathbb{Z}^d} |\rho(\mathbf{r})| < \infty.$$

It can be seen as the sequence of Fourier coefficients of a function g, defined on $L^1(\Pi^d)$, where $\Pi^d = [-\pi, \pi]^d$:

$$\hat{g}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \sum_{\mathbf{r}\in\mathbb{Z}^d} \rho(\mathbf{r})e^{-i\mathbf{r}^T\boldsymbol{\omega}}, \tag{1.53}$$

where the Fourier coefficients are given by:

$$\rho(\mathbf{r}) = \int_{\Pi^d} \hat{g}(\mathbf{x})e^{i\mathbf{r}^T\mathbf{x}}. \tag{1.54}$$

The Fourier Transform of an infinite absolutely summable sequence $\{\rho(\mathbf{r})\}$ is defined by (1.53) and the corresponding Inverse Fourier Transform by (1.54). To sum up, we have defined the Fourier Transform of an $L^1(\mathbb{R}^d)$ function and the Fourier Transform of an absolutely summable sequence. We will deal with these two cases later, when discussing the spectral density of a process.

We will give a final result which will be useful in the next section. Suppose that $g \in L^2(\mathbb{R})$ and g is band-limited. That is to say, g involves only frequencies smaller than some constant Ω

(or equivalently, $\hat{g}$ vanishes outside $[-\Omega, \Omega]$). This result can be extended to higher dimensions.

Theorem. *Suppose that $g \in L^2(\mathbb{R})$ and $\hat{g}(\omega) = 0$ for $|\omega| \geq \Omega$. Then, g is completely determined by its values at the points $x_n = n\pi/\Omega$, $n = 0, \pm 1, \pm 2, \ldots$. In fact:*

$$g(x) = \sum_{n=-\infty}^{\infty} g\left(\frac{n\pi}{\Omega}\right) \frac{\sin(\Omega x - n\pi)}{\Omega x - n\pi}. \tag{1.55}$$

Proof. See Folland (1993), pp.230-231. □

Although in practice the band-limited condition is rarely possible to meet, many functions of practical use are *essentially band-limited* and have a rapid decay rate. A function g is essentially band-limited if there exist constants $\beta, \mu > 0$ such that:

$$|\hat{g}(\boldsymbol{\omega})| \leq \beta(1 + |\boldsymbol{\omega}|)^{-1-\mu},$$

which means that $|\hat{g}(\boldsymbol{\omega})|$ decays faster than $|\boldsymbol{\omega}|^{-1}$ as $|\boldsymbol{\omega}| \to \infty$. For these kind of functions, it is possible to choose a grid spacing sufficiently small that the error in the representation (1.55) is negligible. A deeper and more precise statement about the rate of decay of Fourier Transforms is given by the Paley-Wiener Theorem (see Dautray and Lions (1985), pp.505-506).

1.3.2 Discrete Fourier Transform.

The Discrete Fourier Trasnsform can be introduced from two points of view. We will briefly describe both approaches.

Analytical approach.
Suppose that $g(\mathbf{x})$, with $\mathbf{x} \in \mathbb{R}^d$, is a 2π-periodic function with complex values. That is, $g(\mathbf{x}) = g(x_1, \ldots, x_d) = g(x_1 + 2\pi, \ldots, x_d + 2\pi)$, for all $\mathbf{x} \in \mathbb{R}^d$. Its Fourier series is given by:

$$g(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} c_{\mathbf{k}} e^{i\mathbf{k}^T\mathbf{x}}.$$

From the scalar product:

$$< g, f >= \frac{1}{(2\pi)^d} \int_{\Pi^d} g(\mathbf{x})\overline{f(\mathbf{x})} d\mathbf{x},$$

where $\overline{f(\mathbf{x})}$ denotes the conjugate of $f(\mathbf{x})$, there results the expression:

$$g(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} < g, e^{i\mathbf{k}^T\mathbf{x}} > e^{i\mathbf{k}^T\mathbf{x}}. \tag{1.56}$$

Let S_N be the finite dimensional space generated by the $N = \prod_{i=1}^{d}(2n_i + 1)$ basis:

$$\{e^{i\mathbf{k}^T\mathbf{x}}\}_{\mathbf{k}\in\mathcal{N}}, \quad \mathcal{N} = \{\mathbf{k} = (k_1, \ldots, k_d); k_i = -n_i, \ldots, n_i\}.$$

The function $g_N \in S_N$ obtained by truncating the expression of the Fourier series of g in (1.56), is the best approximation to g (in the sense of the L^2-norm) in S_N. But it is difficult to calculate, since integrals must be evaluated to get the $c_{\mathbf{k}}$ coefficients.

Consider $h_N \in S_N$ given by:

$$h_N(\mathbf{x}) = \sum_{\mathbf{k}\in\mathcal{N}} a_{\mathbf{k}} e^{i\mathbf{k}^T\mathbf{x}},$$

such that h_N coincides with g at the N points

$$\mathbf{x}_{\mathbf{k}} = \left(\frac{2\pi k_1}{2n_1 + 1}, \ldots, \frac{2\pi k_d}{2n_d + 1}\right), \quad 0 \le k_i \le 2n_i, \quad i = 1, \ldots, d \tag{1.57}$$

that is, h_N is the interpolate of g in S_N:

$$\sum_{\mathbf{j}\in\mathcal{N}} a_{\mathbf{j}} e^{i\mathbf{j}^T\mathbf{x}_{\mathbf{k}}} = g(\mathbf{x}_{\mathbf{k}}).$$

Therefore, if we denote by $\mathcal{N}^* = \{\mathbf{k} = (k_1, \ldots, k_d); k_i = 0, \ldots, 2n_i\}$, for $\mathbf{j} \in \mathcal{N}$:

$$a_{\mathbf{j}} = \frac{1}{(2\pi)^d N} \sum_{\mathbf{k}\in\mathcal{N}^*} g(\mathbf{x}_{\mathbf{k}}) e^{i\mathbf{j}^T\mathbf{x}_{\mathbf{k}}}, . \tag{1.58}$$

Definition. *The mapping between $\mathbb{C}^N \to \mathbb{C}^N$ which associates $\{a_{\mathbf{k}}\}_{\mathbf{k}\in\mathcal{N}}$ with $\{g(\mathbf{x}_{\mathbf{k}})\}_{\mathbf{k}\in\mathcal{N}^*}$ given by (1.58) is the Discrete Fourier Transform.*

The Discrete Fourier Transform is, roughly speaking, a linear mapping that operates on $N-$dimensional vectors of $\mathbb{C}^d$ in much the same way that the Fourier Transform operates on functions on $\mathbb{R}^d$.

In dimension $d = 1$, $N = 2n + 1$ is odd. In the case that N is an even number, S_N must be chosen as the space that is generated by:

$$\{e^{ikx}\}_k, \quad -n + 1 \le k \le n,$$

which has dimension $2n$. This also must be extended for the higher dimensional setting.

Remark. The reader must not be confused by the notation in terms of $\mathbf{x}_{\mathbf{k}}$ in (1.57). If g is 2π-periodic, the corresponding Fourier frequencies are integer numbers, and the equations in the

spectral domain are simpler. In a general setting, for a A_l-periodic function g in dimension l, we would consider a grid of points $\boldsymbol{x}_{\mathbf{k}}^T = (x_{k_1}, \ldots, x_{k_d})$; $\quad x_{k_l} = \Delta_l k_l \;\; k_l = -n_l$ where $\Delta_l = \frac{A_l}{(2n_l+1)}$, and the corresponding basis should be $\{e^{i\boldsymbol{\omega}_{\mathbf{k}}^T \mathbf{x}}\}$ with $\boldsymbol{\omega}_{\mathbf{k}}^T = (\omega_{k_1}, \ldots, \omega_{k_d})$; $\quad \omega_{k_l} = \frac{2\pi k_l}{\Delta_l(2n_l+1)}$; $\quad k_l = -n_l, \ldots, n_l, \quad l = 1, \ldots, d$.

Numerical approach.

Based on a quadrature approximation formula, an alternative motivation of the Discrete Fourier Transform can be found in (Briggs and Henson (1995)) and its briefly described below. For simplicity, we will only restrict to the one dimensional case. Assume that $g(x)$, $x \in \mathbb{R}$ vanishes outside $[-A/2, A/2]$. Then, the Fourier Transform of this function with limited extent is given by

$$\hat{g}(\omega) = \frac{1}{2\pi} \int_{\infty}^{\infty} g(x) e^{-ix\omega} dx = \frac{1}{2\pi} \int_{-A/2}^{A/2} g(x) e^{-ix\omega} dx. \tag{1.59}$$

The aim is to approximate this integral numerically. For that purpose, divide the interval $[-A/2, A/2]$ into N subintervals of length $\Delta x = A/N$. Assume for the moment that N is even. That is, the grid we define has $N+1$ equally spaced points $x_n = n\Delta x$ for $n = -N/2, \ldots, N/2$:

$$x_{-N/2} = -A/2, \ldots, x_0 = 0, \ldots, x_{N/2} = A/2.$$

Assume that the fuction g is known at these grid points and name the integrand

$$h(x) = g(x) e^{-ix\omega}.$$

Then, the trapezoid rule leads to the approximation:

$$\hat{g}(\omega) = \frac{1}{2\pi} \int_{-A/2}^{A/2} h(x) dx \approx \frac{1}{2\pi} \frac{\Delta x}{2} \left\{ h\left(-\frac{A}{2}\right) + 2 \sum_{n=-\frac{N}{2}+1}^{\frac{N}{2}-1} h(x_n) + h\left(\frac{A}{2}\right) \right\}.$$

With the assumption that $g(-A/2) = g(A/2)$, the trapezoid rule approximation may be written as

$$\hat{g}(\omega) = \frac{1}{2\pi} \int_{-A/2}^{A/2} h(x) dx \approx \frac{\Delta x}{2\pi}; \qquad \sum_{n=-\frac{N}{2}+1}^{\frac{N}{2}} h(x_n) = \frac{A}{2\pi N} \sum_{n=-\frac{N}{2}+1}^{\frac{N}{2}} g(x_n) e^{-ix_n\omega}.$$

Consider N irregularly spaced data $x_n = n\Delta x$ and denote $g_n = g(x_n)$. Let N be an even positive integer and let g_n be a sequence of N complex numbers where $n = -N/2+1, \ldots, N/2$. Then, its Discrete Fourier Transform is another sequence of N complex numbers given by

$$\begin{aligned} G_k = G(\omega_k) &= \frac{1}{N} \sum_{n=-N/2+1}^{N/2} g_n e^{-ix_n\omega_k} \\ &= \frac{1}{N} \sum_{n=-N/2+1}^{N/2} g_n e^{-2\pi ikn/N}; \; k = -N/2+1, \ldots, N/2 \end{aligned}$$

$x_n\omega_k = 2\pi kn/N$	
$x_n = n\frac{A}{N}$	$\omega_k = k\frac{2\pi}{A}$
$\Delta x = \frac{A}{N}$	$\Delta\omega = \frac{2\pi}{A}$
$\Delta x\Delta\omega = \frac{2\pi}{N}$	
$A\Omega = 2\pi N$	

Table 1.2: Spatial and spectral domain relations.

where $\omega_k = \frac{2\pi k}{\Delta x N} \in [-\Omega/2, \Omega/2]; k = -N/2+1, \ldots, N/2$, and $\Omega = \dfrac{2\pi N}{A}$.

If N is an odd positive integer and g_n is a sequence of N complex numbers where $n = -(N-1)/2, \ldots, (N-1)/2$, the Discrete Fourier Transform is given by

$$\begin{aligned} G_k &= \frac{1}{N}\sum_{n=-(N-1)/2}^{(N-1)/2} g_n e^{-ix_n\omega_k} \\ &= \frac{1}{N}\sum_{n=-(N-1)/2}^{(N-1)/2} g_n e^{-2\pi ikn/N}, \quad \text{for } k = -(N-1)/2, \ldots, (N-1)/2. \end{aligned}$$

An alternate form of the Discrete Fourier Transform, without regarding whether N is odd or even,

$$G_k = \frac{1}{2\pi N}\sum_{n=0}^{N-1} g_n e^{-ink/N}, \quad \text{for } k = 0, \ldots, N-1. \tag{1.60}$$

The equivalence with the centered expressions is obtained assuming that g is A-periodic, $g(x+sA) = g(x); x \in [-A/2, A/2]$.Then, $g_{-n} = g(-x_n) = g(A - x_n) = g_{N-n}$ (recall that $e^{-2\pi ik(N-n)/N} = e^{-2\pi ikn/N}$).

The Inverse Discrete Fourier Transform is given by:

$$g_n = \sum_{k=-N/2+1}^{N/2} G_k e^{ink/N}, \quad \text{for } n = -N/2+1, \ldots, N/2$$

if N is even, and if N is odd:

$$g_n = \sum_{k=-(N-1)/2}^{(N-1)/2} G_k e^{ink/N}, \quad \text{for } n = -(N-1)/2, \ldots, (N-1)/2.$$

It is important to remark the relationship between the spectral and spatial scales. These relations are summarised in Table 1.2. It can be said that there exists an inverse relation between

both scales. For instance, when decreasing the spacing in the spatial domain, the spectral spacing increases and vice versa. Besides, we could also say that the increasing domain asymptotics in the spatial context corresponds to infilling asymptotics in the spectral case.

As we noted at the beginning of this section, if we assume the g has a compact support, we have $g(\omega_k) \simeq \frac{A}{2\pi} G_k$. When the function g does not vanish outside $[-A/2, A/2]$, a limit reasoning must be made, letting $N \to \infty$ and $\Delta x \to 0$, with ω_k held constant. The error in this approximation is a truncation error due to the fact that the interval of integration $(-\infty, \infty)$ has been truncated. Therefore, a second limit is required to eliminate this error and recover the exact value $\hat{g}(\omega_k)$ we must let $A \to \infty$:

$$\lim_{\substack{A \to \infty \\ \Delta\omega \to 0}} \lim_{\substack{N \to \infty \\ \Delta x \to 0}} \frac{\Delta x}{2\pi} \sum_{n=-N/2+1}^{N/2} g_n e^{-i x_n \omega_k} = \hat{g}(\omega_k).$$

And for the Inverse Transform:

$$\lim_{\substack{\Omega \to \infty \\ \Delta x \to 0}} \lim_{\substack{N \to \infty \\ \Delta\omega \to 0}} \frac{2\pi}{N\Delta x} \sum_{k=-N/2+1}^{N/2} G_k e^{i x_n \omega_k} = g(x_n).$$

The Fast Fourier Transform.
The direct computation of the Discrete Fourier Transform involves $\mathcal{O}(N^2)$ operations (complex additions and multiplications), since it one must compute the product of an $N \times N$ matrix by a vector. The Fast Fourier Transform refers to an efficient algorithm for computing the usual Discrete Fourier Transform. The classical algorithm is due to Cooley and Tukey (1965) and applies in the case in which $N = 2^m$, reducing the computational cost to $\mathcal{O}(N \log_2 N)$ operations. In practice, usually the data series is extended to the closest power 2 number by zero padding. This may be result in a additional computational cost. For instance, if $N = 70000$, the closest power 2 number is $2^{17} = 131072$. In some cases, it may be preferable to consider a modification of this algorithm (e.g. FFTPACK library, http://www.netlib.org/fftpack/). Some of these modifications result more efficient when N is a product of small prime factors. Therefore, N values should be approximated to the higher closest k-smooth number (with prime factors $\leq k$), where k is a small integer number (e.g. $k = 5$ corresponds to Hamming numbers or ugly numbers). Hamming numbers can be easily obtained by a similar algorithm to that proposed by Dijkstra (1976). The set of 5-smooth numbers is more dense that the 2-smooth ones (for instance, between 519 and 1024 there are 16 Hamming numbers), with the consequent reduction of computational cost. If N are not chosen carefully, the FFT algorithm may require N^2 operations. Some of these modifications

can be found in Press *et al.* (1989).

As a final comment, FFT algorithms has been designed for the computation of non-centered expressions such that (1.60). In order to obtain centered DFTs, we must use the so-called wrap-around order, defining $g_{N-n} = g_{-n}$.

1.3.3 The Hankel Transform.

Not only the Fourier Transform or its discrete version have played an important role in spatial spectral analysis. (Stein (1999), pp. 44-46) notes that isotropic autocovariance functions can be characterized, based on the representation of the covariance as the Hankel Transform of a spectral density, under certain regularity conditions.

Definition. *The Hankel Transform of order $\nu \in \mathbb{R}$ of the function $g : \mathbb{R}^+ \to \mathbb{C}$ is defined by:*

$$\mathcal{H}_\nu(g)(\omega) = \int_0^\infty g(x) J_\nu(x\omega) x dx, \tag{1.61}$$

where J_ν is the Bessel function of the first kind of order ν. (See Abramowitz and Stegun (1965)).

Theorem. *Let $g : \mathbb{R}^+ \to \mathbb{R}$ be a function such that $x \to \sqrt{x} g(x)$ belongs to $L^1(0, \infty)$. Then, for each real number $\nu > 1/2$, the Hankel Transform $\mathcal{H}_\nu(g)$ exists and*

$$\int_0^\infty \mathcal{H}_\nu(g)(\omega) J_\nu(x\omega) d\omega = \frac{1}{2}(g(x^+) + g(x^-)),$$

where $g(x^+)$ (respectively, $g(x^-)$) denotes the limit from the right.

Corollary. *If g is continuous at x_0, then*

$$\int_0^\infty \mathcal{H}_\nu(g)(\omega) J_\nu(x_0\omega) d\omega = g(x_0).$$

In the case in which a function $g : \mathbb{R}^d \to \mathbb{C}$ is invariant under rotations, the Fourier Transform can be written as a function of quantities invariant under rotation:

$$\hat{g}(\rho) = \frac{(2\pi)^{d/2}}{\rho^{d/2-1}} \int_0^\infty g(r) r^{d/2} J_{d/2-1}(\rho r) dr.$$

The transformation $r^{d/2-1} g(r) \to (2\pi)^{-d/2} \rho^{d/2-1} \hat{g}(\rho)$ is the Hankel Trasnform of order $d/2 - 1$. This result is specially useful for isotropic spatial covariances, where $d = 2$, so the Hankel Transform is defined in terms of the Bessel function of the first kind of order 0.

1.3.4 Spectral representation of stochastic processes.

We restrict our attention on spatial processes $\{Z(\mathbf{s}), \mathbf{s} \in \mathbb{R}^2\}$. Every weakly-stationary process can be represented in the form of a Fourier-Stieltjes integral:

$$Z(\mathbf{s}) = \int_{\mathbb{R}^2} e^{i\mathbf{x}^T\boldsymbol{\omega}} \mathcal{Y}(d\boldsymbol{\omega}), \tag{1.62}$$

(see Yaglom (1987), pp.98-100) where $\mathcal{Y}$ is an orthogonal complex random measure. This is called the spectral representation of Z. Instead of this interval function, one could also consider a random point function $Y(\boldsymbol{\lambda}) = Y((-\infty, \boldsymbol{\lambda}])$, where $(-\infty, \boldsymbol{\omega}] = (-\infty, \omega_1] \times (-\infty, \omega_2]$, and representation (1.62) can be written in the form:

$$Z(\mathbf{s}) = \int_{\mathbb{R}^2} e^{i\mathbf{x}^T\boldsymbol{\omega}} d\mathcal{Y}(\boldsymbol{\omega}). \tag{1.63}$$

This point function $\mathcal{Y}$ is a random function with uncorrelated increments. It is known the identity between the class of continuous covariance functions on $\mathbb{R}^d$ and the class of positive definite functions in $\mathbb{R}^d$, given by Bochner's theorem. Characteristic functions Φ of probability distributions in $\mathbb{R}^d$ can be identified with positive definite continuous functions satisfying $\Phi(\mathbf{0}) = 1$. Khinchin's theorem establishes that a continuous real function C defined in $\mathbb{R}^d$ is a covariance function if and only if it is the Inverse Fourier Transform of a positive bounded symmetric measure $F(d\boldsymbol{\omega})$, that is, for $d = 2$:

$$C(\mathbf{u}) = \int_{\mathbb{R}^2} e^{i\mathbf{u}^T\boldsymbol{\omega}} F(d\boldsymbol{\omega}), \quad \text{with } \int_{\mathbb{R}^2} F(d\boldsymbol{\omega}) < \infty. \tag{1.64}$$

The integral $\int F(d\boldsymbol{\omega})$ of the spectral measure is equal to the total power $C(\mathbf{0})$, that is, the variance of the process, σ^2. If C is the covariance of the process $Z(\mathbf{s})$, and this process admits representation (1.62), then:

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

Besides, if C decays sufficiently rapidly to ensure that $C \in L^1(\mathbb{R}^2)$, the measure F is the integral of a bounded continuous function $f(\boldsymbol{\omega})$, which is called the spectral density of the process:

$$F(d\boldsymbol{\omega}) = f(\boldsymbol{\omega}) d\boldsymbol{\omega}.$$

Therefore, the spectral density can be seen as the Fourier Transform (in the sense we have seen in Section 1.3.1) of the covariance, and these functions form a Fourier pair (see (1.50)):

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} C(\mathbf{u}) e^{-i\mathbf{u}^T\boldsymbol{\omega}} d\mathbf{u}, \quad \boldsymbol{\omega} \in \mathbb{R}^2,$$

$$C(\mathbf{u}) = \int_{\mathbb{R}^2} f(\boldsymbol{\omega}) e^{i\mathbf{u}^T\boldsymbol{\omega}} d\boldsymbol{\omega}.$$

All we have written above involve continuous processes, that is, processes that may be observed continuously over the space. If we turn now our attention to discrete parameter stationary

processes, i.e. $Z(\mathbf{s})$ which only take values at a discrete set of points, say $\mathbf{s} \in \mathbb{Z}^2$, the covariance function is now defined for $\mathbf{u} \in \mathbb{Z}^2$. If the covariance values form an absolutely summable sequence, then we can define its Fourier Transform as (see (1.53)):

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{u}\in\mathbb{Z}^2} C(\mathbf{u}) e^{-i\mathbf{u}^T\boldsymbol{\omega}}, \quad \boldsymbol{\omega} \in \Pi^2,$$

and the covariance function at a lag $\mathbf{u}$ can be recovered by Inverse Fourier Transform as:

$$C(\mathbf{u}) = \int_{\Pi^2} f(\boldsymbol{\omega}) e^{i\mathbf{u}^T\boldsymbol{\omega}} d\boldsymbol{\omega}.$$

The analogue of Khinchin's theorem for discrete parameter processes is known as Wold's theorem. In time series setting, some authors prefer working with the normalized spectra. The normalized power spectral density is defined by

$$h(\boldsymbol{\omega}) = \frac{1}{\sigma^2} f(\boldsymbol{\omega})$$

where σ^2 denotes the variance of the process. The use of the normalized version of the spectral density may help to clarify the physical interpretation of the process and $\int h(\boldsymbol{\omega}) d\boldsymbol{\omega} = 1$. In the same way, the normalized integrated spectrum H is given by:

$$H(\boldsymbol{\omega}) = \int_{-\infty}^{\omega_1} \int_{-\infty}^{\omega_2} f(\boldsymbol{\lambda}) d\boldsymbol{\lambda}.$$

The properties of the normalized integrated spectrum are similar to those of a distribution function. For this reason, H is sometimes referred to as the spectral distribution function, although throughout this manuscript, by abuse of notation, we may referred to F in the same way.

For intrinsic stationary processes, a spectral representation can be obtained for the variogram. Let g be a positive function in $\mathbb{R}^d$, $g(\mathbf{0}) = 0$. Then, the following assertions are equivalent:

1. g is conditional semidefinite negative;
2. $e^{-tg(\mathbf{u})}$ is semidefinite positive, $\forall t > 0$;
3. g can be written as:

$$g(\mathbf{u}) = \int_{\mathbb{R}^d} \frac{1 - \cos(\mathbf{u}^T\boldsymbol{\omega})}{\|\boldsymbol{\omega}\|^2} dF(\boldsymbol{\omega}) + Q(\mathbf{u}), \tag{1.65}$$

where $Q \geq 0$ is a quadratic form and dF is a simmetric positive measure, continuous about $\mathbf{0}$ and satisfying:

$$\int_{\mathbb{R}^d} \frac{1}{1 + \|\boldsymbol{\omega}\|^2} dF(\boldsymbol{\omega}) < \infty.$$

By assertion (2), $C(\mathbf{u}) = e^{-tg(\mathbf{u})}$ is a covariogram, $\forall t > 0$. Besides, it can be proved that if $\gamma(\mathbf{u})$ satisfies assertion (3) with $Q = 0$, then γ is the variogram of an intrinsic stationary process

in $\mathbb{R}^d$ (see Cressie (1993), pp.87-88).

Remark. A special remark should be made about the non-stationary case, since spatial processes in environmental science, for instance, may be non-stationary: the spatial dependence structure depends on location. Yaglom and Yaglom (1987) proposes a generalization of the spectral representation (1.63), considering, for a spatial process Z, a corresponding spectral process $\mathcal{Y}$ that does not have uncorrelated increments. Then, the covariance of Z can be given by:

$$C(\mathbf{s},\mathbf{u}) = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \exp(\mathbf{s}^T\boldsymbol{\omega} + \mathbf{u}^T\boldsymbol{\lambda}) d^2G(\boldsymbol{\omega},\boldsymbol{\lambda}),$$

where $G(\boldsymbol{\omega},\boldsymbol{\lambda})$ is given by:

$$G(\boldsymbol{\omega},\boldsymbol{\lambda}) = E(\mathcal{Y}(\boldsymbol{\omega})\overline{\mathcal{Y}(\boldsymbol{\lambda})}).$$

Similarly to the stationary case, if G has a density with respect to the Lebesgue measure, namely g, it will be given by the Fourier Transform of the covariance function. That is:

$$g(\boldsymbol{\omega},\boldsymbol{\lambda}) = \frac{1}{(2\pi)^4}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \exp(-i(\mathbf{s}^T\boldsymbol{\omega} + \mathbf{u}^T\boldsymbol{\lambda}))C(Z(\mathbf{s}),Z(\mathbf{u}))d\mathbf{s}d\mathbf{u}.$$

1.3.5 Aliasing.

If Z is defined over a continuum (Z takes values on any location $s \in D$, that is, geostatistical context), the spectrum lies on $\boldsymbol{\lambda} \in \mathbb{R}^2$. For a discrete process (D is a discrete set of points), the spectrum can be restrited to the bounded support $\Pi^2 = [-\pi,\pi]\times[-\pi,\pi]$.

In practice, we may aim to recover the spectrum of a continuous process from a discrete realization and therefore, despite the frequency band is the whole space $\mathbb{R}^2$, the frequency behaviour we can recover is restricted to $\Pi^2_\Delta = [-\pi/\Delta_1,\pi/\Delta_1]\times[-\pi/\Delta_2,\pi/\Delta_2]$, where Δ_1 and Δ_2 are the spacing between neighbouring coordinates in each direction. This effect is known as aliasing. Aliasing effect has been analized in time series context. See Priestley (1981), p. 224) or Robinson (1976), among others.

The spectral density f of a continuous process Z, can not be completely restored from a discrete set of observations. It is easy to justify just taking into account that complex exponentials in a frequency $\boldsymbol{\lambda}$ and in $\boldsymbol{\lambda}\pm(2\pi,2\pi)$ are undistinguishable. In this case the spectrum of the observations is concentrated on the band Π^2_Δ and the aliased spectral density is given by:

$$f_\Delta(\boldsymbol{\lambda}) = \sum_{m_1=-\infty}^{\infty}\sum_{m_2=-\infty}^{\infty} f\left(\boldsymbol{\lambda} + 2\pi\left(\frac{m_1}{\Delta_1},\frac{m_2}{\Delta_2}\right)^T\right)$$

$$= f(\boldsymbol{\lambda}) + \sum_{m_1 \neq 0} \sum_{m_2 \neq 0} f\left(\boldsymbol{\lambda} + 2\pi \left(\frac{m_1}{\Delta_1}, \frac{m_2}{\Delta_2}\right)^T\right). \tag{1.66}$$

Therefore, if we want to determine the spectral characteristics of a process from a set of observations, the Nyquist frequency $(\pi/\Delta_1, \pi/\Delta_2)$ must be high enough to guarantee that frequencies above it make negligible contribution to the total power of the process.

Aliasing must be taken into account when trying to estimate the spectral density, and also, as we will see later, when we want to simulate a realization of a continuous process on a discrete set of locations.

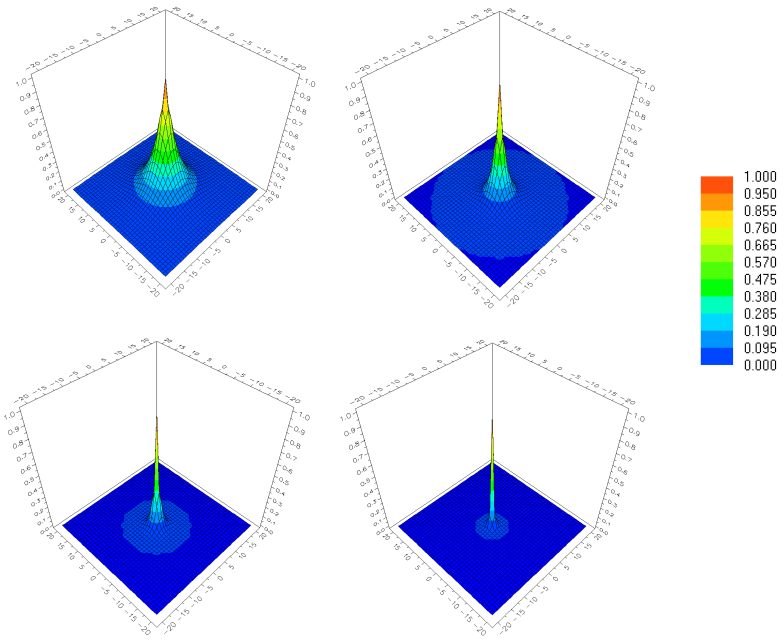

Figure 1.5: Matérn covariances. $c_0 = 0.0$, $\sigma^2 = 1.0$ and $\nu = 0.5$. From left to right and from top to bottom: $\alpha_1 = 0.25$, $\alpha_2 = 0.5$, $\alpha_3 = 1.0$ and $\alpha_4 = 2.0$.

1.3.6 Matérn class of covariances.

Consider a spectral density, for a random field on $\mathbb{R}^d$, of the form:

$$f(\boldsymbol{\lambda}) = \phi(\alpha^2 + \|\boldsymbol{\lambda}\|^2)^{-\nu - d/2}, \quad \phi, \nu, \alpha > 0. \tag{1.67}$$

The corresponding covariance function is given by:

$$C(\mathbf{u}) = \frac{\pi^{d/2}\phi}{2^{\nu-1}\Gamma(\nu + d/2)\alpha^{2\nu}} (\alpha\|\mathbf{u}\|)^{\nu} \mathcal{K}_{\nu}(\alpha\|\mathbf{u}\|), \tag{1.68}$$

where $\mathcal{K}_\nu$ is a modified Bessel function of the second kind. This family of covariances (or spectral densities) is known as the Matérn class of covariances. The corresponding variogram has already been introduced in (1.45). The parameter α can be interpreted as the inverse of the autocorrelation range, ν controls the rate of decay of the spectral density at high frequencies and ϕ is proportional to the variance of the process σ^2 times $\alpha^{2\nu}$:

$$\phi = \sigma^2 \cdot \frac{\alpha^{2\nu}\Gamma(\nu + d/2)}{\Gamma(\nu)\Gamma(d/2)}. \tag{1.69}$$

Some considerations on the parameters must be made. First, the parameter ν given the smoothness of the process since, for instance, Z will be m-times Mean Square differentiable if and only if $\nu > m$ (see Stein (1995), pp.32-33). The particular case $\nu = 1/2$ corresponds to the exponential model (1.40). The Matérn model allows for flexibility in the smoothness of the process, and the number of parameters involved is still manageable. Unfortunately, with the parametrization above, the autocorrelation range is a function of the smoothness and scale parameters, α and ν. There exist other parametrizations which involve paramterers not so highly dependent on the smoothness (see Stein (1995), pp. 49-50).

In Figures (1.5) and (1.6) we have plotted the Matérn covariances and the corresponding log-spectral densities, for smoothness parameter $\nu = 0.5$. We have fixed no nugget effect and unit variance. We can see that, as the parameter α increases, or equivalently, as the autocorrelation range decreases, the covariance becomes more concentrated around zero. The inverse behaviour can be seen for the log-spectral densities. Results are shown in the logarithmic scale since, for the parameters we have chosen, visualization is easier after a logarithmic transform.

1.4 Real data examples

In this section we will introduce two real data sets. The first one, Mercer and Hall wheat data, is a classical example in spatial statistics. The second data set has been kindly provided by the Department of Ecology and Celular Biology of the University of Santiago de Compostela.

These two datasets will be considered along the manuscript for illustration purposes.

1.4.1 Mercer and Hall wheat data (a classical example).

Mercer and Hall experiment carried out in 1911 consisted of a uniform trial on a field of wheat (all the plots received the same treatment) on an area of one acre. Yields of grain were measured

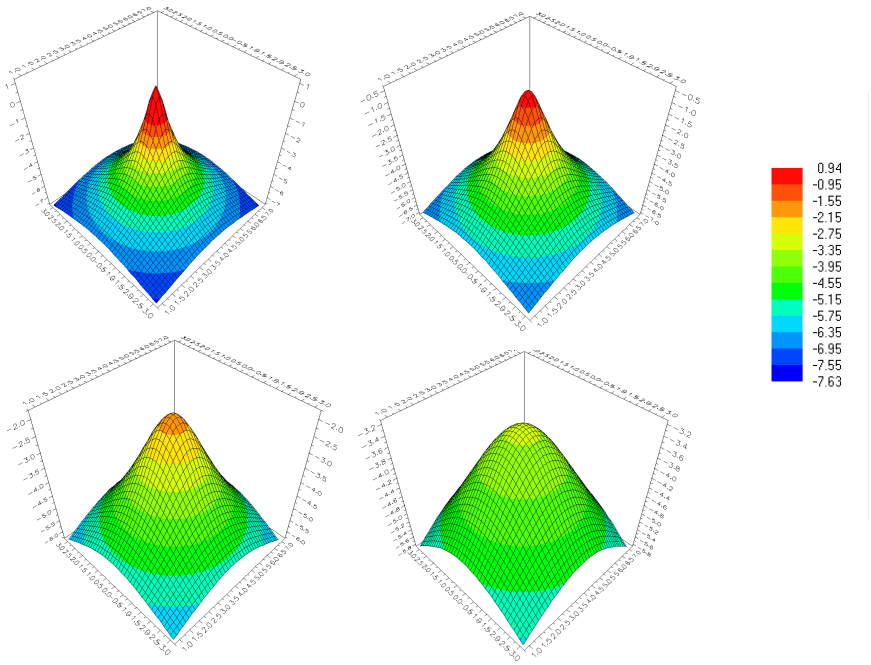

Figure 1.6: Matérn log-spectral densities. $c_0 = 0.0$, $\sigma^2 = 1.0$ and $\nu = 0.5$. From left to right and from top to bottom: $\alpha_1 = 0.25$, $\alpha_2 = 0.5$, $\alpha_3 = 1.0$ and $\alpha_4 = 2.0$.

in pounds. On the 20×25 layout, each of the 20 rows runs in the E-W direction, and each of the 25 columns runs in the N-S direction. Although the exact size of the plots from the original data set seems to be unknown, some researchers have used 3.30 meters east to west, and 2.51 meters north to south. Histogram for these data is plotted in Figure 1.7. This dataset has been broadly studied by different authors (Whittle (1954), Cressie (1993), Young and Young (1998)). Young and Young (1998) conducted an exploratory data analysis on these data and Cressie (1993) shows that data indicate an irregular east-west trend. Some authors have considered lattice data models in order to analyze this dataset but, although the data are given on a spatial lattice, we may think about wheat yields on potential plots located bewteen observed plots. Therefore, a geostatistical analysis could be also conducted (see Cressie (1993), pp. 248-259).

A prediction surface for this dataset was constructed considering model (1.21) with trend function as in (1.22), and $f_0 = 1$, $f_j(\mathbf{s}) = s_j$ with $j = 1, 2$ and exponential variogram for ε. In order to obtain a variogram estimation, we use the iterative algorithm proposed by Neuman and Jacobson (1984). In the universal kriging interpolation method, the estimated parameters are $\boldsymbol{\beta} = (4.1484, -0.144, -0.0035)^T$. The adjusted valid variogram model (by WLS) was an exponential variogram (1.40), with parameters $c_0 = 0.079$, $c_1 = 0.118$ and practical range $a = 4.663$. Empirical semivariogram estimator (1.34) and exponential variogram fit for Mercer and Hall data are plotted in Figure 1.8. In Figure 1.9, we show the number of pairs contributing at each lag

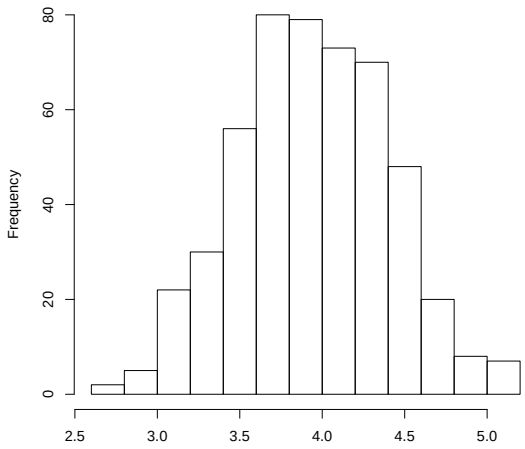

Figure 1.7: Histogram for Mercer and Hall wheat yield data.

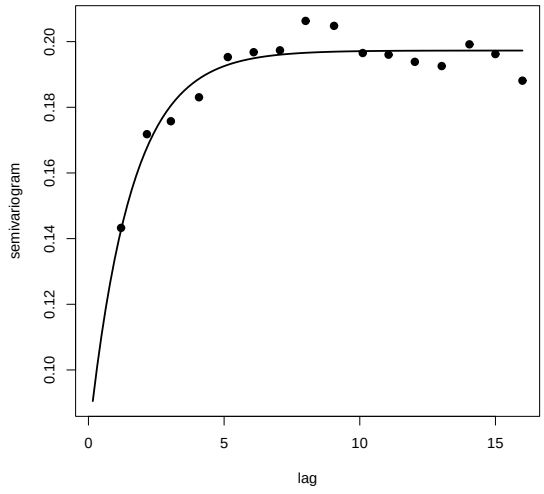

Figure 1.8: Empirical and exponential variogram fit for Mercer and Hall wheat yield data. Solid line: exponential semivariogram. Dots: empirical semivariogram.

in the semivariogram estimation. We have also computed a robust empirical variogram estimator (1.36), with similar results. Results are plotted in Figure 1.10.

In Figure 1.11 we can see the universal kriging surface for Mercer and Hall data. In Figure 1.12 we show the corresponding kriging variance surface.

1.4.2 Heavy metal depositions in Galicia.

One method used for large-scale monitoring of long-range transport is the moss technique, which was developed in Sweden in the late 1960s as a means of surveying atmospheric metal deposition. Following this technique, biomonitoring studies have been hold over the last years in order to determine levels of heavy metal concentration all over Europe. The use of mosses as biomonitors has been proved to be a useful way of determining levels of atmospheric deposition since the uptake of heavy metals in mosses occurs mainly from this source.

The accumulation of heavy metals over large areas and long time periods may cause chronic damage to living organisms and it must be thoroughly controlled. In the particular case of Galicia (NW Spain), mosses have been used as biomonotors. This technique was first used in Galicia in 1995 (Fernández *et al.* (2000)) and measurements of heavy metal concentrations have been taken every two years. In 1995 and 1997, observation points where selected attending a kind of *preferential* sampling. Most locations where placed in the north part of Galicia, where two electricity power plants are located.

In 2000, a new sample grid of regularly spaced locations was considered (see Figure 1.13). In 2004, samples of two types of mosses, *Scleropodium purum* and *Hypnum cupressiforme* were collected on a grid with 148 points. The sampling locations spread over the entire region (29434km^2) and the limiting area. The sampling stations are regularly spaced on a 15×15 lattice. Besides, samples were collected at least 300 meters away from main roads and urban areas, and alt least 100 meters apart from other kind of roads and isolated areas. Concentrations of different heavy metals were measured: Al, As, Co, Cr, Cu, Fe, Hg, Ni, Pb, Se and Zn. For more details, see Fernández *et al.* (2000). From the collection of measured metals above, we will consider data corresponding to Selenium and Mercury (in parts per billion) concentrations in 2004.

Although Selenium is an essential trace element in humans it is toxic if taken in large doses. Symptoms of selenosis (intoxication by Se) include gastrointestinal disorders, fatigue and neurological damage, among others. Extreme cases of selenosis can result in cirrhosis of the liver, pulmonary edema and death. Selenium poisoning of water systems may result whenever new

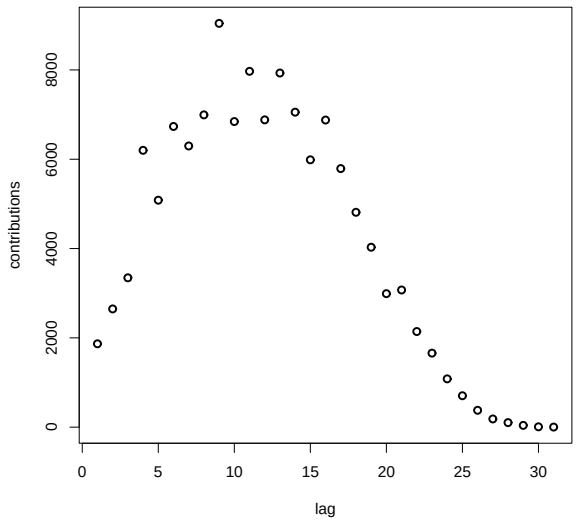

Figure 1.9: Number of contributions in each lag for Mercer and Hall wheat yield data.

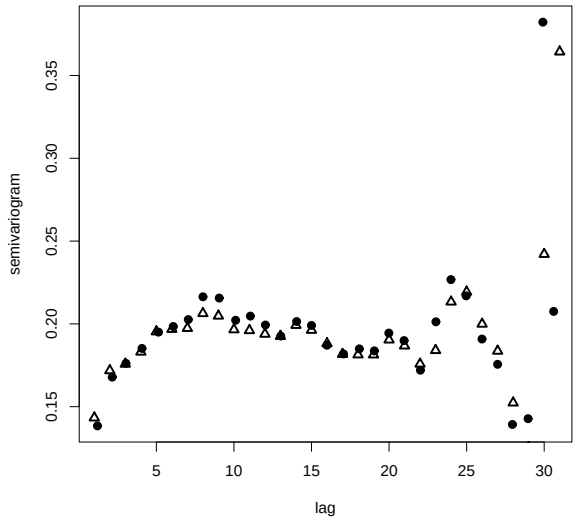

Figure 1.10: Classical and robust variogram estimators for Mercer and Hall wheat yield data. Circles: classical semivariogram (1.34). Triangles: robust semivariogram (1.36).

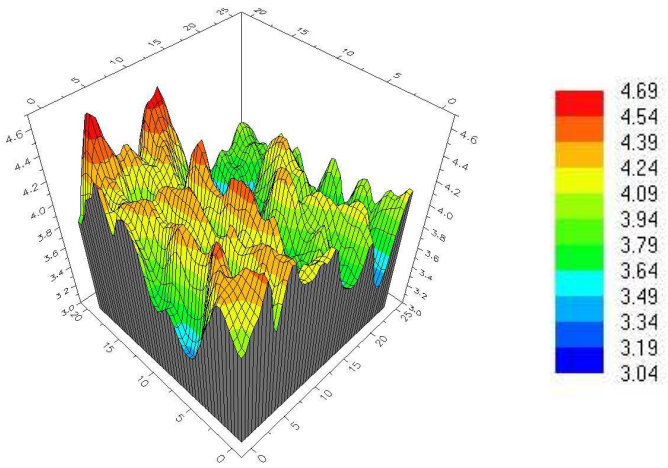

Figure 1.11: Universal kriging surface with linear trend, for Mercer and Hall wheat yield data.

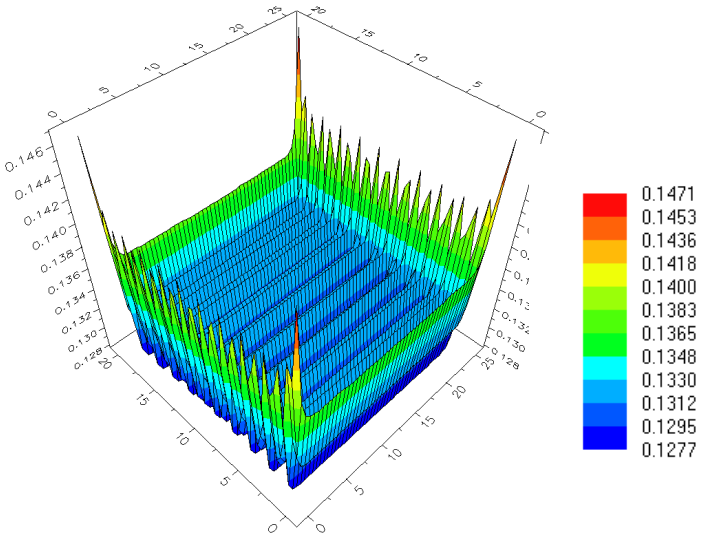

Figure 1.12: Universal kriging variance surface for Mercer and Hall data.

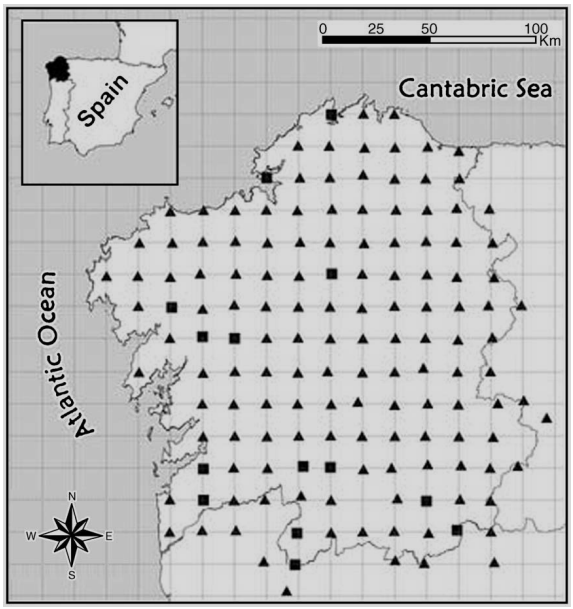

Figure 1.13: Sampling sites in Galicia and limiting area for two types of mosses. Squares: *Scleropodium purum*. Triangles: *Hypnum cupressiforme*.

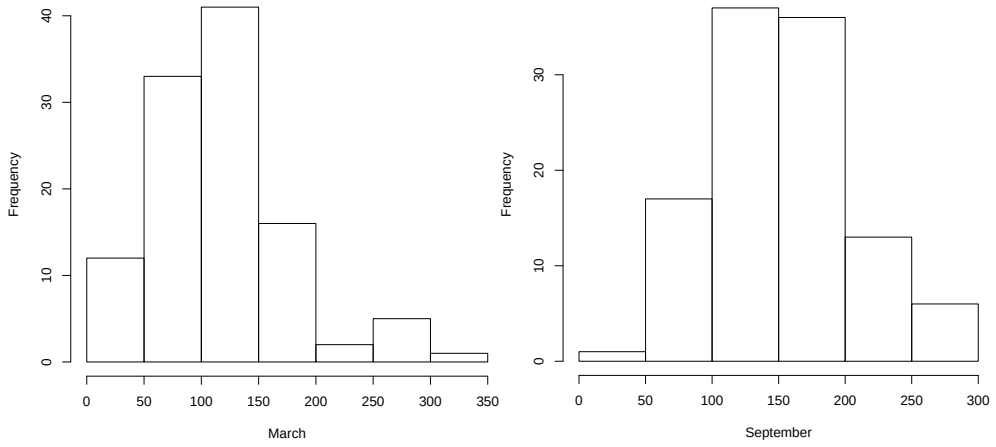

Figure 1.14: Left panel: histogram of $log(Se)$ concentrations in March 2004. Right panel: histogram of $log(Se)$ concentrations in September 2004.

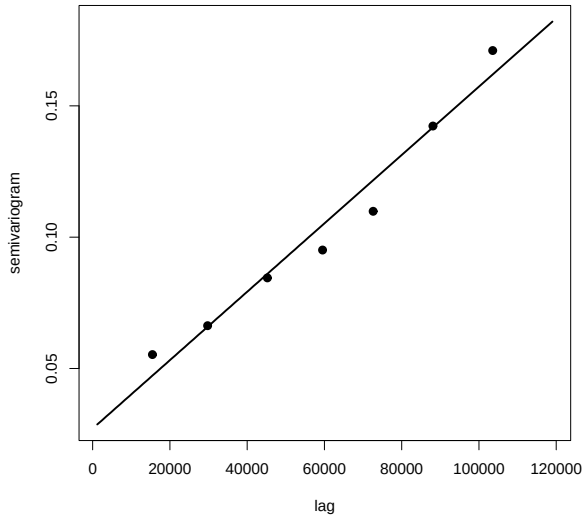

Figure 1.15: Empirical variogram estimation and linear variogram fit for $\log(Se)$ concentrations, March 2004.

agricultural runoff courses through normally-dry undeveloped lands. This process leaches natural soluble Se compounds into the water, which may then be concentrated in new wetlands as it evaporates. In Figure 1.14 we show the distribution of $\log(Se)$ concentrations in March and September 2004. We could see that raw data exhibit asymmetry, and by the logarithmic transform, this asymmetry is corrected.

A linear semivariogram was fitted by WLS over the classical semivariogram estimates provided by (1.34). The estimated parameters were $(c_0, c_1) = (2.71E-02, 1.30E-06)$. Classical semivariogram estimates and linear fit are plotted in Figure 1.15. In Figure 1.16 we show the ordinary kriging surface for logarithms of Se concentrations. The kriging variance surface is plotted in Figure 1.17.

The study of Hg concentrations is particularly interesting since Hg is not a common element in earth's crust. However, since mercury does not blend geochemically with elements in the crustal mass, Hg ores can be highly concentrated. Besides, Hg is a bioaccumulative toxin and it is easily absorbed through the skin, respiratory and gastrointestinal tissues, so the exposure to high Hg concentrations produces toxic effects on human beings.

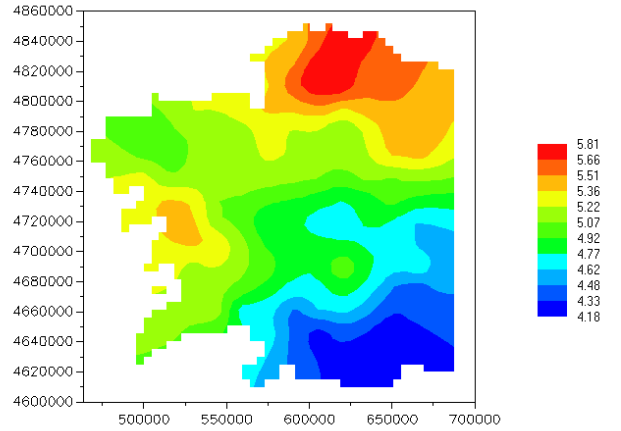

Figure 1.16: Ordinary kriging surface for $\log(Se)$ concentrations, March 2004.

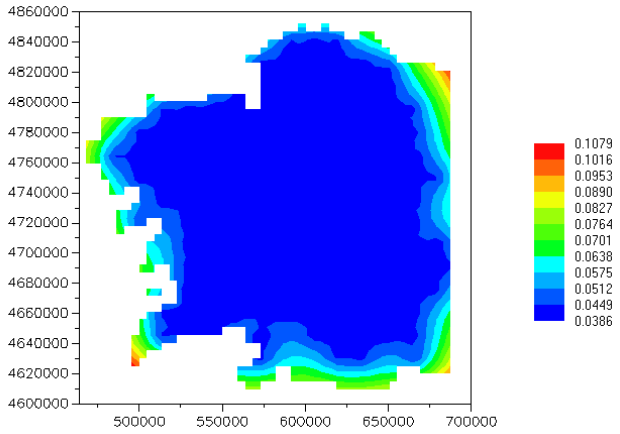

Figure 1.17: Ordinary kriging variance surface for $\log(Se)$ concentrations, March 2004.

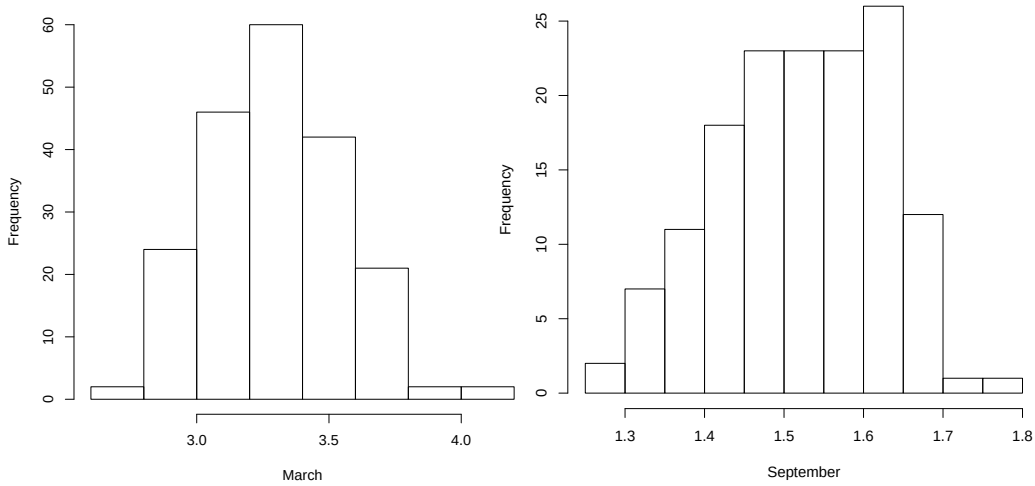

Figure 1.18: Left panel: histogram of $\log(Hg)$ concentrations in March 2004. Right panel: histogram of $log(Hg)$ concentrations in Semptember 2004.

Figure 1.18 show the histograms of the concentrations of Hg in March (left panel) and the logarithmic transformation of the data (right panel). Similar results are obtained for the data corresponding to September sampling. For $\log(Hg)$ concentrations, a linear semivariogram was fitted by WLS over the classical semivariogram estimates provided by (1.34). The estimated parameters were $(c_0, c_1) = (4.452E-021.455E-07)$. Classical and robust semivariogram estimates are plotted in Figure 1.19.

In Figure 1.20 we show the ordinary kriging surface for $\log(Hg)$ concentrations. The kriging variance surface is plotted in Figure 1.21. For Hg we see a hot-spot in the south-west of the region, which does not appear for Se concentrations.

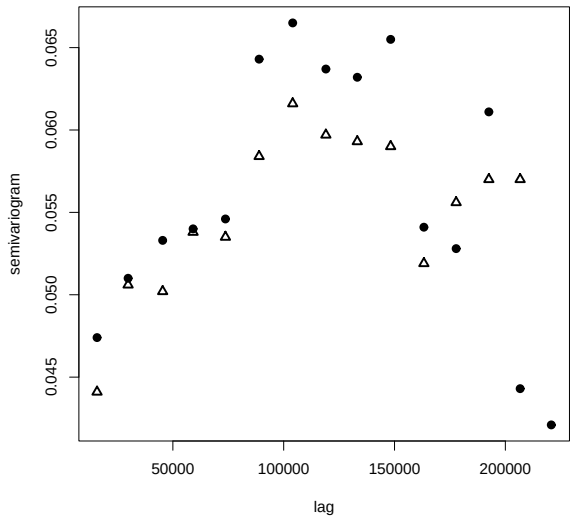

Figure 1.19: Classical and robust variogram estimators for $\log(Hg)$ concentrations, March 2004. Circles: classical semivariogram (1.34). Triangles: robust semivariogram (1.36).

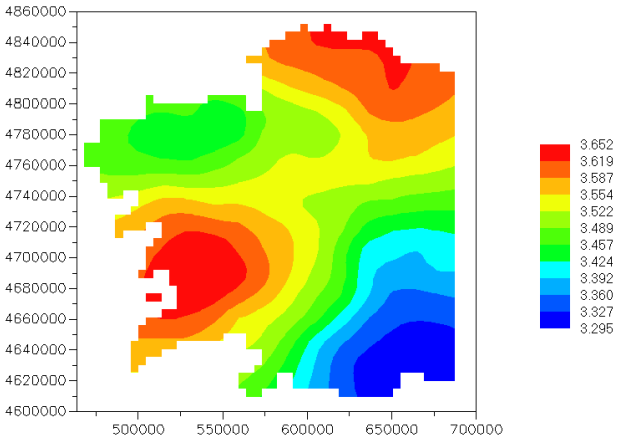

Figure 1.20: Ordinary kriging surface for $\log(Hg)$ concentrations, March 2004.

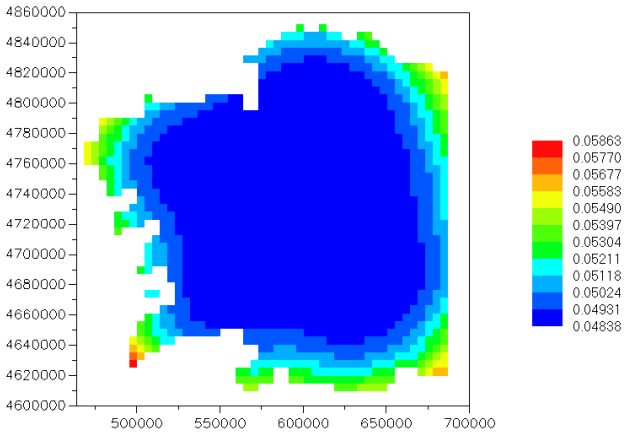

Figure 1.21: Ordinary kriging variance surface for $\log(Hg)$ concentrations, March 2004.

Chapter 2

Spectral techniques for modeling spatial dependence.

Contents

Describing the dependence structure of spatial random processes has been a major topic of discussion in spatial statistics. For geostatistical data, the study of the dependence features has been mainly done by modelling the variogram or the covariogram function associated with the spatial, or even more generally, spatio-temporal, process. Some modelling techniques for the variogram have been introduced in Chapter 1. In recent years, some authors have developed spectral methods in this context (e.g. Stein (1995)), considering the Fourier Transform of the covariogram, the spectral density, as the new target function.

The classical nonparametric estimator of the spectral density is, in time series context, the periodogram. This estimator has been extended to the spatial setting. In fact, the spatial spectral techniques have been inspired in the time series spectral analysis developed in Hannan (1970), Brillinger (1981) or Priestley (1981), among others. In the spatial setting, Stein (1995) investigated periodogram properties for stationary processes under a fixed-domain asymptotic framework. Fuentes (2002) studied asymptotic periodogram properties and proposes a nonstationary periodogram, under shrinking asymptotics. Fuentes (2006a) introduces a modification of Whittle's approximation to the Gaussian loglikelihood, for spatial regular lattices with missing values and for irregularly spaced datasets.

Although asymptotically unbiased, the periodogram is not consistent because its variance at each spectral frequency is proportional to the square of the density at these frequencies. We study a class of consistent estimates of the spectral density for geostatistical processes: smoothed nonparametric kernel spectral estimates. We will refer to these estimators as smoothed-covariances periodograms. These kind of estimates have been broadly studied in time series (Hannan (1970), Priestley (1981)) and in spatial lattice data processes (Robinson (2006)). The basic idea of the smoothed kernel non parametric estimators for the spatial spectral density is to soften the influence of the sample covariance estimate, damping its value for $\mathbf{u}$ with large $|\mathbf{u}|$ by the introduction of a lag window. Examples of such functions will be introduced in this chapter. We confront this problem from the continuous setting, when the dependence structure of a geostatistical process has to be recovered from a discrete spectral signal.

In some cases, the edge-effect bias may not be negligible. We have to deal not only with an edge-effect, but also with the aliasing problem. We observe that bias can be dominated by the edge-effect but, as it happens for lattice processes, this problem can also be mitigated by tapering. The choice of the kernel, bandwidth and spacing parameter presents some implications for the edge-effect bias.

Although not in the scope of our work, parametric spectral density estimation is another al-

ternative. Some remarks on this topic will be made at the end of this chapter.

2.1 The spatial periodogram.

The periodogram (also called sample spectral density) is a classical nonparametric estimator of the spectral density. For a spatial process Z observed on a regular grid $D = \{\mathbf{s} = (s_1, s_2) : s_1 = 0, \ldots, n_1 - 1, s_2 = 0, \ldots, n_2 - 1\}$, $D \subset \mathbb{R}^2$, with $N = n_1 n_2$ points, the spatial periodogram at a frequency $\boldsymbol{\lambda}$ is given by:

$$I(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2 N} \left| \sum_{\mathbf{s} \in D} Z(\mathbf{s}) e^{-i\mathbf{s}^T \boldsymbol{\lambda}} \right|^2, \quad \boldsymbol{\lambda} \in \Pi^2. \tag{2.1}$$

Assume that the spacing between locations is given by Δ_1 in one direction (x-axis direction) and Δ_2 in the other direction (y-axis direction). Then, the frequency band is $\boldsymbol{\lambda} \in \Pi_\Delta^2 = [-\pi/\Delta_1, \pi/\Delta_1] \times [-\pi/\Delta_2, \pi/\Delta_2]$. In this case, the spacing must be taken into account when defining the periodogram. Denote by $\Delta\mathbf{s} = (\Delta_1 s_1, \Delta_2 s_2)$. Then, the spatial periodogram is given by:

$$I_\Delta(\boldsymbol{\lambda}) = \frac{\Delta_1 \Delta_2}{(2\pi)^2 N} \left| \sum_{\mathbf{s} \in D} Z(\Delta\mathbf{s}) e^{-i\Delta\mathbf{s}^T \boldsymbol{\lambda}} \right|^2. \tag{2.2}$$

The periodogram is usually computed at the set of bidimensional Fourier frequencies $\boldsymbol{\lambda}_{\mathbf{k}}^T = (\lambda_{k_1}, \lambda_{k_2})$:

$$\begin{aligned} \lambda_{k_1} &= \tfrac{2\pi k_1}{n_1}, \quad k_1 = 0, \pm 1, \ldots, \pm m_1, \quad \text{where } m_1 = [(n_1 - 1)/2], \\ \lambda_{k_2} &= \tfrac{2\pi k_2}{n_2}, \quad k_2 = 0, \pm 1, \ldots, \pm m_2, \quad \text{where } m_2 = [(n_2 - 1)/2] \end{aligned}$$

for (2.1) and for (2.2)

$$\begin{aligned} \lambda_{k_1} &= \tfrac{2\pi k_1}{\Delta_1 n_1}, \quad k_1 = 0, \pm 1, \ldots, \pm m_1, \quad \text{where } m_1 = [(n_1 - 1)/2], \\ \lambda_{k_2} &= \tfrac{2\pi k_2}{\Delta_2 n_2}, \quad k_2 = 0, \pm 1, \ldots, \pm m_2, \quad \text{where } m_2 = [(n_2 - 1)/2]. \end{aligned}$$

For (2.1), it is known that the spatial periodogram is an asymptotically unbiased estimator of the spectral density, but it is not consistent (Brilinger (1974)), since the variance is proportional to the square of the spectral density at each frequency. Nevertheless, the periodogram characteristic which makes it useful is that periodogram values at different frequencies are asymptotically uncorrelated. This fact allows for dealing with periodogram values as if they were independent data, for a large enough sample.

In the geostatistical data case, Fuentes (2002) proved that the assertions above (asymptotic unbiasedness, inconsistency and asymptotic uncorrelation) also hold for (2.2), with two further

assumptions: (1) the rate of decay of the spectral density $f(\boldsymbol{\lambda})$ is proportional to $\|\boldsymbol{\lambda}\|^{-\tau}$, for $\tau > 2$ and (2) $n_1, n_2 \to \infty$, $\Delta_1 = \Delta_2 = \Delta \to 0$, $n_1/n_2 \to c$ (for a constant c) and $\Delta n_1, \Delta n_2 \to \infty$.

This type of asymptotics is called shrinking asymptotics (Fuentes (2002)) and it is a mixture between increasing-domain (Cressie (1993)) and fixed-domain asymptotics (Stein (1995)). In an increasing-domain asymptotic analysis, one assumes that the number of observations grows, preserving a minimum positive distance between the locations where the observations are taken. In fixed-domain asymptotics, the number of observations grows filling in the observation region.

Theorem. *(Fuentes, 2002) Let Z be a second-order stationary process, with spectral density f and assume that:*

(i) The rate of decay of the spectral density $f(\boldsymbol{\omega})$ at high frequencies is proportional to $\|\boldsymbol{\omega}\|^{-\tau}$, for $\tau > 2$.

(ii) The covariance function satisfies the inequality $\int \|\mathbf{u}\||C(\mathbf{u})|d\mathbf{u} < \infty$.

Then, under shrinking asymptotics:

(i) The expected value of the periodogram $I(\boldsymbol{\omega})$, for $\boldsymbol{\omega} \in [-\pi/\Delta, \pi/\Delta]^2$ is asymptotically $f(\boldsymbol{\omega})$.

(ii) The asymptotic variance of $I(\boldsymbol{\omega})$ given in (2.2) is $f^2(\boldsymbol{\omega})$.

(iii) The periodogram values $I(\boldsymbol{\omega})$ and $I(\boldsymbol{\lambda})$, for $\boldsymbol{\omega} \neq \boldsymbol{\lambda}$ are asymptotically uncorrelated.

If we observe a continuous process and we consider a different asymptotic framework, for instance, increasing domain asymptotics, then, the periodogram is no longer an unbiased estimator of the spectral density but for f_Δ, the aliased spectral density given in (1.66).

Remark. Fuentes (2002) also studied the non-stationary case, considered in Section 1.3. The author defines the periodogram for a non-stationary spatial process Z and proved that, under shrinking asymptotics, the asymptotic expected value of the periodogram at a pair of frequencies $(\boldsymbol{\omega}, \boldsymbol{\lambda})$ is the spectral density $f(\boldsymbol{\omega}, \boldsymbol{\lambda})$. Expressions for the asymptotic variance and the covariance are also given. In order to get these results under non-stationarity assumptions, some conditions on the rate of decay of the spectral density and the integrability of the covariance are needed.

The spatial periodogram has been defined in terms of the observed data, both in (2.1) and (2.2). Since the spectral density is the Fourier Transform of the covariance function, it is not unnatural

to write the periodogram in terms of a covariance estimator. Define the sample covariances for a lattice and a geostatistical process:

$$\hat{C}(\mathbf{u}) = \frac{1}{N} \sum_{\mathbf{s} \in D(\mathbf{u})} Z(\mathbf{u}) Z(\mathbf{s}+\mathbf{u}), \tag{2.3}$$

$$\hat{C}_{\Delta}(\mathbf{u}) = \frac{1}{N} \sum_{\mathbf{s} \in D(\mathbf{u})} Z(\Delta\mathbf{s}) Z(\Delta(\mathbf{s}+\mathbf{u})), \tag{2.4}$$

where $D(\mathbf{u}) = \{\mathbf{s} \in D : \mathbf{s}+\mathbf{u} \in D\}$. The periodogram can be written in terms of the sample covariances as:

$$I(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{u} \in \mathcal{U}} \hat{C}(\mathbf{u}) e^{-i\mathbf{u}^T\boldsymbol{\lambda}}, \tag{2.5}$$

$$I_{\Delta}(\boldsymbol{\lambda}) = \frac{\Delta_1\Delta_2}{(2\pi)^2} \sum_{\mathbf{u} \in \mathcal{U}} \hat{C}_{\Delta}(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}. \tag{2.6}$$

and $\mathbf{u} \in \mathcal{U} = \{(u_1, u_2) : 1-n_1 \leq u_1 \leq n_1 - 1, 1-n_2 \leq u_2 \leq n_2 - 1\}$. Representations (2.5) and (2.6) for the spatial periodogram will be considered when constructing consistent estimators for the spectral density. These estimators will be based on smoothed versions of the sample covariances.

The periodogram is, essentially the same function for the sample covariances $\hat{C}(\mathbf{u})$ as f is of the theoretical covariances $C(\mathbf{u})$. For a spatial process Z on the Euclidean space $\mathbb{R}^d$, the bias of these estimators is of order $N^{-1/d}$ (see Robinson (2006)). When computing the periodogram at a fixed frequency $\boldsymbol{\lambda}$, it includes all the sample covariances and hence, no matter how large N becomes, it always involves a tail effect.

Note also that, if we define the Discrete Fourier Transform of the data as:

$$J(\boldsymbol{\lambda}) = \frac{1}{2\pi\sqrt{n_1 n_2}} \sum_{\mathbf{s} \in D} Z(\mathbf{s}) e^{-i\mathbf{s}^T\boldsymbol{\lambda}}, \tag{2.7}$$

then, the periodogram can be obtained as:

$$I(\boldsymbol{\lambda}) = J(\boldsymbol{\lambda})\overline{J(\boldsymbol{\lambda})} = |J(\boldsymbol{\lambda})|^2, \tag{2.8}$$

where $\overline{(\cdot)}$ denotes the congujate. We can also obtain a similar representation of I_{Δ}, just modifying (2.7) in a suitable way:

$$J_{\Delta}(\boldsymbol{\lambda}) = \frac{1}{2\pi} \sqrt{\frac{\Delta_1\Delta_2}{n_1 n_2}} \sum_{\mathbf{s} \in D} Z(\Delta\mathbf{s}) e^{-i\Delta\mathbf{s}^T\boldsymbol{\lambda}}. \tag{2.9}$$

Despite its lack of consistency as an estimator of the spectral density, the periodogram (2.1) has an attractive feature for some kind of processes. Consider a spatial processes which can be

represented as:

$$Z(\mathbf{s}) = \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_{jl}\varepsilon(s_1 - j, s_2 - l), \quad \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_{jl}^2 < \infty \tag{2.10}$$

where the error variables ε are independent and identically distributed as $N(0, \sigma_\varepsilon^2)$. The condition on the coefficients a_{jl} is neeeded in order to guarantee stationarity. This kind of structure for an underlying model of an observed set of data holds for any Gaussian, stationary process with absolutely continuous spectral density.

Remark. If ε is a continuous white-noise, a general linear process is given by:

$$Z(\mathbf{s}) = \int a(\mathbf{u})\varepsilon(\mathbf{s} - \mathbf{u})d\mathbf{u} \quad \text{with } \int a^2(\mathbf{u})d\mathbf{u} < \infty, \tag{2.11}$$

in order to guarantee the stationarity of the process. Expression (2.10) can be interpreted as an approximation of a general linear process.

The spectral density of (2.10) at a frequency $\boldsymbol{\lambda} \in \Pi^2$ is given by:

$$f(\boldsymbol{\lambda}) = \left| \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_{jl} e^{-i(j,l)\boldsymbol{\lambda}} \right|^2 \frac{\sigma^2}{(2\pi)^2} = |A(\boldsymbol{\lambda})|^2 f_\varepsilon(\boldsymbol{\lambda}), \tag{2.12}$$

where

$$A(\boldsymbol{\lambda}) = \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_{jl} e^{-i(j,l)\boldsymbol{\lambda}}, \quad f_\varepsilon(\boldsymbol{\lambda}) = \sigma^2/(2\pi)^2.$$

and $(j,l)\boldsymbol{\lambda} = j\lambda_1 + l\lambda_2$. Therefore, f can be written in terms of f_ε, the spectral density of the innovation process. A similar expression is obtained for the periodogram of Z, when its covariance C is axial and diagonal symmetric, which can be written in terms of the periodogram of ε:

$$I(\boldsymbol{\lambda}) = |A(\boldsymbol{\lambda})|^2 I_\varepsilon(\boldsymbol{\lambda}) + R_N(\boldsymbol{\lambda}), \tag{2.13}$$

where, the residual term is uniformly bounded and I_ε denotes the periodogram for ε. Expression (2.13) can be written as

$$I(\boldsymbol{\lambda}_\mathbf{k}) = f(\boldsymbol{\lambda}_\mathbf{k})V_\mathbf{k} + R_N(\boldsymbol{\lambda}_\mathbf{k}) \tag{2.14}$$

where each $\boldsymbol{\lambda}_\mathbf{k}$ denotes a Fourier frequency and $V_\mathbf{k}$'s are independent identically distributed random variables with standard exponential distribution (see Brockwell and Davis (1991)). Then, applying logarithms in (2.14), we have

$$Y_\mathbf{k} = m(\boldsymbol{\lambda}_\mathbf{k}) + z_\mathbf{k} + r_\mathbf{k} \tag{2.15}$$

where $m = \log f$ and

$$r_\mathbf{k} = \log\left[1 + \frac{R_N(\boldsymbol{\lambda}_\mathbf{k})}{f(\boldsymbol{\lambda}_\mathbf{k})V_\mathbf{k}}\right]. \tag{2.16}$$

The variables $z_{\mathbf{k}}$ are independently and identically distributed with density function:

$$h(x) = e^{-e^x+x}. \tag{2.17}$$

The mean is the Euler constant $E(z_{\mathbf{k}}) = C_0 = -0.57721$ and the variance is $\mathrm{Var}(z_{\mathbf{k}}) = \pi^2/6$. This is a particular case of the Gumbel distribution, with position and scale parameters 0 and 1 ($Gum(0,1)$), respectively.

We will take advantage of the representation (2.14) for the periodogram and (2.15) for the log-periodogram in next chapters, for simulation purposes and in order to build goodness-of-fit tests for the spatial spectral density.

Remark. Under additional assumptions on the rate of decay of the coefficients a_{jl} in (2.10), the residual term $R_N(\boldsymbol{\lambda}_{\mathbf{k}})$ in (2.14) can be uniformly bounded (Brockwell and Davis (1991)).

2.2 Some modifications on the periodogram.

In this section, we revise two spectral density estimation approaches, based on modifications of the periodogram. The first one is based on tapering techniques and its goal is to reduce the bias of the periodogram for finite samples. The second approach aims to build consistent spectral density estimators, by smoothing the sample covariances. A review on these two approaches, in time series context, can be found in Robinson (1983).

2.2.1 Tapered periodogram.

The periodogram has an asymptotically negligible bias, but for a fixed grid of size N, there are biases due to leakage. In order to reduce this bias, data tapers (or faders) are introduced (see Brillinger (1970)). The tapered spatial periodogram, for discrete processes, is defined by:

$$I^{tap}(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2 H} \left| \sum_{\mathbf{s}\in D} h(\mathbf{s}) Z(\mathbf{s}) e^{-i\mathbf{s}^T\boldsymbol{\lambda}} \right|^2, \tag{2.18}$$

and for continuous processes:

$$I^{tap}_{\Delta}(\boldsymbol{\lambda}) = \frac{\Delta_1\Delta_2}{(2\pi)^2 H} \left| \sum_{\mathbf{s}\in D} h(\mathbf{s}) Z(\Delta\mathbf{s}) e^{-i\Delta\mathbf{s}^T\boldsymbol{\lambda}} \right|^2, \tag{2.19}$$

where h denotes the bidimensional taper and $H = \sum_{\mathbf{s}} h^2(\mathbf{s})$. Asymptotic distributional aspects of $I^{tap}(\boldsymbol{\lambda})$ have been studied in Brillinger (1970).

We will focus our attention on the discrete context. Discrete Fourier Transform of the tapered observations, in a similar way to (2.7) as:

$$J^{tap}(\boldsymbol{\lambda}) = \sum_{\mathbf{s}\in D} h(\mathbf{s})Z(\mathbf{s})e^{-i\mathbf{s}^T\boldsymbol{\lambda}}, \tag{2.20}$$

and the tapered spatial periodogram (2.18) can be written as:

$$I^{tap}(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2 H}|J^{tap}(\boldsymbol{\lambda})|^2. \tag{2.21}$$

Once again, we can also obtain a similar representation of I^{tap}_{Δ}, just modifying (2.20) in a suitable way.

The bidimensional taper function h is usually obtained as the tensor product of one-dimensional tapers. Data tapers are usually required to be measurable functions, bounded with bounded support, L^2-integrable and Lipschitz continuous. Examples of data tapers, for dimension one, can be seen in Priestley (1981), pp. 561-562.

As a general example, a one-dimensional taper $h(u)$, for $u \in [0,1]$, with smoothness parameter ρ, could be given by (see Dahlhaus and Künsch (1987)):

$$h(u) = \begin{cases} w(2u/\rho) & 0 \le u < \rho/2, \\ 1 & \rho/2 \le u \le 1/2, \\ h(1-u) & 1/2 < u \le 1. \end{cases} \tag{2.22}$$

The function w is chosen to be differentiable on $[0,1]$, with a Lipchitz-continuous derivative. For instance, for $w(\mathbf{u}) = 1/2(1-\cos(u\pi))$, the Tukey-Hanning taper is obtained (see Priestley (1981), pp. 442). In an unpublished Technical Report of North Carolina State University, Fuentes introduces the so-called rounded-taper, designed for giving more tapering to the grid corner observations. The rounded taper presents discontinuities when wheighting the borders, which may result unnatural. These taper functions can be considered in a similar way to those weight functions introduced in Section 2.2.2.

Data tapers help in removing the edge-effect in high dimensions. For that purpose, these estimators have been also used by Dahlhaus and Künsch (1987) in order to obtain $\sqrt{N}$ consistent parametric Whittle estimators, as we will see later.

Remark. In Fourier analysis, it is known that the partial sums of a Fourier series are not necesarily a good approximation of a function of interest. Data tapers appeared in Fourier analysis to improve the approximation of the partial sums of a Fourier series to a continuous function. These tapers (convergence factors) usually involve a maximum of 1 at $\mathbf{s} = \mathbf{0}$ and then they decrease to 0 as $\|\mathbf{s}\|$ increases, where $\|\mathbf{s}\| = \sqrt{s_1^2 + s_2^2}$.

2.2.2 Smoothed-covariances periodogram

One way of reducing the variance in the spatial-periodogram is simply to omit some of the terms which cause the tail-effect in (2.5) and (2.6). But this procedure will affect the expected value of the new expression. However, if the process has continuous spectrum, the covariances tend to zero as $\mathbf{s}$ increases and hence, if we omit only those terms which correspond to the tail of the sample covariance function, then the bias will not be so seriously affected.

Consider the following estimator for the spatial spectral density, obtained by truncating the covariances involved in the spatial periodogram estimation:

$$\tilde{f}(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2} \sum_{u_1=-m_1}^{m_1} \sum_{u_2=-m_2}^{m_2} \hat{C}(\mathbf{u}) \exp(-i\mathbf{u}^T\boldsymbol{\lambda}) \tag{2.23}$$

where m_1 and m_2 (truncation points) are such that $m_1 < n_1 - 1$ and $m_2 < n_2 - 1$. In this case, the variance of the estimate is $\mathcal{O}(M/N)$ ($M = m_1 m_2$, $N = n_1 n_2$). Thus, choosing $N \to \infty$, $M \to \infty$ and $M/N \to 0$, the bias and the variance tend to zero.

In order to solve the lack of consistency problem in high dimensions, Robinson (2006) studied a class of nonparametric spectral density estimates, for lattice processes, based on weighted sample covariances:

$$\tilde{f}(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{u}\in\mathcal{U}} k_n(\mathbf{u}) \hat{C}(\mathbf{u}) e^{-i\mathbf{u}^T\boldsymbol{\lambda}}, \tag{2.24}$$

which can be extended to the geostatistical data case as:

$$\tilde{f}_\Delta(\boldsymbol{\lambda}) = \frac{\Delta_1\Delta_2}{(2\pi)^2} \sum_{\mathbf{u}\in\mathcal{U}} k_n(\mathbf{u}) \hat{C}_\Delta(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}, \tag{2.25}$$

where $k_n(\mathbf{u})$ are weight functions (data windows or lag windows) obtained as the tensor product of one-dimensional windows:

$$k_n(\mathbf{u}) = k_{n_1}\left(\frac{u_1}{m_1}\right) k_{n_2}\left(\frac{u_2}{m_2}\right). \tag{2.26}$$

The truncated periodogram estimator given by (2.23) can be regarded as a particular case of (2.24). If $k_n(\mathbf{u})$ are chosen such that they decrease gradually, the contribution of the tail of the

sample covariances would be reduced, rather than eliminated. Provided that $k_n(\mathbf{u})$ decreases at a suitable rate, we may expect $\tilde{f}(\boldsymbol{\lambda})$ still being consistent estimates.

We consider bidimensional windows, obtained as the tensor product of one-dimensional windows. Some common lag windows used in the literature can be found in Hannan (1970) and Priestley (1981).

Chosing m_j, $j = 1, 2$ small relative to n_j, $j = 1, 2$, the variability of the estimated is controlled, and letting m_j for $j = 1, 2$ grow, the estimate becomes unbiased. For the geostatistical data case, define also, for $j = 1, 2$, the corresponding spectral window:

$$W_{n,\Delta}(\boldsymbol{\lambda}) = K_{m_1,\Delta_1}(\lambda_1) K_{m_2,\Delta_2}(\lambda_2), \tag{2.27}$$

where

$$K_{m_j,\Delta}(\lambda_j) = \frac{\Delta_j}{2\pi} \sum_{u_j} k_{n_j}\left(\frac{u_j}{m_j}\right) e^{-i\Delta u_j \lambda_j}, \quad j = 1, 2. \tag{2.28}$$

This kind of spectral windows are called *scale parameter* forms. The data windows in (2.26) are even, continuous, $k_{n_j}(0) = 1$ and L^2 integrable and by Parseval's Theorem:

$$(2\pi)^2 \int_{\Pi^2_\Delta} W^2(\boldsymbol{\theta}) d\boldsymbol{\theta} = \sum_{\mathbf{u} \in \mathcal{U}} k_n^2(\mathbf{u}), \tag{2.29}$$

where W denotes W_n (corresponding to $\Delta = 1$) or $W_{n,\Delta}$, by abuse of notation. These kind of estimators can be seen as weighted integrals of the periodogram, as we will see later.

$$\tilde{f}_\Delta(\boldsymbol{\lambda}) = \int_{\Pi^2_\Delta} W_{n,\Delta}(\boldsymbol{\lambda} - \boldsymbol{\omega}) I_\Delta(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{2.30}$$

We will give some examples of these lag-window functions, following (2.26). We will also give the corresponding Fourier Transform in (2.29). For simplicity, we will denote by k the weight functions k_{n_j}, $j = 1, 2$.

- Truncated periodogram window.

$$k(u_j) = \begin{cases} 1 & |u_j| \leq m_j \\ 0 & |u_j| > m_j, \end{cases} \tag{2.31}$$

with $m_j \leq n_j - 1$ (for $j = 1, 2$). Then, its Fourier Transform is given by the product of two Dirichlet kernels:

$$W(\boldsymbol{\theta}) = D_{m_1}(\theta_1) D_{m_2}(\theta_2), \quad D_{m_j}(\theta_j) = \frac{1}{2\pi} \frac{\sin\left(\left(m_j + \frac{1}{2}\right)\theta_j\right)}{\sin\left(\frac{\theta_j}{2}\right)}. \tag{2.32}$$

- Bartlett Window.

$$k(u_j) = \begin{cases} 1 - \frac{|u_j|}{m_j} & |u_j| \leq m_j \\ 0 & \text{otherwise.} \end{cases} \tag{2.33}$$

Its Fourier Transform is the product of two Fejer kernels:

$$W(\boldsymbol{\theta}) = F_{m_1}(\theta_1)F_{m_2}(\theta_2), \quad F_{m_j}(\theta_j) = \frac{1}{2\pi}\frac{\sin^2\left(\frac{m_j\theta_j}{2}\right)}{m\sin\left(\frac{\theta_j}{2}\right)}, \quad j = 1, 2. \tag{2.34}$$

- Tukey window.

$$k(u_j) = \begin{cases} 1 - 2a + 2a\cos\left(\frac{\pi u_j}{m_j}\right) & |u_j| \leq m_j \\ 0 & \text{otherwise.} \end{cases} \tag{2.35}$$

The Fourier Transform:

$$W(\boldsymbol{\theta}) = F_{m_1}(\theta_1)F_{m_2}(\theta_2), \tag{2.36}$$

$$F_{m_j}(\theta_j) = (1-2a)D_{m_j}(\theta_j) + aD_{m_j}\left(\theta_j - \frac{\pi}{m_j}\right) + aD_{m_j}\left(\theta_j + \frac{\pi}{m_j}\right), \tag{2.37}$$

where $D_{m_j}(\theta_j)$, $j = 1, 2$, is the Dirichlet kernel given by (2.32).

- Tukey-Hanning window (it is a particular case of Tukey window, taking $a = 0.25$).

$$k(u_j) = \begin{cases} \frac{1}{2}\left(1 + \cos\left(\frac{\pi u_j}{m_j}\right)\right) & |u_j| \leq m_j \\ 0 & \text{otherwise.} \end{cases} \tag{2.38}$$

$$W(\boldsymbol{\theta}) = F_{m_1}(\theta_1)F_{m_2}(\theta_2), \tag{2.39}$$

$$F_{m_j}(\theta_j) = \frac{1}{4}D_{m_j}\left(\theta_j - \frac{\pi}{m_j}\right) + \frac{1}{2}D_{m_j}(\theta_j) + \frac{1}{4}D_{m_j}\left(\theta_j + \frac{\pi}{m_j}\right). \tag{2.40}$$

- Parzen window.

$$k(u_j) = \begin{cases} 1 - 6\left(\frac{u_j}{m_j}\right)^2 + 6\left(\frac{|u_j|}{m_j}\right)^3 & |u_j| \leq \frac{m_j}{2}, \\ 2\left(1 - \frac{|u_j|}{m_j}\right)^3 & \frac{m_j}{2} \leq |u_j| \leq m_j, \\ 0 & \text{otherwise.} \end{cases} \tag{2.41}$$

An approximation of its Fourier Transform is given by:

$$W(\boldsymbol{\theta}) \approx F_{m_1}(\theta_1)F_{m_2}(\theta_2), \quad j = 1, 2, \tag{2.42}$$

$$F_{m_j}(\theta_j) = \frac{3}{8\pi m_j^3}\left(\frac{\sin\left(\frac{m_j\theta_j}{4}\right)}{\frac{1}{2}\sin\left(\frac{\theta_j}{2}\right)}\right)^4. \tag{2.43}$$

- Priestley-Bartlett window.

$$k(u_j) = \frac{3m_j^2}{(\pi u_j)^2}\left(\frac{m_j}{\pi u_j}\sin\left(\frac{\pi u_j}{m_j}\right) - \cos\left(\frac{\pi u_j}{m_j}\right)\right). \tag{2.44}$$

The Fourier Transform is given by:

$$W(\boldsymbol{\theta}) = F_{m_1}(\theta_1)F_{m_2}(\theta_2), \tag{2.45}$$

$$F_{m_j}(\theta_j) = \begin{cases} \frac{3m_j}{4\pi}\left(1 - \left(\frac{m_j\theta_j}{\pi}\right)^2\right) & |\theta_j| \leq \frac{\pi}{m_j}, \\ 0 & \text{otherwise}, \quad j = 1, 2. \end{cases} \tag{2.46}$$

Other lag windows are Daniell window, where $k(u) = sin(\pi u)/(\pi u)$ and the Cosine window.

Lemma 1. *Consider Z a zero-mean weakly stationary geostatistical process, observed on a regular lattice D, with $N = n_1 n_2$ points. Assume k_n is a bidimensional lag-window and W_n and K_{m_j} are defined as in (2.27) and (2.28) with unit spacement, respectively. Then, the smooth kernel estimator of the spectral density on the lattice, $\tilde{f}$ given by (2.24) can be written as:*

$$\tilde{f}(\boldsymbol{\lambda}) = \int_{\Pi^2} W_n(\boldsymbol{\lambda} - \boldsymbol{\omega}) I_n(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{2.47}$$

Lemma 2. *Consider Z a zero-mean weakly stationary geostatistical process, observed on a regular lattice D, with $N = n_1 n_2$ points and spacing $\Delta_1 = \Delta_2 = \Delta$. Assume k_n is a bidimensional lag-window and $W_{n,\Delta}$ and $K_{m_j,\Delta}$ are defined as in (2.27) and (2.28), respectively. Then, the smooth kernel estimator of the spectral density on the lattice, $\tilde{f}_\Delta$ given by (2.25) can be written as:*

$$\tilde{f}_\Delta(\boldsymbol{\lambda}) = \int_{\Pi^2_\Delta} W_{n,\Delta}(\boldsymbol{\lambda} - \boldsymbol{\omega}) I_\Delta(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{2.48}$$

Therefore, the device of weighting the sample covariance function so as to reduce the contribution from the tail has exactly the same effect as smoothing the periodogram by a weighted integral. Extension of this lemma for different spacements Δ_1 and Δ_2 is straightforward.

For most of the sequences $k_n(\mathbf{u})$ which we shall use in practice, the function W_n is typically concentrated around $\boldsymbol{\lambda} = 0$; the more slowly $k_n(\mathbf{u})$ decays, the more concentrated is W_n. The same holds for $W_{n,\Delta}$. The introduction of these weights may produce negative estimates of the spatial spectral density (see Section 2.5).

2.3 Expectation and covariance on finite grids.

In practice, the periodogram is computed just over the Fourier frequencies and thus, the characteristics of unbiasedness and independence do not hold, unless N is large enough. This fact implies that the error involved in considering the periodogram as an estimator of the spectral density can not be treated as a random noise component. These characteristics of the periodogram are useful whenever working on a dense grid, but in practice we usually work on finite grids with sparse data. Thus, the uncertainty of how much information is lost under the assumption of independence arises. The problem is how to quantify this loss of information. A first step would be to evaluate the expected value and the covariance of the periodogram with and without tapering, for a finite grid D.

Assumption 1. *Assume that for each* $\mathbf{s} \in \mathbb{R}^2$*, the taper* $h(\mathbf{s})$ *is measurable in* $\mathbf{s}$*, bounded, with compact support,* $\int h(\mathbf{s})^2 d\mathbf{s} \neq 0$*, and there exists a finite constant* $L \in \mathbb{R}$ *such that*

$$\int |h(\mathbf{s}+\mathbf{t}) - h(\mathbf{t})| d\mathbf{t} < L|\mathbf{s}|. \tag{2.49}$$

This condition is a form of integrated Lipschitz condition. As it is pointed out in Brillinger (1981) and in Brockwell and Davis (1991), this condition is satisfied by functions with uniformly bounded first derivatives and by functions of bounded variation.

The results shown below can be found in Porcu *et al.* (2005).

2.3.1 A brief note on cumulants.

In the same way that the generating function of a random variable generates its moments, the logarithm of the generating function, provides the cumulants. Cumulants are symmetric and multilinear in their arguments (Brillinger (1981)) and if any subset of $\{X_1, \ldots, X_r\}$ is statistically independent of the remaining set, then $cum(X_1, \ldots, X_r) = 0$. From this property, cumulants may be used to measure the statistical dependence of variables. In Brillinger (1970) cumulants theory is the fundamental tool to develop the frequency analysis of spatial series.

Cumulants have also been called semi-invariants and they admit a representation in terms of a Toeplitz form. They have been used in the context of long-memory processes. Recent applications of cumulants theory may be also found in applied physics (thermostatistics, signal processes...) and in probability theory. A complete description of the cumulants for spatial series properties can be found in Brillinger (1970). In this section, we just introduce the definition and remark some properties.

Definition 1. *Let $(X_1, \ldots, X_r)$ be a r-variate random variable. The joint cumulant of $X_1, \ldots, X_r$ is the coefficient of $t_1, \ldots, t_r$ in the expansion of the cumulant generating function*

$$\log E\left(e^{\sum_j X_j t_j}\right) \tag{2.50}$$

An alternative definition is given by

$$cum(X_1, \ldots, X_r) = \sum (-1)^{p-1}(p-1)! E(\prod_{j \in \mu_1} X_j) \ldots E(\prod_{j \in \mu_p} X_j) \tag{2.51}$$

where the sum and products extend over all partitions $(\mu_1, \ldots, \mu_p), \quad p = 1, \ldots, r$ of $(1, \ldots, r)$. From this definition, an inverse relation is obtained

$$E(X_1 \ldots X_r) = \sum (cum\{X_j, j \in \mu_1\}) \ldots (cum\{X_j, j \in \mu_p\}) \tag{2.52}$$

Definition 2. *Suppose that moments of all orders exist for $X_j, j = 1, \ldots, r$. The joint cumulant function of order k is defined as*

$$c_{1,\ldots,k}(\mathbf{s}_1, \ldots, \mathbf{s}_{(k-1)}) = cum\{X_1(\mathbf{s}_1 + \mathbf{s}), \ldots, X_{k-1}(\mathbf{s}_{k-1} + \mathbf{s}), X_k(\mathbf{s})\} \tag{2.53}$$

Note that

$$c_{1,\ldots,k}(\mathbf{s}_1, \ldots, \mathbf{s}_{k-1}) = c_{1,\ldots,k}(\mathbf{s}_1, \ldots, \mathbf{s}_{k-1}, \mathbf{0}) \tag{2.54}$$

Then, the cumulant of a single variable is its expectation and the covariance between X and Y is the joint cumulant of X and $\overline{Y}$. Therefore, the spectral density can be recovered from a certain cumulant.

If X is stationary and its moments exist and satisfy

$$\int \int \{|\mathbf{s}_1| + \ldots + |\mathbf{s}_{k-1}|\} \cdot |c_{1,\ldots,k}|(\mathbf{s}_1, \ldots, \mathbf{s}_{k-1}) d\mathbf{s}_1 \ldots d\mathbf{s}_{k-1} < \infty \tag{2.55}$$

then, we can define the cumulant spectra of order k

$$f_{1,\ldots,k}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{k-1}) = (2\pi)^{-2(k-1)} \int \int c_{1,\ldots,k}(\mathbf{s}_1, \ldots, \mathbf{s}_{k-1}) e^{-i \sum_{j=1}^{k-1} \boldsymbol{\lambda}_j^T \mathbf{s}_j} d\mathbf{s}_1 \ldots d\mathbf{s}_{k-1}$$

For the particular case of $k = 2$, the spectral density is obtained. Cumulants theory will be used in order to prove the results shown in this section.

2.3.2 Expected value of the periodogram and the tapered periodogram.

The following results are concerned with the expected value of the periodogram and the tapered periodogram. We will see that, in both cases, the expected value of the periodogram depends on the grid dimensions.

The expectation of the (non-tapered) periodogram depends directly on the grid dimensions, but the corresponding expected value of the tapered periodogram depends on the grid dimensions through the associated taper. This property is proved by the next results.

Define

$$H_r(\boldsymbol{\lambda}) = \sum_{\mathbf{s}\in D} h(\mathbf{s})^r e^{-i\mathbf{s}^T\boldsymbol{\lambda}}, \quad r \in \mathbb{N}.$$

Proposition 1. *Let Z be a second-order stationary random field, with mean μ and covariance function C. Assume Z is observed on a regular grid of size $N = n_1 n_2$. Let I be the periodogram with no taper as in equation (2.1). Suppose $\sum_{\mathbf{u}\in\mathcal{U}} |C(\mathbf{u})| < \infty$. Then, for $\boldsymbol{\lambda} \in \Pi^2$*

$$E(I(\boldsymbol{\lambda})) = \tag{2.56}$$

$$\frac{1}{(2\pi)^2 n_1 n_2}\int_{\Pi^2}\left[\frac{\sin n_1(\lambda_1-\omega_1)/2 \sin n_2(\lambda_2-\omega_2)/2}{\sin(\lambda_1-\omega_1)/2 \quad \sin(\lambda_2-\omega_2)/2}\right]^2 f(\boldsymbol{\omega})d\boldsymbol{\omega} \tag{2.57}$$

$$+\frac{1}{(2\pi)^2 n_1 n_2}\left[\frac{\sin n_1\lambda_1/2}{\sin\lambda_1/2}\right]^2\left[\frac{\sin n_2\lambda_2/2}{\sin\lambda_2/2}\right]^2 \mu^2. \tag{2.58}$$

Proposition 2. *Let Z be a second-order stationary random field, with mean μ and covariance function C. Assume Z is observed on a regular grid of size $N = n_1 n_2$. Let I^{tap} be the tapered periodogram as in equation (2.18). Suppose $\sum_{\mathbf{u}\in\mathcal{U}} |C(\mathbf{u})| < \infty$. Then, for $\boldsymbol{\lambda} \in \Pi^2$*

$$E(I^{tap}(\boldsymbol{\lambda})) = \tag{2.59}$$

$$\left((2\pi)^2\int_{\Pi^2}|H(\boldsymbol{\omega})|^2 d\boldsymbol{\omega}\right)^{-1}\int_{\Pi^2}|H(\boldsymbol{\lambda}-\boldsymbol{\omega})|^2 f(\boldsymbol{\omega})d\boldsymbol{\omega} \tag{2.60}$$

$$+\left((2\pi)^2\int_{\Pi^2}|H(\boldsymbol{\omega})|^2 d\boldsymbol{\omega}\right)^{-1}|H(\boldsymbol{\lambda})|^2\mu^2. \tag{2.61}$$

2.3.3 Covariance of the periodogram and the tapered periodogram.

The next results are concerned with the covariance structure associated with the periodogram and the tapered periodogram. These results show that the covariance between periodogram values is non-stationary.

Proposition 3. *Let Z be a second-order stationary random field, with mean μ and covariance function C. Assume Z is observed on a regular grid of size $N = n_1 n_2$. Let I be the periodogram with no taper as in equation (2.1). Then, the covariance structure associated with the periodogram I is given by*

$$\begin{aligned} Cov(I(\boldsymbol{\omega}), I(\boldsymbol{\lambda})) &= \left\{ \left(\frac{\sin n_1(\omega_1+\lambda_1)/2}{n_1 \sin(\omega_1+\lambda_1)/2} \cdot \frac{\sin n_2(\omega_2+\lambda_2)/2}{n_2 \sin(\omega_2+\lambda_2)/2} \right)^2 \right. \\ &+ \left. \left(\frac{\sin n_1(\omega_1-\lambda_1)/2}{n_1 \sin(\omega_1-\lambda_1)/2} \cdot \frac{\sin n_2(\omega_2-\lambda_2)/2}{n_2 \sin(\omega_2-\lambda_2)/2} \right)^2 \right\} f(\boldsymbol{\omega})^2 \\ &+ \mathcal{O}(N^{-1}). \end{aligned} \tag{2.62}$$

Proposition 4. *Let Z be a second-order stationary random field, with mean μ and covariance function C. Assume Z is observed on a regular grid of size $N = n_1 n_2$. Let I^{tap} be the tapered periodogram as in equation (2.18). Then, the covariance structure associated to the tapered periodogram I^{tap} is given by*

$$\begin{aligned} Cov(I^{tap}(\boldsymbol{\omega}), I^{tap}(\boldsymbol{\lambda})) &= |H_2(\mathbf{0})|^{-2} \left\{ |H_2(\boldsymbol{\omega}-\boldsymbol{\lambda})|^2 + |H_2(\boldsymbol{\omega}+\boldsymbol{\lambda})|^2 \right\} f(\boldsymbol{\omega})^2 \\ &+ \mathcal{O}(N^{-1}). \end{aligned} \tag{2.63}$$

2.3.4 The spatio-temporal case.

Consider now a stationary spatio-temporal process $Z(\mathbf{s}, t)$, $\mathbf{s} \in D \subset \mathbb{R}^2, t \in T \subset \mathbb{R}$. The process is observed on a spatial regular grid $n_1 \times n_2$ and at n time observations. Define the spatio-temporal periodogram, $I : \mathbb{R}^3 \to \mathbb{R}$, in a spatio-temporal frequency $(\boldsymbol{\omega}, \tau) \in \mathbb{R}^3$ as:

$$I(\boldsymbol{\lambda}, \tau) = \frac{1}{(2\pi)^3 n_1 n_2 n} \left| \sum_{(\mathbf{s},t) \in D \times T} Z(\mathbf{s}, t) \exp(-i\mathbf{s}^T \boldsymbol{\lambda}) \exp(-it\tau) \right|^2. \tag{2.64}$$

By abuse of notation, we will denote by I the spatio-temporal periodogram and J for the discrete Fourier Transform of the spatio-temporal data. Define then $J : \mathbb{R}^3 \to \mathbb{R}$

$$J(\boldsymbol{\lambda}, \tau) = \frac{1}{(2\pi)^{2/2} \sqrt{n_1 n_2 n}} \sum_{(\mathbf{s},t) \in D \times T} Z(\mathbf{s}, t) \exp(-i\mathbf{s}^T \boldsymbol{\lambda}) \exp(-it\tau). \tag{2.65}$$

Then, the spatio-temporal periodogram can be obtained as:

$$I(\boldsymbol{\lambda}, \boldsymbol{\tau}) = |J(\boldsymbol{\lambda}, \boldsymbol{\tau})|^2. \tag{2.66}$$

Similarly to the spatial case, we may consider the tapered spatio-temporal periodogram. The data taper, in this setting, is obtained as the product of two data tapers: h, the spatial component

taper and g the temporal taper. Thus, we define the tapered spatio-temporal peridogram as

$$I^{tap}(\boldsymbol{\lambda},\tau) = \frac{1}{(2\pi)^3 HG}\left|\sum_{(\mathbf{s},t)\in D\times T} h(\mathbf{s})g(t)e^{-i\mathbf{s}^T\boldsymbol{\lambda}}e^{-it\tau}\right|^2 \tag{2.67}$$

where

$$H=\sum_{\mathbf{s}\in D} h^2(\mathbf{s}) \quad \text{and} \quad G=\sum_{t\in T} g^2(t). \tag{2.68}$$

If we define

$$J^{tap}(\boldsymbol{\lambda},\tau) = \sum_{(\mathbf{s},t)\in D\times T} h(\mathbf{s})g(t)e^{-i\mathbf{s}^T\boldsymbol{\lambda}}e^{-it\tau}Z(\mathbf{s},t), \tag{2.69}$$

then, the spatio-temporal tapered periodogram periodogram can be written as the square modulus of J^{tap}, similarly to (2.66). The same considerations about the spectral density apply for the spatio-temporal setting. Thus, the spatio-temporal periodogram will be an asymptotically unbiased estimator of

$$f(\boldsymbol{\lambda},\tau) = \frac{1}{(2\pi)^3}\sum_{\mathbf{u}\in\mathcal{U}}\sum_{t\in\mathcal{T}} C(\mathbf{u},t)\exp(-i\mathbf{u}^T\boldsymbol{\lambda})\exp(-it\tau)$$

with $\mathbf{u}\in\mathcal{U}$ and $t\in\mathcal{T}=\{1-n,\ldots,n-1\}$.

Proposition 5. *Let Z be a second-order stationary space-time random field, with mean μ and covariance function $C(\mathbf{u},t)$. Assume Z is observed on a regular grid $n_1\times n_2\times n$ and let $N=n_1n_2n$. Let I be the periodogram with no taper as in equation (2.64). Suppose $\sum_{(\mathbf{u},t)\in\mathcal{U}\times\mathcal{T}}|C(\mathbf{u},t)|<\infty$. Then,*

$$E(I(\boldsymbol{\lambda},\tau)) = \frac{1}{(2\pi)^3 N}\int_{\Pi^3} B_{\boldsymbol{\lambda},\tau}(\boldsymbol{\omega},\nu)f(\boldsymbol{\omega},\nu)d\boldsymbol{\omega}d\nu \tag{2.70}$$

$$+ \frac{1}{(2\pi)^3 N}B_{\lambda,\tau}(\mathbf{0},0)\mu^2 \quad \textit{where } \Pi^3=[-\pi,\pi]^3 \tag{2.71}$$

and

$$B_{\boldsymbol{\lambda},\tau}(\boldsymbol{\omega},\nu) = \left(\frac{\sin n_1(\lambda_1-\omega_1)/2}{\sin(\lambda_1-\omega_1)/2}\right)^2\left(\frac{\sin n_2(\lambda_2-\omega_2)/2}{\sin(\lambda_2-\omega_2)/2}\right)^2\left(\frac{\sin n(\tau-\nu)/2}{\sin(\tau-\nu)/2}\right)^2. \tag{2.72}$$

Proposition 6. *Let Z be a second-order stationary space-time random field, with mean μ and covariance function $C(\mathbf{u},t)$. Assume Z is observed on a regular grid $n_1\times n_2\times n$ and let $N=n_1n_2n$. Let I^{tap} be the spatio-temporal tapered periodogram as in equation (2.67). Suppose $\sum_{(\mathbf{u},t)\in\mathcal{U}\times\mathcal{T}}|C(\mathbf{u},t)|<\infty$. Then,*

$$E(I^{tap}(\boldsymbol{\lambda},\tau)) = \tag{2.73}$$

$$\frac{1}{(2\pi)^3 HG}\int_{\Pi^3}|H(\boldsymbol{\lambda}-\boldsymbol{\omega})|^2|G(\tau-\nu)|^2 f(\boldsymbol{\omega},\nu)d\boldsymbol{\omega}d\nu \tag{2.74}$$

$$+\frac{1}{(2\pi)^3 HG}|H(\boldsymbol{\lambda})|^2|G(\tau)|^2\mu^2. \tag{2.75}$$

2.4 Bias of periodogram estimators.

The nonparametric estimators of the spectral density introduced in Section 2.2.2 (equations 2.24 and 2.25) may present edge-effect biases, due to the choice of kernel and bandwidth when smoothing the covariances. For geostatistical data, we will study the asymptotic properties of the smoothed-covariances estimators (2.25) for the spatial spectral density. The asymptotic framework considered is shrinking asymptotics, as in (Fuentes (2002)). These estimators have been called smooth non-parametric kernel estimators in Robinson (2006), for the lattice data case.

Expressions for the bias and the covariance structure are obtained and the implications on the edge-effect bias of the choice of the kernel, bandwidth and spacing parameter in the design are also discussed, both for tapered and untapered smoothed-covariances estimators.

2.4.1 Bias of smoothed-covariances periodogram estimators.

Consider Z a zero-mean weakly stationary spatial process, observed on a region $D \subset \mathbb{R}^2$, with covariance function C and spectral density f. Consider a shrinking-asymptotics framework, and for simplicity, take equal spacing parameters $\Delta_1 = \Delta_2 = \Delta$. Extensions for different spacings are straightforward. We restrict our attention to spatial processes on $\mathbb{R}^2$, but generalizations for $\mathbb{R}^d$ present no challenges.

Assumption 1: $k(v)$ is a real, even function such that $|k(v)| \leq 1$, for some $q > 0$ (the characteristic exponent, Priestley (1981), p.459), $0 < k_q < \infty$

$$\lim_{v \to 0} \frac{1 - k(v)}{|v|^q} = k_q \quad \text{and} \quad \int_{-\infty}^{\infty} |k(v)| dv < \infty.$$

Assumption 2: Let $n_j \to \infty$ and $j = 1, 2$. Then,

$$m_j \to \infty, \quad \Delta \to 0, \text{ and } \Delta n_j \to \infty.$$

Assumption 3: Z is a covariance stationary process and

$$\int_{\mathbb{R}^2} \left(\sum_{j=1}^{2} |u_j|^{\max(q,1)} \right) |C(\mathbf{u})| d\mathbf{u} < \infty.$$

Assumption 4: The rate of decay of the spectral density $f(\boldsymbol{\omega})$ at high frequencies is proportional to $|\boldsymbol{\omega}|^{-\tau}$, for $\tau > 2$.

Proposition 7. *Let Assumptions 1-4 above hold. Then, as* $n_j \to \infty, \quad j = 1, 2$, *the expectation of the smoothed kernel estimate (2.25) is given by:*

$$E(\tilde{f}_\Delta(\boldsymbol{\omega})) = f(\boldsymbol{\omega}) + \alpha_{1n,\Delta} + \alpha_{2n,\Delta} + o\left(\Delta^2 \sum_{j=1}^{2}(m_j^{-q} + n_j^{-1})\right) + \mathcal{O}(\Delta^\tau), \tag{2.76}$$

where

$$\alpha_{1n,\Delta} = \frac{\Delta^2}{(2\pi)^2} k_q \sum_{j=1}^{2} m_j^{-q} \sum_{\mathbf{u}\in\mathbb{Z}^2} |u_j|^q C_\Delta(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\omega}}, \tag{2.77}$$

$$\alpha_{2n,\Delta} = \frac{-\Delta^2}{(2\pi)^2} \sum_{j=1}^{2} n_j^{-1} \sum_{\mathbf{u}\in\mathbb{Z}^2} |u_j| C_\Delta(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\omega}}. \tag{2.78}$$

If we focus on the magnitude of $\alpha_{1n,\Delta}$ (the bias term), $\alpha_{2n,\Delta}$ (the edge-effect) and Δ^τ (from the approximation of the continuous spectral density through the spectral in the lattice) we will have to study the relations between n_j, m_j, Δ for different values of τ (decay of the spectral density) and q, we have that

$$\alpha_{1n,\Delta} = \mathcal{O}\left(\Delta^2 \sum_{j=1}^{2} m_j^{-q}\right), \quad \alpha_{2n,\Delta} = \mathcal{O}\left(\Delta^2 \sum_{j=1}^{2} n_j^{-1}\right) \quad \text{and} \quad \mathcal{O}(\Delta^\tau),$$

with $\tau > 2$. The last term corresponds to the aliasing part. In order to achieve consistency, we require

$$\frac{m_j}{n_j} \to 0, \quad j = 1, 2.$$

In order to guarantee that the bias is the leading term, we have to impose the following condition on the spacing parameter, Δ:

$$\Delta^\beta m_j^q \to 0, \quad \beta > 0, \quad j = 1, 2.$$

Therefore, when $q \leq 1$, the edge-effect $\alpha_{2n,\Delta}$ is dominated by the bias $\alpha_{1n,\Delta}$.

Remark. The characteristic exponent for a weight function k is defined as the largest exponent $q > 0$ such that:

$$k^{(q)} = \lim_{u\to 0}\left(\frac{1 - k(u)}{|u|^q}\right)$$

exists, is finite and non-zero. For instance, Bartlett window has characteristic exponent $q = 1$. Parzen and Daniell windows have characteristic exponent $q = 2$. In the case of the truncated periodogram, the characteristic exponent is not finite.

The behaviour of k near the origin is related to the behaviour of its Fourier Transform for large frequencies and the characteristic exponent is a measure of the width of the Fourier Transform of k. The larger q, the slower does the function k decay.

2.4.2 Bias of smoothed-covariances tapered-periodogram estimators.

We will introduce a spectral density estimator, obtained from the combination of tapering and smoothing-covariances approaches. Consider a taper function $h(v)$ satisfaying the following condition:

Assumption 5. $h(v)$ is Lipschitz-continuous on $[0,1]$ and satisfies

$$h(0) = 0, \quad h(1-v) = h(v), \quad 0 \leq v < \frac{1}{2},$$

and

$$\int_0^1 h^2(v)dv > 0.$$

For an integer t, define

$$h_{j,t} = h\left(\frac{t-1/2}{n_j}\right), \quad j = 1,2$$

and define the tapered sample covariances

$$\hat{C}^h_\Delta(\mathbf{u}) = \frac{1}{H_n} \sum_{\mathbf{s}\in D(\mathbf{u})} \left(\prod_{j=1}^{2} h_{j,s_j} h_{j,s_j+u_j}\right) Z(\Delta\mathbf{s}) Z(\Delta\mathbf{s} + \Delta\mathbf{u}) \tag{2.79}$$

where

$$H_n = \prod_{j=1}^{2} \sum_{s_j=1}^{n_j} h^2_{j,s_j}.$$

The smoothed-covariances tapered spectral density estimate is defined by:

$$\tilde{f}^h_\Delta(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{u}\in\mathcal{U}} k_n(\mathbf{u}) \hat{C}^h_\Delta(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}.$$

Consider also the following conditions on the spectral density and the spectral window.

Assumption 6. $f(\boldsymbol{\omega})$ is twice boundedly differentiable on $\mathbb{R}^2$.

Assumption 7. For n large enough (which implies, Δ small),

$$K_{m_j,\Delta}(\omega_j) \geq 0 \quad j = 1,2.$$

Proposition 8. *If Assumptions 1-7 hold, then as $n_j \to \infty$, for $j = 1,2$, the expectation of the smoothed kernel estimate, using the tapered covariances, is given by:*

$$E\left(\tilde{f}^h_\Delta(\boldsymbol{\lambda})\right) = f(\boldsymbol{\lambda}) + \alpha_{1n,\Delta}(1 + o(1)) + \mathcal{O}\left(\sum_{j=1}^{2} n_j^{-2}\right) + \mathcal{O}(\Delta^\tau) \tag{2.80}$$

where $\alpha_{1n,\Delta}$ is given by (2.77).

From this result, we can see the improvement obtained by using a taper. For $q \leq 2$, we have to require $\Delta^{\beta} m_j^q \to 0$ to guarantee that the bias term is the dominant one. By the consistency requirement, we ensure that it dominates the remainder term ($\mathcal{O}(\sum_{j=1}^{2} n_j^{-2})$).

2.4.3 Optimal bandwidth selection.

The bandwidth parameter for smoothed-covariances estimators (2.25) is given by the values m_1 and m_2. We will give an asymptotically optimal (in the sense of the Mean Square Error) bandwidth. For that purpose, we will obtain asymptotic expressions for the bias and variance of such estimators.

Lemma 3. *Consider estimator (2.25) and assume that conditions in Theorem 1 hold. Then, the asymptotic expectation of (2.25) as an estimator of $f(\boldsymbol{\lambda})$ can be approximated by:*

$$E(\tilde{f}_{\Delta}(\boldsymbol{\lambda})) - f(\boldsymbol{\lambda}) =_{as} \frac{-k_q}{2\Delta^2} \mathrm{tr}\left(\mathcal{H} f_{\Delta}(\boldsymbol{\lambda})\right) \sum_{j=1}^{2} m_j^{-q}, \tag{2.81}$$

where $\mathcal{H} f_{\Delta}$ denotes the Hessian matrix and tr *denotes the trace operator.*

From now, we will assume that the process we observe is a linear sequence as in (2.10).

Lemma 4. *Consider estimator (2.25) and assume that conditions in Theorem 1 hold. Assume also that Z can be represented as (2.10). Then, the asymptotic variance of (2.25) as an estimator of $f(\boldsymbol{\lambda})$ can be approximated by:*

$$Var(\tilde{f}_{\Delta}(\boldsymbol{\lambda})) =_{as} (1 + \delta_{0,\pi}) \frac{(2\pi)^2 \Delta^2 m_1 m_2}{n_1 n_2} \int_{\Pi_{\Delta}^2} f_{\Delta}^2(\boldsymbol{\omega}) W_{n,\Delta}^2(\boldsymbol{\lambda} - \boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{2.82}$$

The function $\delta_{0,\pi}$ vanishes unless the frequency $\boldsymbol{\lambda}$ belongs to $\{-\pi, 0, \pi\} \times \{-\pi, 0, \pi\}$, in which case, it takes value 1.

For scale-parameter windows (the ones we are considering), by Parseval's Theorem, we can approximate the variance by:

$$Var(\tilde{f}_{\Delta}(\boldsymbol{\lambda})) =_{as} (1 + \delta_{0,\pi}) \frac{(2\pi)^2 \Delta^2 m_1 m_2}{n_1 n_2} f_{\Delta}^2(\boldsymbol{\lambda}) \left(\int k^2(\mathbf{u}) d\mathbf{u} \right)^2. \tag{2.83}$$

In order to evaluate the goodness of these spectrum estimators, one can use several criteria. We will try to minimize the asymptotic Mean Square Error, where $Bias$ denotes the asymptotic bias of the estimator (2.25) and Var denotes the asymptotic variance:

$$AMSE = Bias^2(\tilde{f}_{\Delta}(\boldsymbol{\lambda})) + Var(\tilde{f}_{\Delta}(\boldsymbol{\lambda})).$$

Proposition 9. *Consider estimator (2.25) and assume that conditions in Theorem 1 hold. Assume also that Z can be represented as (2.10). Then, the AMSE is minimized by:*

$$m_1 = m_2 = m = c(4q\Delta n_1 n_2)^{\frac{1}{2q+2}},$$

where

$$c = \left(\frac{q^{\frac{q-1}{q+1}} k_q^2 tr(H f_\Delta(\boldsymbol{\lambda}))^2 (n_1 n_2)^{\frac{2q}{q+1}}}{2^{\frac{3q+7}{q+1}} \pi \Delta^5 f_\Delta(\boldsymbol{\lambda}) \left(\int k^2(u) du \right)^2} \right)^{\frac{1}{2q+2}}.$$

For $q = 1$, the optimal bandwidth is given by:

$$m = \left(\frac{k_1^2 \mathrm{tr}(H f_\Delta(\boldsymbol{\lambda}))^2 (n_1 n_2)^2}{8\pi f_\Delta(\boldsymbol{\lambda}) \left(\int k^2(u) du \right)^2 \Delta^4} \right)^{1/4}.$$

2.5 Parametric estimation of the spectral density

In the spectral parametric context, Whittle parameter estimation (introduced in Whittle (1954)) is the most popular method. This estimation procedure is based on an approximation to the Gaussian log-likelihood and it uses the periodogram as a pilot estimate (e.g. Guyon (1982)).

For a parametric model of the spatial spectral density f_θ with $\theta \in \Theta \subset \mathbb{R}^p$, the Whittle parameter estimator $\hat{\theta}$ is given by:

$$\hat{\theta} = \arg\min_\theta L(\theta, I), \tag{2.84}$$

where $L(\theta, I)$ denotes the Whittle log-likelihood

$$L(\theta, I) = \int_{\Pi^2} \left(\log f_\theta(\boldsymbol{\lambda}) + \frac{I(\boldsymbol{\lambda})}{f_\theta(\boldsymbol{\lambda})} \right) d\boldsymbol{\lambda}. \tag{2.85}$$

The log-likelihood function (2.85) can be interpreted as the Kullback-Leibler divergence between I and f_θ. This estimator shows good consistency properties in the one dimensional case.

Note that, in practice, this log-likelihood (2.85) is approximated by a discretized version:

$$\sum_{\mathbf{k}} \left(\log f_\theta(\boldsymbol{\lambda}_\mathbf{k}) + \frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_\theta(\boldsymbol{\lambda}_\mathbf{k})} \right), \tag{2.86}$$

where the sums extends over all Fourier frequencies.

Suppose that we have a random process on the Euclidean space $\mathbb{R}^d$. Assume that it is observed on a grid D with N points, the number of points tending to infinity at the same speed in all directions. In that case, we find an edge-effect bias of order $N^{-1/d}$ (Guyon (1982)). For dimension $d = 1$, this effect is negligible but it becames important for dimension $d \geq 2$.

A first attempt to correct this edge-effect bias in $d = 2$ is proposed by Guyon (1982). In order to obtain a $\sqrt{N}$-consistent estimator of θ, an unbiased version of the periodogram can be used in the Whittle log-likelihood expression. The unbiased periodogram is obtained from (2.5), replacing the sample covariances $\hat{C}(\mathbf{v})$ by the unbiased sample covariances, namely $\tilde{C}(\mathbf{v})$, with $\mathbf{v}^T = (v_1, v_2)$, given by

$$\tilde{C}(\mathbf{v}) = \sum_{\mathbf{s}\in D(\mathbf{v})} \frac{1}{(n_1 - s_1 + v_1)(n_1 - s_2 + v_2)} Z(\mathbf{s})Z(\mathbf{s}+\mathbf{v}). \tag{2.87}$$

Although the use of unbiased covariances in Whittle log-likelihood approximation provides consistent estimators, $\tilde{C}$ present some unpleasant features. The first drawback of this unbiased estimator for the covariance is that it may not be positive definite. This fact implies that spectral estimates obtained from $\tilde{C}$ may present negative values. Besides, the positive definite character allows to interpret the Whittle estimator as a minimun distance estimator. This intuitive idea is lost when considering $\tilde{C}$ Another problem is that $\tilde{C}$ has larger variance than $\hat{C}$, specially for large lags.

Dahlhaus and Künsch (1987) proved that the inconsistency problem of Whittle estimates for multidimensional settings can be solved by tapering the data. Introducing a data taper as (2.22) and considering a tapered-periodogram as in (2.18), the parameter estimators obtained by minimizing(2.85) are $\sqrt{N}$-consistent, assuming that the smoothness parameter in the taper (2.22) $\rho = \rho_N$ satisfies: $\rho_N \to \rho_0$ and $\rho_N = o\left(N^{2/d-1}\right)$, for a constant ρ_0.

A more recent study by Robinson and Vidal Sanz (2006) shows that the edge-effect problem in Whittle estimation for dimension $d \geq 2$ can be overcome by optimizing a Whittle function by a Newton method, using a smoothed-covariances periodogram. We will see (in Chapter 4, when treating the goodness-of-fit test) that this bias can be also corrected by Bootstrap procedures.

2.5.1 A nonparametric estimator based on Whittle's log-likelihood.

Based on the discrete approximation to Whittle's log-likelihood (2.86), it is possible to obtain a nonparametric estimator for the log-spectral density $m_\theta = \log f_\theta$. It is easy to see that, minimizing

(2.86) is equivalent to maximize in θ:

$$\sum_{\mathbf{k}} \left(Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} \right), \tag{2.88}$$

where $Y_{\mathbf{k}}$ denotes the log-periodogram value at the Fourier frequency $\boldsymbol{\lambda}_{\mathbf{k}}$. Besides, this equations corresponds to the log-likelihood associated to (2.15) when ignoring the residual term $r_{\mathbf{k}}$.

From a non parametric approach, we consider the estimator obtained for the log-spectral density function m_θ by a multidimensional local linear kernel estimator. For any $\mathbf{x} \in \mathbb{R}^2$, we approximate $m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})$ by the plane $a + \mathbf{b}^T(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x})$. Then, we construct the local loglikelihood function

$$\sum_{\mathbf{k}} \left[Y_{\mathbf{k}} - a - \mathbf{b}^T(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x}) - e^{Y_{\mathbf{k}} - a - \mathbf{b}^T(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x})} \right] K_H(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x}), \tag{2.89}$$

where the function K_H is a reescaled bidimensional kernel, H is a bidimensional bandwidth matrix and $K_H(\mathbf{x}) = |H|^{-1/2} K(H^{-1/2}\mathbf{x})$. The local maximum likelihood estimator $\hat{m}_{LK}(H, x) \equiv \hat{m}_{LK}(\mathbf{x})$ of $m(\mathbf{x})$ is $\hat{a}$ in the maximizer $(\hat{a}, \hat{b})$ of (2.89).

For Mercer and Hall wheat yield data, introduced in Section 1.4.1, in Figure (2.1) we show the spatial periodogram, a parametric estimation obtained Whittle's method and the nonparametric estimation obtained by maximun local log-likelihood (2.89). The parametric estimation corresponds to a first-order autoregressive model (see Whittle (1954)):

$$Z(\mathbf{s}) = \alpha_1(Z(s_1 + 1, s_2) + Z(s_1 - 1, s_2)) + \alpha_2(Z(s_1, s_2 + 1) + Z(s_1, s_2 - 1)) + \varepsilon(\mathbf{s}), \tag{2.90}$$

where $\varepsilon(\mathbf{s})$ are zero-mean independent Gaussian random variables, with variance σ_ε^2. The corresponding spectral density is given by

$$f(\boldsymbol{\lambda}) = \frac{\sigma^2}{(2\pi)^2} \left(1 - 2\alpha_1 \cos(\lambda_1) - 2\alpha_2 \cos(\lambda_2)\right)^{-2}, \quad \boldsymbol{\lambda} \in \Pi^2. \tag{2.91}$$

We will refer to model (2.90) as the spatial autoregressive model ($SAR(1)$ model). We obtain as estimated parameters $\widehat{\alpha_1} = 0.23217$, $\widehat{\alpha_2} = 0.09267$ and variance 0.12452. The estimates we got are quite similar to those obtained by Whittle (1954), which gives $\tilde{\alpha}_1 = 0.213$ and $\tilde{\alpha}_2 = 0.102$.

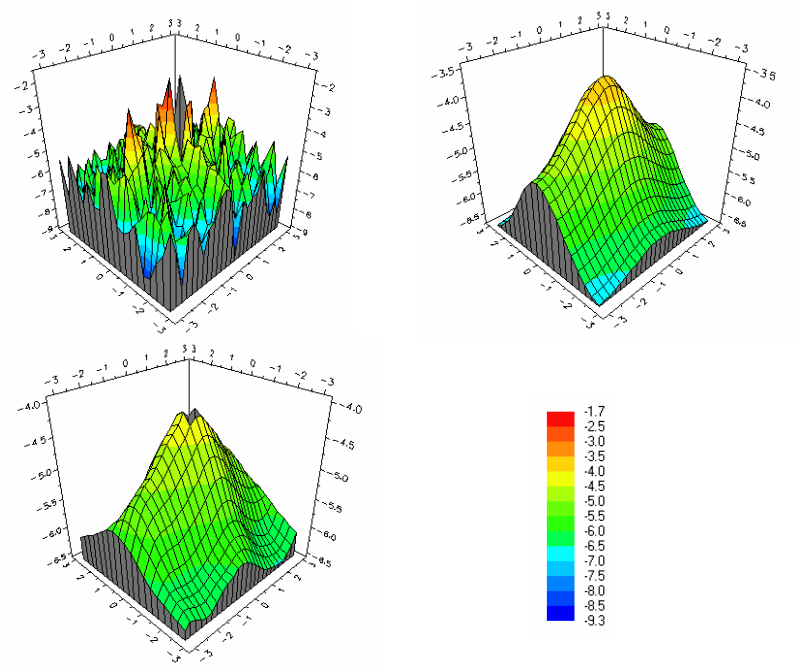

Figure 2.1: Spatial spectral density estimation for Mercer and Hall wheat yield data. From top-left to right-bottom: periodogram, parametric SAR model and local log-likelihood estimator from (2.89).

2.6 An illustrative simulation study.

We consider a particular case of the linear-by-linear processes: the doubly geometric process (Martin (1979)). This type of process will be also used in next chapters. The spectral density of such a process is given by:

$$f(\boldsymbol{\lambda}) = \frac{\sigma^2}{(2\pi)^2} \cdot \frac{1}{1+\beta_1^2 - 2\beta_1 \cos(\lambda_1)} \cdot \frac{1}{1+\beta_2^2 - 2\beta_2 \cos(\lambda_2)}, \tag{2.92}$$

with $\beta_1, \beta_2 \in [0,1)$. One thousand simulations have been carried out on a 20×20 regular grid, using the algorithm provided by Alonso *et al.* (1996). We compute different estimations for the spectral density: tapered and non-tapered periodogram, smoothed-covariance estimator and smoothed-covariance tapered-periodogram. We present the results from the estimation of the spectral density (2.92), with autoregression parameters $(0.5, 0.5)$.

In Figures 2.2 and 2.3, we show the Mean and Mean Square Errors in the estimation of the spectral density. Results in the logarithmic scale are shown in Figures 2.4 and 2.5, for Mean and Mean Square Errors, respectively.

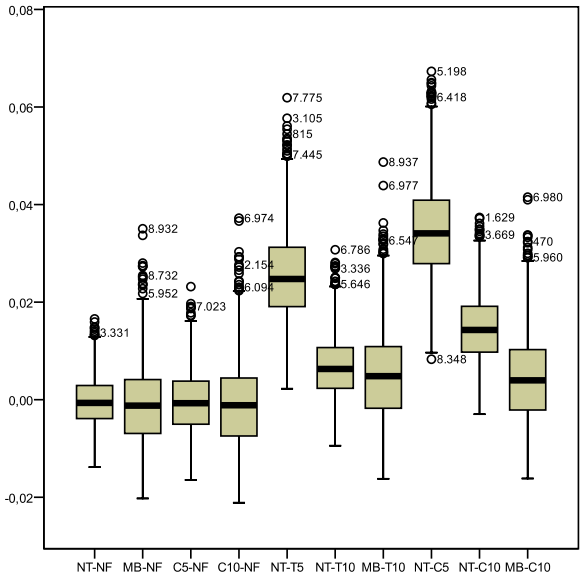

Figure 2.2: Mean Error. Legend: NT-NF=No tapering and no smoothing; MB-NF=Multiplicative Bartlett taper and no smoothing; C5(10)-NF= Cosine taper with $m_1 = m_2 = 5(10)$ and no smoothing; NT-T5(10)=No tapering and truncated covariances ($m_1 = m_2 = 5(10)$); MB-C10=Multiplicative Bartlett taper and cosine kernel for smoothing covariances; NT-C5(10)=No tapering and cosine smoothing in the covariances ($m_1 = m_2 = 10$); MB-C10=Multiplicative Bartlett taper and cosine smoothing in the covariances ($m_1 = m_2 = 10$).

In these figures, each box-plot corresponds to a different spectral density estimator. The first one is for the spatial periodogram (no tapering and no smoothing). The second box-plot is for a multiplicative Bartlett taper, with no smoothing on the covariances. The third and fourth ones use a cosine taper, with parameters 5 and 10, respectively. The fifth and sixth box-plots are for the truncated periodogram. The seventh is for multiplicative Bartlett taper and truncation in the covariances. The eighth and ninth box-plots are for no-tapering and cosine smoothing and the last one is for multiplicative Bartlett taper and cosine smoothing. Then, we could divide these box-plots in four cases: spatial periodogram (box-plot 1), tapered periodogram (box-plots 2, 3 and 4), smoothed covariances (box-plots 5, 6, 8 and 9) and tapered and smoothed-covariances (box-plots, 7 and 10).

From Figure 2.2, we can see that the bias of the smoothed-covariance estimates is higher than the bias of the tapered periodograms, although this bias can be corrected by tapering (as we can see in the box-plots corresponding to MB-T10 and MB-C10).

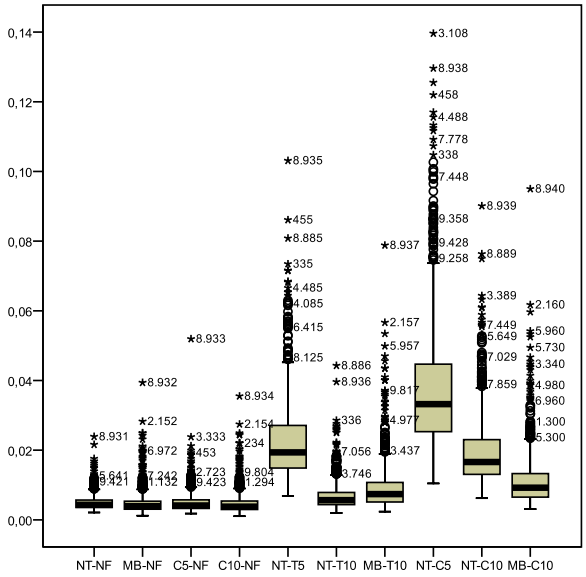

Figure 2.3: Mean Square Error. Legend: NT-NF=No tapering and no smoothing; MB-NF=Multiplicative Bartlett taper and no smoothing; C5(10)-NF= Cosine taper with $m_1 = m_2 = 5(10)$ and no smoothing; NT-T5(10)=No tapering and truncated covariances ($m_1 = m_2 = 5(10)$); MB-C10=Multiplicative Bartlett taper and cosine kernel for smoothing covariances; NT-C5(10)=No tapering and cosine smoothing in the covariances ($m_1 = m_2 = 10$); MB-C10=Multiplicative Bartlett taper and cosine smoothing in the covariances ($m_1 = m_2 = 10$).

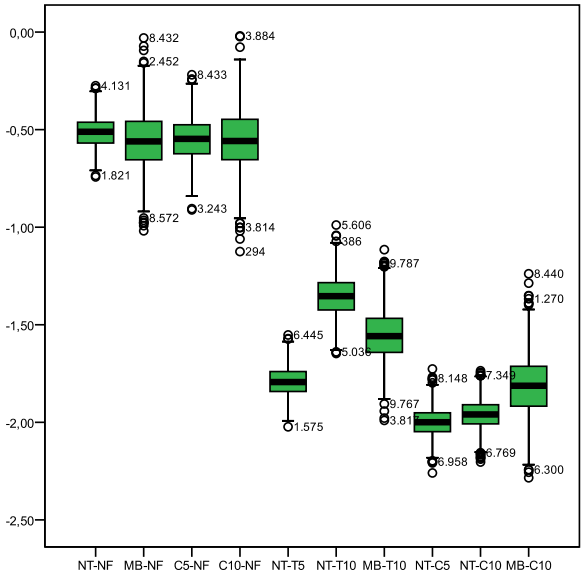

Figure 2.4: Mean Error. Logarithmic scale. Legend: NT-NF=No tapering and no smoothing; MB-NF=Multiplicative Bartlett taper and no smoothing; C5(10)-NF= Cosine taper with $m_1 = m_2 = 5(10)$ and no smoothing; NT-T5(10)=No tapering and truncated covariances ($m_1 = m_2 = 5(10)$); MB-C10=Multiplicative Bartlett taper and cosine kernel for smoothing covariances; NT-C5(10)=No tapering and cosine smoothing in the covariances ($m_1 = m_2 = 10$); MB-C10=Multiplicative Bartlett taper and cosine smoothing in the covariances ($m_1 = m_2 = 10$).

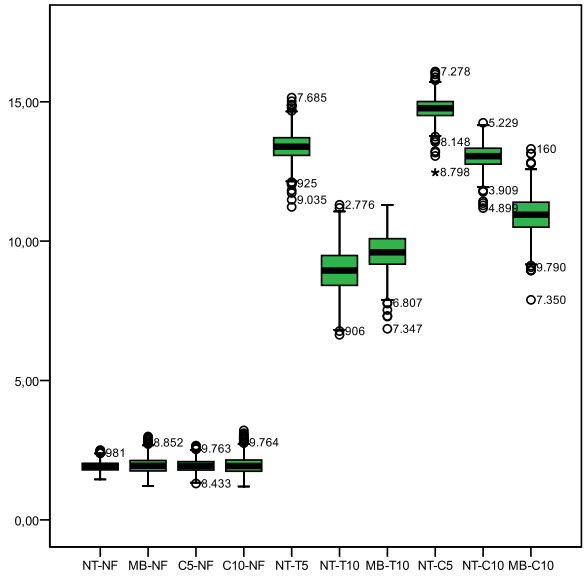

Figure 2.5: Mean Square Error. Logarithmic scale. Legend: NT-NF=No tapering and no smoothing; MB-NF=Multiplicative Bartlett taper and no smoothing; C5(10)-NF= Cosine taper with $m_1 = m_2 = 5(10)$ and no smoothing; NT-T5(10)=No tapering and truncated covariances ($m_1 = m_2 = 5(10)$); MB-C10=Multiplicative Bartlett taper and cosine kernel for smoothing covariances; NT-C5(10)=No tapering and cosine smoothing in the covariances ($m_1 = m_2 = 10$); MB-C10=Multiplicative Bartlett taper and cosine smoothing in the covariances ($m_1 = m_2 = 10$).

2.7 Appendix Chapter 2.

2.7.1 Proofs for Section 2.2

Proof of Lemma 1. The periodogram can be written as:

$$I(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2}\frac{1}{n_1 n_2}\sum_{x_1=1}^{n_1}\sum_{x_2=1}^{n_2}\sum_{y_1=1}^{n_1}\sum_{y_2=1}^{n_2} Z(\mathbf{x})Z(\mathbf{y})e^{-i(\mathbf{x}-\mathbf{y})^T\boldsymbol{\lambda}}$$

This last form will be the one that we use to prove the expression of the $\tilde{f}(\boldsymbol{\omega})$ estimate in terms of the periodogram:

$$\begin{aligned}
&\tilde{f}(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2}\sum_{\mathbf{u}\in\mathcal{U}} k_n(\mathbf{u})C(\mathbf{u})e^{-i\mathbf{u}^T\boldsymbol{\lambda}} = \\
&\frac{1}{(2\pi)^2}\sum_{\mathbf{u}\in\mathcal{U}} k\left(\frac{u_1}{m_1}\right)k\left(\frac{u_2}{m_2}\right)C(\mathbf{u})e^{-i\mathbf{u}^T\boldsymbol{\lambda}} = \\
&\frac{1}{(2\pi)^2}\sum_{\mathbf{u}\in\mathcal{U}} k\left(\frac{u_1}{m_1}\right)k\left(\frac{u_2}{m_2}\right)\frac{1}{n_1 n_2}\sum_{\mathbf{s}\in D(\mathbf{u})} Z(\mathbf{s})Z(\mathbf{s}+\mathbf{u})e^{-i\mathbf{u}^T\boldsymbol{\lambda}} = \\
&\frac{1}{2\pi}\sum_{u_1=1-n_1}^{n_1-1} k\left(\frac{u_1}{m_1}\right)\frac{1}{2\pi}\sum_{u_2=1-n_2}^{n_2-1} k\left(\frac{u_2}{m_2}\right)\frac{1}{n_1 n_2}\sum_{\mathbf{s}\in D(\mathbf{u})} Z(\mathbf{s})Z(\mathbf{s}+\mathbf{u})e^{-i\mathbf{u}^T\boldsymbol{\lambda}} = \\
&\int_{\Pi^2} K_{m_1}(\lambda_1-\omega_1)K_{m_2}(\lambda_2-\omega_2)\frac{1}{(2\pi)^2 n_1 n_2}\sum_{\mathbf{s}\in D(\mathbf{u})} Z(\mathbf{s})Z(\mathbf{s}+\mathbf{u})e^{-i\mathbf{u}^T\boldsymbol{\omega}}d\boldsymbol{\omega}
\end{aligned}$$

where

$$K_{m_j}(\lambda_j) = \frac{1}{2\pi}\sum_{u_j} k\left(\frac{u_j}{m_j}\right)e^{-iu_j\lambda_j}, \quad j=1,2.$$

Defining $W_n(\boldsymbol{\lambda}) = K_{m_1}(\lambda_1)K_{m_2}(\lambda_2)$ we can write:

$$\tilde{f}(\boldsymbol{\lambda}) = \int_{\Pi^2} W_n(\boldsymbol{\lambda}-\boldsymbol{\omega})I(\boldsymbol{\omega})d\boldsymbol{\omega}.$$

□

Proof of Lemma 2. The proof of this Lemma is inmediate, recalling that product and convolution operators for a Fourier pair. The smooth kernel estimator

$$\tilde{f}_\Delta(\boldsymbol{\lambda}) = \frac{\Delta^2}{(2\pi)^2}\sum_{\mathbf{u}\in\mathcal{U}} k_n(\mathbf{u})C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}, \tag{2.93}$$

is given by the Discrete Fourier Transform (DFT) of the product of k_n and C_Δ. Then $\tilde{f}_\Delta$ can be writen as the convolution of the Discrete Fourier Transforms of k_n and C_Δ. The DFT of k_n is given by $W_{n,\Delta}$ in (2.27), and the DFT of C_Δ is the periodogram in (2.2). □

2.7.2 Proofs for Section 2.3

Proof of Proposition 1. The proof of Proposition 1 is obtained just applying elementary properties of the cumulants. Recall representation (2.8) for the periodogram. Then,

$$E(I(\boldsymbol{\omega})) = E(J(\boldsymbol{\omega})\overline{J(\boldsymbol{\omega})}) = cum(J(\boldsymbol{\omega})\overline{J(\boldsymbol{\omega})}) + |E(J(\boldsymbol{\omega}))|^2,$$

where $J(\boldsymbol{\omega})$ is defined as in equation (2.7). Then, we obtain

$$\begin{aligned}
&E(I(\boldsymbol{\omega})) = \\
&cum(J(\boldsymbol{\omega})\overline{J(\boldsymbol{\omega})}) + |cum(J(\boldsymbol{\omega}))|^2 = \\
&\frac{1}{(2\pi)^2 n_1 n_2} \sum_{\mathbf{s}\in D} \sum_{\mathbf{x}\in D} \exp(-i\boldsymbol{\omega}^T(\mathbf{s}-\mathbf{x}))C(\mathbf{s}-\mathbf{x}) \\
&+\frac{1}{(2\pi)^2 n_1 n_2} \left| \sum_{\mathbf{s}\in D} \exp(-i\boldsymbol{\omega}^T\mathbf{s}) \right|^2 \mu^2.
\end{aligned}$$

Computing the differences $\mathbf{s}-\mathbf{x}=\mathbf{u}$, the difference vector belongs to the index set $\mathbf{u}\in\mathcal{U}$. Each vector $\mathbf{u}$ appears $\left(1-\frac{|u_1|}{n_1}\right)\left(1-\frac{|u_2|}{n_2}\right)$ times and plays the same role as a taper in a DFT. Note now that the Bartlett's kernel given by:

$$\frac{1}{(2\pi)^2 n_1 n_2} \left[\frac{\sin(n_1\omega_1)/2}{\sin(\omega_1)/2}\right]^2 \left[\frac{\sin(n_2\omega_2)/2}{\sin(\omega_2)/2}\right]^2 \tag{2.94}$$

can be obtained from

$$\sum_{\mathbf{u}\in\mathcal{U}} \left(1-\frac{|u_1|}{n_1}\right)\left(1-\frac{|u_2|}{n_2}\right) g(\mathbf{u}) \exp(-i\boldsymbol{\omega}^T\mathbf{u}) =$$

$$= \int_{\Pi^2} \frac{1}{(2\pi)^2 n_1 n_2} \left[\frac{\sin n_1\lambda_1/2}{\sin\lambda_1/2}\right]^2 \left[\frac{\sin n_2\lambda_2/2}{\sin\lambda_2/2}\right]^2 G(\boldsymbol{\omega}-\boldsymbol{\lambda})d\boldsymbol{\lambda}.$$

Thus, we finally complete the proof by having

$$
\begin{aligned}
&E(I(\boldsymbol{\omega})) \\
&= \frac{1}{(2\pi)^2 N} \sum_{\mathbf{u}\in\mathcal{U}} \left(1 - \frac{|u_1|}{n_1}\right)\left(1 - \frac{|u_2|}{n_2}\right) \exp(-i\mathbf{h}^T\boldsymbol{\omega}) C(\mathbf{h}) \\
&+ \frac{1}{(2\pi)^2 N} \left| \sum_{\mathbf{s}\in D} \exp(-i\mathbf{s}^T\boldsymbol{\omega}) \right|^2 \\
&= \frac{1}{(2\pi)^2 N} \int_{\Pi^2} \left(\frac{\sin n_1\lambda_1/2}{\sin \lambda_1/2}\right)^2 \left(\frac{\sin n_2\lambda_2/2}{\sin \lambda_2/2}\right)^2 f_\theta(\boldsymbol{\omega} - \boldsymbol{\lambda}) d\boldsymbol{\lambda} \\
&+ \frac{1}{(2\pi)^2 N} \left(\frac{\sin n_1\omega_1/2}{\sin \omega_1/2}\right)^2 \left(\frac{\sin n_2\omega_2/2}{\sin \omega_2/2}\right)^2 \mu^2 \\
&= \frac{1}{(2\pi)^2 N} \int_{\Pi^2} \left(\frac{\sin n_1(\omega_1 - \lambda_1)/2}{\sin(\omega_1 - \lambda_1)/2}\right)^2 \left(\frac{\sin n_2(\omega_2 - \lambda_2)/2}{\sin(\omega_2 - \lambda_2)/2}\right)^2 f_\theta(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \\
&+ \frac{1}{(2\pi)^2 N} \left(\frac{\sin n_1\omega_1/2}{\sin \omega_1/2}\right)^2 \left(\frac{\sin n_2\omega_2/2}{\sin \omega_2/2}\right)^2 \mu^2.
\end{aligned}
$$

□

Proof of Proposition 2. Note that the taper function satisfies the Parseval's identity:

$$\int_{\Pi^2} |H(\boldsymbol{\lambda})|^2 d\boldsymbol{\lambda} = (2\pi)^2 \sum_{\mathbf{s}\in D} h^2(\mathbf{s}).$$

Then, the expectation of the tapered periodogram:

$$E(I^{tap}(\boldsymbol{\omega})) = \frac{1}{(2\pi)^2 H} E(J^{tap}(\boldsymbol{\omega})\overline{J^{tap}(\boldsymbol{\omega})}) = \tag{2.95}$$

$$\frac{1}{(2\pi)^2 H} cum(J^{tap}(\boldsymbol{\omega}), \overline{J^{tap}(\boldsymbol{\omega})}) + \tag{2.96}$$

$$\frac{1}{(2\pi)^2 H} |E(J^{tap}(\boldsymbol{\omega}))|^2, \tag{2.97}$$

where J^{tap} is given by (2.20). Now, doing computations in the first term in (2.97) we have

$$\begin{aligned}
& \frac{1}{(2\pi)^2 H} cum(J^{tap}(\boldsymbol{\omega}), \overline{J^{tap}(\boldsymbol{\omega})}) \\
= & \frac{1}{(2\pi)^2 H} \sum_{\mathbf{s}\in D} \sum_{\mathbf{x}\in D} h(\mathbf{s}) h(\mathbf{x}) \exp(-i\mathbf{s}^T\boldsymbol{\omega}) \exp(i\mathbf{x}^T\boldsymbol{\omega}) C(\mathbf{s}-\mathbf{x}) \\
= & \frac{1}{(2\pi)^2 H} \int_{\Pi^2} \sum_{\mathbf{s}\in D} \sum_{\mathbf{x}\in D} h(\mathbf{s}) h(\mathbf{x}) \exp(-i(\mathbf{s}-\mathbf{x})^T\boldsymbol{\omega}) \\
& \qquad \times \exp(i(\mathbf{s}-\mathbf{x})^T\boldsymbol{\lambda}) f_\theta(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \\
= & \frac{1}{(2\pi)^2 H} \int_{\Pi^2} \sum_{\mathbf{s}\in D} \sum_{\mathbf{x}\in D} h(\mathbf{s}) h(\mathbf{x}) \exp(-i(\mathbf{s}-\mathbf{x})^T(\boldsymbol{\omega}-\boldsymbol{\lambda})) f_\theta(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \\
= & \frac{1}{(2\pi)^2 H} \int_{\Pi^2} \left\{ \sum_{\mathbf{s}\in D} h(\mathbf{s}) \exp(-i\mathbf{s}^T(\boldsymbol{\omega}-\boldsymbol{\lambda})) \right. \\
& \qquad \left. \times \sum_{\mathbf{x}\in D} h(\mathbf{x}) \exp(i\mathbf{x}^T(\boldsymbol{\omega}-\boldsymbol{\lambda})) \right\} f_\theta(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \\
= & \frac{1}{(2\pi)^2 H} \int_{\Pi^2} |H(\boldsymbol{\omega}-\boldsymbol{\lambda})|^2 f_\theta(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \\
= & \frac{1}{(2\pi)^2 H} \int_{\Pi^2} |H(\boldsymbol{\lambda})|^2 f_\theta(\boldsymbol{\omega}-\boldsymbol{\lambda}) d\boldsymbol{\lambda}
\end{aligned}$$

Finally, doing computations now on the second term, we have

$$\begin{aligned}
& \frac{1}{(2\pi)^2 H} |E(J^{tap}(\boldsymbol{\omega}))|^2 \\
= & \frac{1}{(2\pi)^2 H} \left| \sum_{\mathbf{s}\in D} \exp(-i\mathbf{s}^T\boldsymbol{\omega}) \right|^2 \mu^2 = \frac{1}{(2\pi)^2 |H(\mathbf{0})} |H(\boldsymbol{\omega})|^2 \mu^2
\end{aligned}$$

which completes the proof. □

Proof of Proposition 3. We derive the expression for the covariance using the following cumulants-based equality

$$Cov(I(\boldsymbol{\omega}), I(\boldsymbol{\lambda})) = cum(J(\boldsymbol{\omega})\overline{J(\boldsymbol{\omega})}, J(\boldsymbol{\lambda})\overline{J(\boldsymbol{\lambda})})$$

Denote:

$$A(\boldsymbol{\lambda}) = \sum_{\mathbf{s}\in D} \exp(-i\mathbf{s}^T\boldsymbol{\lambda}).$$

Since the cumulant of product variables can be expressed in terms of the sum of cumulants, we have that

$$
\begin{aligned}
& Cov(I(\boldsymbol{\omega}), I(\boldsymbol{\lambda})) = ((2\pi)^2)^3 N f_4(\boldsymbol{\omega}, -\boldsymbol{\omega}, \boldsymbol{\lambda}) + \mathcal{O}(1) \\
+ \;& (A(\boldsymbol{\omega}) f_1 + \mathcal{O}(1)) \left((2\pi)^2 A(-\boldsymbol{\omega}) f_3(-\boldsymbol{\omega}, \boldsymbol{\lambda}) + \mathcal{O}(1)\right) \\
& + \text{three similar terms} \\
+ \;& (A(\boldsymbol{\omega}) f_1 + \mathcal{O}(1)) \left(A(\boldsymbol{\lambda}) f_1 + \mathcal{O}(1)\right) \left((2\pi)^2 A(-\boldsymbol{\omega} - \boldsymbol{\lambda}) f_2(-\boldsymbol{\omega}) + \mathcal{O}(1)\right) \\
+ \;& \text{three similar terms} \\
+ \;& \left((2\pi)^2 A(\boldsymbol{\omega} + \boldsymbol{\lambda}) f_2(\boldsymbol{\omega}) + \mathcal{O}(1)\right) \left((2\pi)^2 A(-\boldsymbol{\omega} - \boldsymbol{\lambda}) f_2(-\boldsymbol{\omega}) + \mathcal{O}(1)\right) \\
+ \;& \left((2\pi)^2 A(\boldsymbol{\omega} - \boldsymbol{\lambda}) f_2(\boldsymbol{\omega}) + \mathcal{O}(1)\right) \left((2\pi)^2 A(-\boldsymbol{\omega} + \boldsymbol{\lambda}) f_2(-\boldsymbol{\omega}) + \mathcal{O}(1)\right)
\end{aligned}
$$

From these two last addends

$$
\begin{aligned}
& \left((2\pi)^2 A(\boldsymbol{\omega} + \boldsymbol{\lambda}) f_2(\boldsymbol{\omega}) + \mathcal{O}(1)\right) \left((2\pi)^2 A(-\boldsymbol{\omega} - \boldsymbol{\lambda}) f_2(-\boldsymbol{\omega}) + \mathcal{O}(1)\right) \\
+ \;& \left((2\pi)^2 A(\boldsymbol{\omega} - \boldsymbol{\lambda}) f_2(\boldsymbol{\omega}) + \mathcal{O}(1)\right) \left((2\pi)^2 A(-\boldsymbol{\omega} + \boldsymbol{\lambda}) f_2(-\boldsymbol{\omega}) + \mathcal{O}(1)\right)
\end{aligned}
$$

we obtain

$$
\begin{aligned}
& (2\pi)^4 |A(\boldsymbol{\omega} + \boldsymbol{\lambda})|^2 f_2(\boldsymbol{\omega})^2 + (2\pi)^4 |A(\boldsymbol{\omega} - \boldsymbol{\lambda})|^2 f_2(\boldsymbol{\omega})^2 = \\
& \left[\left(\frac{\sin n_1(\omega_1 + \lambda_1)/2}{\sin(\omega_1 + \lambda_1)/2} \cdot \frac{\sin n_2(\omega_2 + \lambda_2)/2}{\sin(\omega_2 + \lambda_2)/2} \right)^2 + \right. \\
& \left. \left(\frac{\sin n_1(\omega_1 - \lambda_1)/2}{\sin(\omega_1 - \lambda_1)/2} \cdot \frac{\sin n_2(\omega_2 - \lambda_2)/2}{\sin(\omega_2 - \lambda_2)/2} \right)^2 \right] f_2(\boldsymbol{\omega})^2
\end{aligned}
$$

the leading term in expression (2.62). The $\mathcal{O}(N^{-1})$ term comes from some computations on the first addend, writing f_4 in terms of cumulants. Assuming $f_1 = E(Z) = 0$, the other terms in the sum cancel. □

Proof of Proposition 4. For the complete expression of the tapered periodogram, we must consider the fractional factors where H is involved. Recall the definition of H_r:

$$
H_r(\boldsymbol{\lambda}) = \sum_{\mathbf{s} \in D} h(\mathbf{s})^r e^{-i\mathbf{s}^T \boldsymbol{\lambda}}. \tag{2.98}
$$

Thus,

$$
\begin{aligned}
& cum(J^{tap}(\boldsymbol{\omega}) \overline{J^{tap}(\boldsymbol{\omega})}) \\
= \;& ((2\pi)^2)^3 H_4(\mathbf{0}) f_4(\boldsymbol{\omega}, -\boldsymbol{\omega}, \boldsymbol{\lambda}) + \mathcal{O}(1) \\
+ \;& \left[(2\pi)^2 H_2(\boldsymbol{\omega} + \boldsymbol{\lambda}) f_2(\boldsymbol{\lambda}) + \mathcal{O}(1)\right] \left[(2\pi)^2 H_2(-\boldsymbol{\omega} - \boldsymbol{\lambda}) f_2(\boldsymbol{\lambda})\right] \\
+ \;& \left[(2\pi)^2 H_2(\boldsymbol{\omega} - \boldsymbol{\lambda}) f_2(\boldsymbol{\lambda}) + \mathcal{O}(1)\right] \left[(2\pi)^2 H_2(-\boldsymbol{\omega} + \boldsymbol{\lambda}) + \mathcal{O}(1)\right],
\end{aligned}
$$

where J^{tap} is defined as in (2.20). Then, the expression for the covariance is obtained

$$Cov(I^{tap}(\boldsymbol{\omega}), I^{tap}(\boldsymbol{\lambda})) = |H_2(\mathbf{0})|^{-2} \left\{ |H_2(\boldsymbol{\omega}+\boldsymbol{\lambda})|^2 + |H_2(\boldsymbol{\omega}-\boldsymbol{\lambda})|^2 \right\} f_2(\boldsymbol{\lambda}) + \mathcal{O}(N^{-1})$$

□

Proof of Proposition 5. By standard properties of cumulants, we obtain

$$\begin{aligned} E(I(\boldsymbol{\omega},\tau)) &= \frac{1}{(2\pi)^3 N} E(J(\boldsymbol{\omega},\tau)\overline{J(\boldsymbol{\omega},\tau)}) \\ &= \frac{1}{(2\pi)^3 N} cum(J(\boldsymbol{\omega},\tau)\overline{J(\boldsymbol{\omega},\tau)}) + \frac{1}{(2\pi)^3 N} \left| E(J(\boldsymbol{\omega},\tau)) \right|^2, \end{aligned}$$

where J is defined as in equation (2.65). By properties of the sum of exponentials, the second term in the above expression can be further developed to give:

$$\begin{aligned} \frac{1}{(2\pi)^3 N} \left| E(J(\boldsymbol{\omega},\tau)) \right|^2 &= \frac{1}{(2\pi)^3 N} \left| \sum_{(\mathbf{s},t)\in D\times T} \exp(-i\mathbf{s}^T\boldsymbol{\omega}) \exp(-it\tau) \right|^2 \mu^2 \\ = \frac{1}{(2\pi)^3 N} \left| \sum_{\mathbf{s}\in D} \exp(-i\mathbf{s}^T\boldsymbol{\omega}) \sum_{t\in T} \exp(-it\tau) \right|^2 \mu^2 \\ = \frac{1}{(2\pi)^3 N} \left(\frac{\sin n_1\omega_1/2}{\sin \omega_1/2} \right)^2 \left(\frac{\sin n_2\omega_2/2}{\sin \omega_2/2} \right)^2 \left(\frac{\sin n\tau/2}{\sin \tau/2} \right)^2 \mu^2 \end{aligned}$$

Finally, the first term can also be decomposed as follows

$$\begin{aligned} &\frac{1}{(2\pi)^3 N} cum(J(\boldsymbol{\omega},\tau)\overline{J(\boldsymbol{\omega},\tau)}) \\ = &\frac{1}{(2\pi)^3 N} \sum_{(\mathbf{s},t)\in D\times T} \sum_{(\mathbf{s}^*,t^*)\in D\times T} e^{(-i\mathbf{s}^T\boldsymbol{\omega})} e^{(-it\tau)} e^{(-i\mathbf{s}^{*T}\boldsymbol{\omega})} e^{(-it^*\tau)} Cov(\mathbf{s}-\mathbf{s}^*, t-t^*) \\ = &\frac{1}{(2\pi)^3 N} \int_{\Pi^3} \sum_{(\mathbf{s},t)\in D\times T} \sum_{(\mathbf{s}^*,t^*)\in D\times T} e^{-i(\mathbf{s}-\mathbf{s}^*)^T\boldsymbol{\omega}} e^{-i(t-t^*)\tau} e^{i(\mathbf{s}-\mathbf{s}^*)^T\boldsymbol{\lambda}} e^{i(t-t^*)\nu} f_\theta(\boldsymbol{\lambda},\nu) d\boldsymbol{\lambda} d\nu \\ = &\frac{1}{(2\pi)^3 N} \int_{\Pi^3} \sum_{(\mathbf{s},t)\in D\times T} \sum_{(\mathbf{s}^*,t^*)\in D\times T} e^{-i(\mathbf{s}-\mathbf{s}^*)^T(\boldsymbol{\omega}-\boldsymbol{\lambda})} e^{-i(t-t^*)(\tau-\nu)} f_\theta(\boldsymbol{\lambda},\nu) d\boldsymbol{\lambda} d\nu \\ = &\frac{1}{(2\pi)^3 N} \int_{\Pi^3} \sum_{\mathbf{s},\mathbf{s}^*\in D} e^{-i(\mathbf{s}-\mathbf{s}^*)^T(\boldsymbol{\omega}-\boldsymbol{\lambda})} \sum_{t,t^*\in T} e^{-i(t-t^*)(\tau-\nu)} f_\theta(\boldsymbol{\lambda},\nu) d\boldsymbol{\lambda} d\nu \\ = &\frac{1}{(2\pi)^3 N} \int_{\Pi^3} B_{\boldsymbol{\omega},\tau}(\boldsymbol{\lambda},\nu) f_\theta(\boldsymbol{\lambda},\nu) d\boldsymbol{\lambda} d\nu, \end{aligned}$$

which completes the proof. □

Proof of Proposition 6. The expectation of the tapered spatio-temporal periodogram is given by

$$\begin{aligned} E(I^{tap}(\boldsymbol{\omega},\tau)) &= \frac{1}{(2\pi)^3 HG} E(J^{tap}(\boldsymbol{\omega},\tau)\overline{J^{tap}(\boldsymbol{\omega},\tau)}) = \\ &= \frac{1}{(2\pi)^3 HG} \sum_{(\mathbf{s},t)\in D\times T} \sum_{(\mathbf{s}^*,t^*)\in D\times T} h(\mathbf{s})h(\mathbf{s}^*)g(t)g(t^*) \\ &\qquad \times e^{-i(\mathbf{s}-\mathbf{s}^*)^T\boldsymbol{\omega}} e^{-i(t-t^*)\tau} Cov(Z(\mathbf{s},t), Z(\mathbf{s}^*,t^*)) \\ &+ \frac{1}{(2\pi)^3 HG} \left| \sum_{(\mathbf{s},t)\in D\times T} h(\mathbf{s})g(t)e^{-i\mathbf{s}^T\boldsymbol{\omega}} e^{-it\tau} \right|^2 \mu^2 \end{aligned}$$

Doing computations in the second term of the above equation, and considering H and G obtained as in (2.98), we have

$$\begin{aligned} &\frac{1}{(2\pi)^3 H} \left| \sum_{(\mathbf{s},t)\in D\times T} h(\mathbf{s})g(t)e^{-i\mathbf{s}^T\boldsymbol{\omega}} e^{-it\tau} \right|^2 \mu^2 \\ &= \frac{1}{(2\pi)^3 HG} \left| \sum_{\mathbf{s}\in D} h(\mathbf{s})e^{-i\mathbf{s}^T\boldsymbol{\omega}} \right|^2 \left| \sum_{t\in T} g(t)e^{-it\tau} \right|^2 \mu^2 \\ &= \frac{1}{(2\pi)^3 HG} |H(\boldsymbol{\omega})|^2 |G(\tau)|^2 \mu^2. \end{aligned}$$

Finally, for the first term (without the fraction) we have

$$\begin{aligned} &\sum_{(\mathbf{s},t)\in D\times T} \sum_{(\mathbf{s}^*,t^*)\in D\times T} h(\mathbf{s})h(\mathbf{s}^*)g(t)g(t^*)e^{-i(\mathbf{s}-\mathbf{s}^*)^T\boldsymbol{\omega}} e^{-i(t-t^*)\tau} Cov(Z(\mathbf{s},t), Z(\mathbf{s}^*,t^*)) \\ &= \sum_{(\mathbf{s},t)\in D\times T} \sum_{(\mathbf{s}^*,t^*)\in D\times T} h(\mathbf{s})h(\mathbf{s}^*)g(t)g(t^*)e^{-i(\mathbf{s}-\mathbf{s}^*)^T\boldsymbol{\omega}} e^{-i(t-t^*)\tau} C(\mathbf{s}-\mathbf{s}^*, t-t^*) \\ &= \int_{\Pi^3} \sum_{(\mathbf{s},t)\in D\times T} \sum_{(\mathbf{s}^*,t^*)\in D\times T} h(\mathbf{s})h(\mathbf{s}^*)g(t)g(t^*)e^{-i(\mathbf{s}-\mathbf{s}^*)^T(\boldsymbol{\omega}-\boldsymbol{\lambda})} e^{-i(t-t^*)(\tau-\nu)} f_\theta(\boldsymbol{\lambda},\nu) d\boldsymbol{\lambda} d\nu \\ &= \int_{\Pi^3} \left| \sum_{\mathbf{s}\in D} h(\mathbf{s})e^{-i\mathbf{s}^T(\boldsymbol{\omega}-\boldsymbol{\lambda})} \right|^2 \left| \sum_{t\in T} g(t)e^{-it(\tau-\nu)} \right|^2 f_\theta(\boldsymbol{\lambda},\nu) d\boldsymbol{\lambda} d\nu \\ &= \int_{\Pi^3} |H(\boldsymbol{\omega}-\boldsymbol{\lambda})|^2 |G(\tau-\nu)|^2 f_\theta(\boldsymbol{\lambda},\nu) d\boldsymbol{\lambda} d\nu \end{aligned}$$

which completes the proof. □

2.7.3 Proofs of Section 2.4.

This section is devoted to the introduction of the smoothed kernel estimates of the spectral density and it also includes some results that will be used later. The lag-**u** sample autocovariances are

defined in equation (2.4). The sample autocovariance estimate is not an unbiased estimator of the theoretical covariance $C_\Delta(\mathbf{u})$. The expectation of this estimator is given by:

$$E(\hat{C}_\Delta(\mathbf{u})) = C_\Delta(\mathbf{u})(\eta_n(\mathbf{u}) + 1),$$

where

$$\eta_n(\mathbf{u}) = \left(1 - \frac{|u_1|}{n_1}\right)\left(1 - \frac{|u_2|}{n_2}\right) - 1.$$

For a fixed $\mathbf{u}$, we can see that, as $n_1, n_2 \to \infty$:

$$\eta_n(\mathbf{u}) = \left(-\sum_{j=1}^{2} \frac{|u_j|}{n_j}\right) + (1 + o(1))$$

For $\mathbf{u}$ such that

$$\left|\sum_{j=1}^{2} \frac{|u_j|}{n_j}\right| \geq c \sum_{j=1}^{2} \frac{1}{n_j}$$

we can apply the inequality between arithmetic and geometric means and conclude that the bias of the estimate is of order $1/\sqrt{N}$, where $N = n_1 n_2$. This is the so-called edge-effect (Guyon (1982)).

In order to obtain the covariance structure of the sample covariance estimates, assume that Z is stationary to the fourth moment. Under this assumption, the fourth order cumulants behave like a stationary sequence and the expression of the covariance is much simpler. For $\mathbf{s}, \mathbf{r}$ in the grid

$$\begin{aligned}
&Cov(\hat{C}_\Delta(\mathbf{s}), \hat{C}_\Delta(\mathbf{r})) = E(\hat{C}_\Delta(\mathbf{s}) \cdot \hat{C}_\Delta(\mathbf{r})) - E(\hat{C}_\Delta(\mathbf{s}))E(\hat{C}_\Delta(\mathbf{r})) = \\
&\frac{1}{n_1^2 n_2^2} E\left(\sum_{\boldsymbol{\tau} \in \mathcal{T}} \sum_{\mathbf{u} \in \mathcal{U}} Z(\Delta\mathbf{s}) Z(\Delta\mathbf{s} + \Delta\boldsymbol{\tau}) Z(\Delta\mathbf{r}) Z(\Delta\mathbf{r} + \Delta\mathbf{u})\right) - \\
&\left(1 - \frac{|s_1|}{n_1}\right)\left(1 - \frac{|s_2|}{n_2}\right)\left(1 - \frac{|r_1|}{n_1}\right)\left(1 - \frac{|r_2|}{n_1}\right) C_\Delta(\mathbf{s}) C_\Delta(\mathbf{r})
\end{aligned}$$

Applying the following result for quadrivariate zero-mean distributions:

$$\begin{aligned}
&E(Z(\Delta\boldsymbol{\tau})Z(\Delta\boldsymbol{\tau} + \Delta\mathbf{s})Z(\Delta\mathbf{u})Z(\Delta\mathbf{u} + \Delta\mathbf{r})) = \\
&= E(Z(\Delta\boldsymbol{\tau})Z(\Delta\boldsymbol{\tau} + \Delta\mathbf{s}))E(Z(\Delta\mathbf{u})Z(\Delta\mathbf{u} + \Delta\mathbf{r})) + \\
&E(Z(\Delta\boldsymbol{\tau})Z(\Delta\mathbf{u}))E(Z(\Delta\boldsymbol{\tau} + \Delta\mathbf{s})Z(\Delta\mathbf{u} + \Delta\mathbf{r})) + \\
&E(Z(\Delta\boldsymbol{\tau})Z(\Delta\mathbf{u} + \Delta\mathbf{r}))E(Z(\Delta\boldsymbol{\tau} + \Delta\mathbf{s})Z(\Delta\mathbf{u})) + \\
&k_4(\Delta\mathbf{r} - \Delta\mathbf{s}, \Delta\boldsymbol{\tau}, \Delta\mathbf{u}),
\end{aligned}$$

where k_4 is the finite fourth-order cumulant (Priestley (1981), p. 404). The first addend in the covariance is given by:

$$\frac{1}{n_1^2 n_2^2}\sum_{\boldsymbol{\tau}\in\mathcal{T}}\sum_{\mathbf{r}\in\mathcal{U}}\{C_\Delta(\mathbf{s})C_\Delta(\mathbf{r})+C_\Delta(\mathbf{r}-\mathbf{s}+\mathbf{u}-\boldsymbol{\tau})C_\Delta(\mathbf{u}-\boldsymbol{\tau})+$$
$$C_\Delta(\mathbf{r}+\mathbf{u}-\boldsymbol{\tau})C_\Delta(\mathbf{u}-\mathbf{s}-\boldsymbol{\tau})+k_4(\mathbf{r}-\mathbf{s},\boldsymbol{\tau},\mathbf{u})\}$$

and the first part cancels with the second addend. In order to obtain a simpler expression, consider

$$Cov(\hat{C}_\Delta(\mathbf{s}),\hat{C}_\Delta(\mathbf{s}+\mathbf{r}))=$$
$$\frac{1}{n_1 n_2}\sum_{u_1=1-n_1+s_1}^{n_1-s_1-r_1-1}\sum_{u_2=1-n_2+s_2}^{n_2-s_2-r_2-1}\left(1-\frac{\eta(u_1)+s_1+r_1}{n_1}\right)\left(1-\frac{\eta(u_2)+s_2+r_2}{n_2}\right)$$
$$\{C_\Delta(\mathbf{u})C_\Delta(\mathbf{u}+\mathbf{r})+C_\Delta(\mathbf{u}+\mathbf{s}+\mathbf{r})C_\Delta(\mathbf{u}-\mathbf{s})+k_4(\mathbf{u}+\mathbf{s}+\mathbf{r})\}$$

where

$$\eta(u_j)=\begin{cases}u_j & \text{if} \quad u_j>0\\ 0 & \text{if} \quad -r_j\le u_j\le 0\\ -u_j-r_j & \text{if} \quad 1-n_j+s_j\le u_j\le -r_j.\end{cases}$$

As a particular case, the variance of the sample-covariance is obtained:

$$Var(\hat{C}_\Delta(\mathbf{s}))=$$
$$\frac{1}{n_1 n_2}\sum_{u_1=1-n_1+s_1}^{n_1-s_1-1}\sum_{u_2=1-n_2+s_2}^{n_2-s_2-1}\left(1-\frac{|u_1|+s_1}{n_1}\right)\left(1-\frac{|u_2|+s_2}{n_2}\right)$$
$$\left\{C_\Delta^2(\mathbf{u})+C_\Delta(\mathbf{u}+\mathbf{s})C_\Delta(\mathbf{u}-\mathbf{s})+k_4(\mathbf{u}+\mathbf{s})\right\}.$$

Proof of Lemma 3. We will just give a sketch of the proof. It can be obtained following standard arguments from spectral density estimation. The asymptotic bias is dominated by the $\alpha_{1n,\Delta}$ term, which is given by:

$$\alpha_{1n,\Delta}=\frac{\Delta^2}{(2\pi)^2}k_q\sum_{j=1}^{2}m_j^{-q}\sum_{\mathbf{u}}|u_j|^2C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}$$

Since the trace of the Hessian matrix of f_Δ in $\boldsymbol{\lambda}$ is

$$\operatorname{tr}\left(Hf_\Delta(\boldsymbol{\lambda})\right)=\frac{-2\Delta^4}{(2\pi)^2}\sum_{\mathbf{u}}(u_1^2+u_2^2)C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}},$$

then, the expression of the bias can be simplified by

$$\alpha_{1n,\Delta}=\frac{-k_q}{2\Delta^2}\operatorname{tr}\left(Hf_\Delta(\boldsymbol{\lambda})\right)\sum_{j=1}^{2}m_j^{-q}.$$

□

Proof of Lemma 4. Recall the representation of the smoothed-kernel estimates in terms of the periodogram given by (2.30). The covariance structure for these estimates, for $\boldsymbol{\omega}, \boldsymbol{\nu} \in \Pi^2_\Delta$ can be approximated by:

$$\begin{aligned}
&Cov(\tilde{f}_\Delta(\boldsymbol{\omega}), \tilde{f}_\Delta(\boldsymbol{\nu})) \approx \\
&\frac{(2\pi)^2\Delta^2 m_1 m_2}{n_1 n_2} \int W_{n,\Delta}(\boldsymbol{\omega}-\boldsymbol{\theta}) W_{n,\Delta}(\boldsymbol{\nu}-\boldsymbol{\theta}') \\
&\cdot \left(F_{N,\Delta}(\boldsymbol{\omega}+\boldsymbol{\nu}) + F_{N,\Delta}(\boldsymbol{\omega}-\boldsymbol{\nu})\right) f_\Delta(\boldsymbol{\theta}) f_\Delta(\boldsymbol{\theta}') d\boldsymbol{\theta} d\boldsymbol{\theta}',
\end{aligned}$$

integrating over $\Pi^2_\Delta \times \Pi^2_\Delta$. As a particular case, an approximation of the variance for $\boldsymbol{\omega} \in \Pi^2_\Delta$:

$$Var(\tilde{f}_\Delta(\boldsymbol{\omega})) \approx (1+\delta_{0,\pi}) \frac{(2\pi)^2\Delta^2 m_1 m_2}{n_1 n_2} \int_{\Pi^2_\Delta} f^2_\Delta(\boldsymbol{\theta}) W^2_{n,\Delta}(\boldsymbol{\omega}-\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

For scale-parameter windows, by the Parseval's Theorem, we can approximate the variance by:

$$Var(\tilde{f}_\Delta(\boldsymbol{\omega})) \approx (1+\delta_{0,\pi}) \frac{(2\pi)^2\Delta^2 m_1 m_2}{n_1 n_2} f^2_\Delta(\boldsymbol{\omega}) \left(\int k^2(\mathbf{u}) d\mathbf{u} \right)^2.$$

□

Proof of Proposition 7. Consider the difference between the smooth kernel estimator bias and the aliased spectral density f_Δ, at a frequency $\boldsymbol{\lambda}$

$$\begin{aligned}
&E(\tilde{f}_\Delta(\boldsymbol{\lambda})) - f_\Delta(\boldsymbol{\lambda}) = \\
&\frac{\Delta^2}{(2\pi)^2} \sum_{\mathbf{u}\in\mathcal{U}} (k_n(\mathbf{u})-1) C_\Delta(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}} \\
&+ \frac{\Delta^2}{(2\pi)^2} \sum_{\mathbf{u}\in\mathcal{U}} k_n(\mathbf{u}) \eta_n(\mathbf{u}) C_\Delta(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}} \\
&- \frac{\Delta^2}{(2\pi)^2} \sum_{\mathbf{u}\in\mathbb{Z}^2-\mathcal{U}} C_\Delta(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}} = (A)+(B)+(C).
\end{aligned}$$

First, we proof that $(A) = \alpha_{1n,\Delta}(1+o(1))$. It is easy to see that

$$k_n(\mathbf{u}) - 1 = \sum_{j=1}^{2} \left(k\left(\frac{u_j}{m_j}\right) - 1 \right) + v_n$$

where v_n is linear in products of two $k\left(\frac{u_j}{m_j}\right)-1$ (or more, in the case $d>2$). Then:

$$\frac{(2\pi)^2}{\Delta^2}(A)=\sum_{\mathbf{u}\in\mathcal{U}}(k_n(\mathbf{u})-1)C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}=$$
$$\sum_{\mathbf{u}\in\mathcal{U}}\sum_{j=1}^{2}\left(k\left(\frac{u_j}{m_j}\right)-1\right)C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}+\sum_{\mathbf{u}}v_nC_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}=$$
$$\sum_{\mathbf{u}\in\mathcal{U}}v_nC_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}\left(1+\frac{\sum_{\mathbf{u}}\sum_{j=1}^{2}\left(k\left(\frac{u_j}{m_j}\right)-1\right)C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}}{\sum_{\mathbf{u}}v_nC_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}}\right)=$$
$$\sum_{\mathbf{u}\in\mathcal{U}}v_nC_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}(1+o(1))$$

For any subset L of $\{1,2\}$, proceeding as in (Hannan (1970), p.284)

$$\sum_{\mathbf{u}\in\mathcal{U}}\prod_{j\in L}\left(k\left(\frac{u_j}{m_j}\right)-1\right)C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}=$$

$$\prod_{j\in L}m_j^{-q}\sum_{\mathbf{u}\in\mathcal{U}}\left(\prod_{j\in L}\frac{k(u_j/m_j)-1}{|u_j/m_j|^q}|u_j|^q\right)C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}$$

Then

$$(A)=\frac{\Delta^2}{(2\pi)^2}\sum_{j=1}^{2}m_j^{-q}k_q\sum_{\mathbf{u}\in\mathcal{U}}|u_j|^qC_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}(1+o(1)). \tag{2.99}$$

For the second addend, (B), we must take into account that:

$$\eta_n(\mathbf{u})=-\sum_{j=1}^{2}\frac{|u_j|}{n_j}+s_n \tag{2.100}$$

where s_n is linear in products of two $|u_j|/n_j$ (two or more, in the case that $d>2$). Then:

$$\frac{(2\pi)^2}{\Delta^2}(B)=\sum_{\mathbf{u}\in\mathcal{U}}k_n(\mathbf{u})\left(-\sum_{j=1}^{2}\frac{|u_j|}{n_j}+s_n\right)C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}=$$
$$\sum_{\mathbf{u}\in\mathcal{U}}k_n(\mathbf{u})\left(-\sum_{j=1}^{2}\frac{|u_j|}{n_j}\right)C_\Delta(\mathbf{u})+\sum_{\mathbf{u}}k_n(\mathbf{u})s_nC_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}=$$
$$\sum_{\mathbf{u}\in\mathcal{U}}k_n(\mathbf{u})s_nC_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}\left(1+\frac{\sum_{\mathbf{u}}k_n(\mathbf{u})\left(-\sum_{j=1}^{2}\frac{|u_j|}{n_j}\right)C_\Delta(\mathbf{u})}{\sum_{\mathbf{u}}k_n(\mathbf{u})s_nC_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}}\right)=$$
$$\sum_{\mathbf{u}\in\mathcal{U}}k_n(\mathbf{u})s_nC_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}(1+o(1))$$

Now, by Assumption 1, we have that $k_n(\mathbf{u}) \to 1$ for every $\mathbf{u}$ as $n \to \infty$. Therefore, the first addend can be approximated by:

$$\prod_{j\in L} n_j^{-1} \sum_{\mathbf{u}\in\mathcal{U}} \left(\prod_{j\in L} |u_j| \right) C_\Delta(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}$$

Then:

$$(B) = \frac{\Delta^2}{(2\pi)^2} \sum_{j=1}^{2} n_j^{-1} \sum_{\mathbf{u}\in\mathcal{U}} |u_j| C_\Delta(\mathbf{u}) e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}} (1 + o(1)) \tag{2.101}$$

The last addend is the tail effect, and it can be bounded by:

$$\mathcal{O}\left(\frac{\Delta^2}{(2\pi)^2} \sum_{j=1}^{2} n_j^{-q} \sum_{|u_j|>n_j} |u_j|^q |C_\Delta(\mathbf{u})| \right) = o\left(\Delta^2 \sum_{j=1}^{2} n_j^{-q} \right) \tag{2.102}$$

By Assumption 4, we ensure the convergence of $f_\Delta(\boldsymbol{\lambda})$ to $f(\boldsymbol{\lambda})$, at a rate $\mathcal{O}(\Delta^\tau)$, for $\tau > 2$ (Stein (1999)), and the result is proved. □

Proof of Proposition 8. In order to proof Proposition 8, define the tapered periodogram as

$$I_\Delta^h(\boldsymbol{\lambda}) = \frac{\Delta^2}{(2\pi)^2 H_n} \left| \sum_{\mathbf{s}\in D} \left(\prod_{j=1}^{2} h_{j,s_j} \right) Z(\Delta\mathbf{s}) e^{-i\Delta\mathbf{s}^T\boldsymbol{\lambda}} \right|^2. \tag{2.103}$$

The smooth kernel tareped estimate for the spectral density can be written as:

$$\tilde{f}_\Delta^h(\boldsymbol{\lambda}) = \int_{\Pi_\Delta^2} W_{n,\Delta}(\boldsymbol{\lambda} - \boldsymbol{\nu}) I_{n,\Delta}^h(\boldsymbol{\nu}) d\boldsymbol{\nu}, \tag{2.104}$$

and the expectation of (2.103) is:

$$E\left(I_\Delta^h(\boldsymbol{\lambda}) \right) = \int_{\Pi_\Delta^2} f_\Delta(\boldsymbol{\nu}) \prod_{j=1}^{2} g_j(\omega_j - \nu_j) d\boldsymbol{\nu},$$

where

$$g_j(\omega_j) = \left(2\pi \sum_{s_j=1}^{n_j} h_{j,s_j}^2 \right)^{-1} \left| \sum_{s_j=1}^{n_j} h_{j,s_j} e^{is_j\omega_j} \right|^2$$

The proof is analogous as Theorem 2 in (Robinson (2006)).

□

Proof of Proposition 9. A first strategy for finding an optimal bandwidth, in the sense that the MSE is minimized, is considering the order of the bias and the variance and try to minimize the order of the MSE. Considering the equation of the MSE orders, we have:

$$\phi(m_1, m_2) = \left(\Delta^2 \sum_{j=1}^{2} m_j^{-q} \right)^2 + \frac{\Delta^2 m_1 m_2}{n_1 n_2}.$$

To minimize, this function, we have to solve the following equations:

$$\frac{\partial \phi}{\partial m_1} = -2q\Delta^4 m_1^{-2q-1} - 2q\Delta^4 m_1^{-q-1} m_2^{-q} + \frac{\Delta^2 m_2}{n_1 n_2},$$

$$\frac{\partial \phi}{\partial m_2} = -2q\Delta^4 m_2^{-2q-1} - 2q\Delta^4 m_2^{-q-1} m_1^{-q} + \frac{\Delta^2 m_1}{n_1 n_2}.$$

The solution of this equation system implies that $m_1 = m_2 = m$, and

$$m = c\,(4q\Delta n_1 n_2)^{\frac{1}{2q+2}}\,.$$

Therefore, we have to find c such that the MSE is minimized. The MSE is given by:

$$\text{MSE} = \frac{k_q^2}{4\Delta^4}\text{tr}^2\left(H f_\Delta(\boldsymbol{\lambda})\right)^2 \left(\frac{2}{m^q}\right)^2 + (1+\delta_{0,\pi})\frac{4\pi^2\Delta^2}{n_1 n_2} m^2 f_\Delta^2(\boldsymbol{\lambda}) \left(\int k^2(u)du\right)^2.$$

Replacing m by its expression in terms of n_1, n_2 and Δ, the constant c which minimizes the MSE is:

$$c = \left(\frac{q^{\frac{q-1}{q+1}} k_q^2 \text{tr}(H f_\Delta(\boldsymbol{\lambda}))^2 (n_1 n_2)^{\frac{2q}{q+1}}}{2^{\frac{3q+7}{q+1}} \pi \Delta^5 f_\Delta(\boldsymbol{\lambda}) \left(\int k^2(u)du\right)^2}\right)^{\frac{1}{2q+2}}.$$

□

2.7.4 A note on the order of the periodogram bias.

The main drawback of the periodogram as a estimate of the spectrum is its lack of consistency. It can be easily proved that its variance does not tend to zero as the number of observations increases. For data from a process Z observed on a lattice, the periodogram is defined as:

$$I(\boldsymbol{\lambda}) = \frac{\Delta^2}{(2\pi)^2 n_1 n_2}\left|\sum_{\mathbf{s}\in D} Z(\Delta\mathbf{s}) e^{-i\Delta\mathbf{s}^T\boldsymbol{\lambda}}\right|^2.$$

Consider also the shrinking asymptotic model defined above. Asymptotic bias and covariance structure of the periodogram in this setting have been studied by Fuentes (2002). We give here an alternative study in order to analyze the bias.

The bias of the periodogram as an estimator of the spectral density f is given by:

$$E(I(\boldsymbol{\lambda})) - f(\boldsymbol{\lambda}) = E(I(\boldsymbol{\lambda})) - f_\Delta(\boldsymbol{\lambda}) + f_\Delta(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda}) = \text{Bias } + \text{Aliasing}$$

The bias term:

$$
\begin{aligned}
&E(I(\boldsymbol{\lambda})) - f_\Delta(\boldsymbol{\lambda}) = \\
&\frac{\Delta^2}{(2\pi)^2 n_1 n_2} \sum_{\mathbf{s}\in D} \sum_{\mathbf{x}\in D} E(Z(\Delta\mathbf{s})Z(\Delta\mathbf{x}))e^{-i\Delta(\mathbf{s}-\mathbf{x})^T\boldsymbol{\lambda}} - f_\Delta(\boldsymbol{\lambda}) = \\
&\frac{\Delta^2}{(2\pi)^2} \sum_{\mathbf{u}\in\mathcal{U}} \left(1 - \frac{|u_1|}{n_1}\right)\left(1 - \frac{|u_2|}{n_2}\right) C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}} \\
&-\frac{\Delta^2}{(2\pi)^2} \sum_{\mathbf{u}\in\mathbb{Z}^2} C_\Delta(\mathbf{u})e^{-i\mathbf{u}^T\boldsymbol{\lambda}} = \\
&\frac{\Delta^2}{(2\pi)^2} \sum_{\mathbf{u}\in\mathcal{U}} (k_n(\mathbf{u}) - 1)C_\Delta(\mathbf{u})e^{-i\mathbf{u}^T\boldsymbol{\lambda}} - \frac{\Delta^2}{(2\pi)^2} \sum_{\mathbb{Z}^2-\mathcal{U}} C_\Delta(\mathbf{u})e^{-i\mathbf{u}^T\boldsymbol{\lambda}}
\end{aligned}
$$

where

$$
k_n(\mathbf{u}) = \left(1 - \frac{|u_1|}{n_1}\right)\left(1 - \frac{|u_2|}{n_2}\right).
$$

The second of the addends is the tail effect. So, we are working with Bartlett-type weights (no window), and proceeding as in the proof of Theorem 1, we have:

$$
E(I(\boldsymbol{\lambda})) - f(\boldsymbol{\lambda}) = \alpha_{2n,\Delta} + o\left(\Delta^2 \sum_{j=1}^{2} n_j^{-1}\right) + o\left(\Delta^2 \sum_{j=1}^{2} n_j^{-q}\right) + \mathcal{O}(\Delta^\tau),
$$

where

$$
\alpha_{2n,\Delta} = \frac{\Delta^2}{(2\pi)^2} \sum_{j=1}^{2} n_j^{-1} \sum_{\mathbf{u}\in\mathbb{Z}^2} |u_j| C_\Delta(\mathbf{u})e^{-i\Delta\mathbf{u}^T\boldsymbol{\lambda}}.
$$

Therefore, the order of the bias is given by:

$$
\mathcal{O}\left(\Delta^2 \sum_{j=1}^{2} n_j^{-1}\right) + \mathcal{O}\left(\Delta^{-\tau}\right), \quad \tau > 2.
$$

Chapter 3

Simulation of spatial dependence structures

Contents

In most applied works in spatial statistics, one can not avoid the use of simulation techniques for spatial (lattice or geostatistical) dependent data, or even in more theoretical developments, when trying to illustrate the performance of a certain statistical technique. For instance, when the approximation of any distributional characteristic of a statistic is required (e.g. a p-value). Spatial random fields simulation has been an important research topic in spatial statistics.

In the geostatistical context, Gaussian process generation, with a certain covariance structure, can be done using the Cholesky factorization (Cressie (1993), pp.201-203) of the variance-covariance matrix, but such a matrix factorization may be computationally expensive. The most well-known method for generating a multidimensional stationary process, avoiding the factorization of the variance-covariance matrix, is the Turning-Bands method (e.g. Chilès and Delfiner (1999), pp.472-477). The success of this method relies on the fact that it simplifies the multidimensional simulations to the one-dimensional case.

In the Markov random field context, Moura and Balram (1992) consider the problem of generating a non-causal Gaussian-Markov random field defined on finite lattices. The characterization of the field structure is not given in terms of its covariance matrix, but on its potential or precision matrix (the inverse of the covariance matrix). A recursive structure is developed for this type of processes, consisting of two equivalent one-sided representations obtained by the Cholesky factorization of the potential matrix.

Also based on the potential matrix, Rue (2001) proposes an algorithm which takes advantages of the Markov properties of the field, applying numerical techniques for sparse matrices. For regularly-spaced observations with Gaussian correlations, Martin (2000) obtains the theoretical autoregressive and moving-average representations. This fact allows for the exact simulation of a set of observations, given a certain vector of innovations. The author also points out that the moving-average form is preferable for simulation but the autoregression and moving-average coefficients are difficult to approximate.

The methods introduced above, both for geostatistical or lattice data contexts, involve the covariance matrix. An alternative to these techniques is spectral simulation. On this context, Shinozuka (1971) proposes a method for simulating multivariate and multidimensional random processes, with a specified spectral density. Another method for generating a stationary random field with an imposed model of covariance function is the so-called Fourier Integral Method (Borgman *et al.* (1984), Pardo-Igúzquiza and Chica-Olmo (1993), Yao (1998) and Yao (2004)). For instance, Pardo-Igúzquiza and Chica-Olmo (1993) describe this algorithm in the multidimensional case and their results are compared with Shinozuka's method, in one-dimension, and with

Turning-Bands in two and three dimensions. One of the main advantages of these methods is their computational efficiency, since the computations involved can be done using the Fast Fourier Transform algorithm.

We may be interested in the simulation of spatial processes realizations, with a certain covariance (known or unknown) structure. If our aim is to obtain a realization of a spatial process from which we have a set of observations and the underlying covariance function is not known, we must estimate first the covariance from these data.

Simulation methods that do honor the observed data (simulated values at observed locations agree with observed values) are known as conditional simulation methods. However, simulation procedures that do not honor the data (maybe because no data has been collected) are known as unconditional simulation methods. We may have obtained some observations of the process, and our aim could be simulate the process in such a way that the new realizations are consistent with the observed data. This is conditional simulation which is not in the scope of this chapter. Note that we focus on non-conditional simulation, although conditiona simulation procedures can be obtained from unconditional methdos (see Cressie (1993), Section 3.6.2).

On the other hand, in many situations, one only needs to simulate statistics related to the dependence structure of the process. For instance, simulate covariance or spectral density estimators, in order to make inference on these functions. Concretly, one may be interested in approximating the distribution of the classical nonparametric estimator of the spectral density, the periodogram (or different estimators derived from this one). In this case, it is worth it to have an adequate method for generating periodogram values. Different bootstrap approaches, based on resampling the periodogram, have been proposed in time series context. For instance, Franke and Härdle (1992) introduce a bootstrap technique for kernel spectral estimates, considering the periodogram as the response in an approximate multiplicative regression model. This method is extended in Dahlhaus and Janas (1996) for ratio statistics and Whittle estimates. In Paparoditis and Politis (1999), a local bootstrap method is proved to be consistent for kernel estimates, ratio statistics and Whittle estimates. A more complex procedure is given by Kreiss and Paparoditis (2003), where the authors propose a combination of time domain parametric and frequency domain non parametric bootstrap. Instead of considering periodogram values, Fan and Zhang (2004) propose a parametric method for generating log-periodogram values, regarding the fact that the log-periodogram can be obatined as the response in an additive regression model. Extensions of these methods to the multidimensional setting must be done carefully. Apart from some challenges in the theoretical developments, the results obtained from straightforward extensions may not be as satisfactory as in the one-dimensional case, as we will be shown in a simulation study. Another

difficulty that we find when constructing simulation methods for spatial process is the continuous character of geostatistical data. In this case, the aliasing problem arises.

In this chapter, we focus our attention on spectral simulation methods and we propose a modification of the Fourier Integral Method which exhibits a better performance. Our purpose is to generate unconditional simulations. The contents of this chapter can be found in Crujeiras and Fernández-Casal (2006).

3.1 Some background on simulation techniques for spatial processes

3.1.1 Spectral simulation methods.

We have already seen that any stationary random field has a spectral representation (1.62). Therefore, simulating a stationary process with a certain covariance (or equivalently, a certain spectral density), can be done by simulation the corresponding spectral process $\mathcal{Y}$ satisfying

$$E(|\mathcal{Y}(B)|^2) = F(B), \quad \text{for any Borel set } B \in \mathbb{R}^2.$$

The spectral process $\mathcal{Y}$ can be decomposed in its real and imaginary parts, namely $U(\boldsymbol{\lambda}) = Re(\mathcal{Y}(\boldsymbol{\lambda}))$ and $V(\boldsymbol{\lambda}) = Im(\mathcal{Y}(\boldsymbol{\lambda}))$, where U and V satisfy certain conditions that will be described later.

As we have already seen, if the spatial process Z is defined over a continuum (Z takes values on any location $\mathbf{s} \in D$, that is, geostatistical context), the spectrum lies on $\boldsymbol{\lambda} \in \mathbb{R}^2$. For a discrete process (D is a discrete set of points), we can define the spectrum bounded in Π^2. However, in practice we may aim to recover the spectrum of a continuous process from a discrete realization and therefore, despite the frequency band is the whole space $\mathbb{R}^2$, the frequency behaviour we can recover is restricted to $\Pi^2_\Delta = [-\pi/\Delta_1, \pi/\Delta_1] \times [-\pi/\Delta_2, \pi/\Delta_2]$, where Δ_l, $l = 1, 2$, is the spacing between neighbouring coordinates in the corresponding direction. This effect is known as aliasing and it has been discussed in Section 1.3.5. The aliased spectral density has been defined in (1.66).

It is important to note that in the discrete case the aliasing problem does not arise ($f_\Delta(\boldsymbol{\lambda}) \equiv f(\boldsymbol{\lambda})$). Spectral simulation techniques, as well as most part of the spectral theory for spatial processes, are based on generalizations of spectral procedures for time series. Therefore, the extension of one-dimensional algorithms must be made carefully, regarding the possible continuous character of the spatial process.

In order to capture the continuous character of the spatial process, consider Z observed at locations on a regular grid:

$$D = \{0, \ldots, \Delta_1(n_1 - 1)\} \times \{0, \ldots, \Delta_2(n_2 - 1)\}$$

and denote by $N = n_1 n_2$ the number of observations. The periodogram (2.2) is usually computed at the set of Fourier frequencies $\boldsymbol{\lambda}_{\mathbf{k}} = (\lambda_{k_1}, \lambda_{k_2})^T$:

$$\lambda_{k_l} = \frac{2\pi k_l}{\Delta_l n_l}; \quad k_l = 0, \pm 1, \ldots, \pm[(n_l - 1)/2], \quad l = 1, 2. \tag{3.1}$$

For simplicity, write the periodogram in terms of the sample covariances as follows:

$$I(\boldsymbol{\lambda}_{\mathbf{k}}) = \frac{\Delta_1 \Delta_2}{(2\pi)^2} \sum_{j_1=1-n_1}^{n_1-1} \sum_{j_2=1-n_2}^{n_2-1} \hat{C}(\mathbf{u_j}) e^{-i\boldsymbol{\lambda}_{\mathbf{k}}^T \mathbf{u_j}}, \tag{3.2}$$

where

$$\hat{C}(\mathbf{u_j}) = \frac{1}{N} \sum_{k_1=0}^{n_1-|j_1|} \sum_{k_2=0}^{n_2-|j_2|} Z(\mathbf{s_k}) Z(\mathbf{s_{k+|j|}}), \tag{3.3}$$

$$|\mathbf{j}| = (|j_1|, |j_2|), \quad \mathbf{u}_{\mathbf{j}}^T = (u_{j_1}, u_{j_2}), \quad u_{j_l} = \Delta_l j_l; \quad j_l = 1 - n_l, \ldots, n_l - 1,$$

and $l = 1, 2$. In practice, the periodogram is usually computed from equation (3.2), using an FFT algorithm and with corresponding frequencies given in (3.1). Nevertheless, from this frequency set it is not possible to recover the complete set of sample covariances $\{\hat{C}(\mathbf{u_j}) : j_l = 0, \ldots, n_l - 1, \quad l = 1, 2\}$ (see e.g. Priestley (1981), pp. 577-579, for more details on the one-dimensional case). Therefore, it may be preferable to compute the periodogram at a larger set of frequencies, given by $\boldsymbol{\lambda}_{\mathbf{k}}^T = (\lambda_{k_1}, \lambda_{k_2})$:

$$\lambda_{k_l} = \frac{2\pi k_l}{\Delta_l (2n_l - 1)}; \quad k_l = 0, \pm 1, \ldots, \pm(n_l - 1), \quad l = 1, 2. \tag{3.4}$$

In order to use an FFT algorithm, it would be necessary to obtain a $(2n_1 - 1) \times (2n_2 - 1)$ dataset by zero padding. One could find in the literature different expressions for the Fourier frequency set. With representation (3.4), the Fourier frequencies are symmetric in Π_{Δ}^2 and the boundary is never reached (avoiding some complications).

Any stationary random field admits the Fourier-Stieltjes representation (1.62), as we have already commented, and this fact is the key point in spectral simulation. This integral can be approximated by a discrete transformation. Considering a regular grid with $\{0, \ldots, m_1 - 1\} \times \{0, \ldots, m_2 - 1\}$ observations (for simplicity, assume that m_1 and m_2 are odd), we can define:

$$J(\boldsymbol{\lambda}_{\mathbf{k}}) = \frac{1}{M} \sum_{j_1=0}^{m_1-1} \sum_{j_2=0}^{m_2-1} Z(\mathbf{s_j}) e^{-i\boldsymbol{\lambda}_{\mathbf{k}}^T \mathbf{s_j}}, \tag{3.5}$$

where $M = m_1 m_2$ y $\lambda_{k_l} = \frac{2\pi k_l}{\Delta_l m_l}, k_l = 0, \ldots, m_l - 1, l = 1, 2$. The observations of the process in the grid points can be recovered by an Inverse Fourier Transform:

$$Z(\mathbf{s_j}) = \sum_{k_1=0}^{m_1-1} \sum_{k_2=0}^{m_2-1} J(\boldsymbol{\lambda_k}) e^{i \boldsymbol{\lambda}_\mathbf{k}^T \mathbf{s_j}}, \tag{3.6}$$

where $J(\boldsymbol{\lambda_k})$ are complex random variables:

$$J(\boldsymbol{\lambda_k}) = U(\boldsymbol{\lambda_k}) + iV(\boldsymbol{\lambda_k}),$$

such that $J(\boldsymbol{\lambda_{m-k}}) = J(\boldsymbol{\lambda_{-k}}) = J(\boldsymbol{\lambda_k})^c$, or equivalently, its real and imaginary parts verify:

$$U(\boldsymbol{\lambda_{m-k}}) = U(\boldsymbol{\lambda_{-k}}) = U(\boldsymbol{\lambda_k}),$$
$$V(\boldsymbol{\lambda_{m-k}}) = V(\boldsymbol{\lambda_{-k}}) = -V(\boldsymbol{\lambda_k}).$$

Asymptotic properties for U and V have been studied in Brilinger (1974) (as an extension of Theorem 4.4.2 in Brillinger (1981)), for the particular case of $\Delta_1 = \Delta_2 = 1$. Under the assumption that well separated values of the process are weakly dependent (a kind of mixing condition), it can be proved that asymptotically :

(i) $U(\boldsymbol{\lambda_k})$ and $V(\boldsymbol{\lambda_k})$ are independent.

(ii) $U(\boldsymbol{\lambda_k})$ and $U(\boldsymbol{\lambda_j})$ are independent, for $\mathbf{k} \neq \pm\mathbf{j}$. This assertion also holds for V.

(iii) $E(U(\boldsymbol{\lambda_k})) = E(V(\boldsymbol{\lambda_k})) = 0$, for $\boldsymbol{\lambda_k} \neq \mathbf{0}$ and $E(U(\mathbf{0})) = E(Z(\mathbf{s}))$ (note that $V(\mathbf{0}) = 0$).

(iv) $Var\left(\sqrt{\frac{M\Delta_1\Delta_2}{(2\pi)^2}} U(\boldsymbol{\lambda_k})\right) = \frac{f_\Delta(\boldsymbol{\lambda_k})}{2}$, for $\boldsymbol{\lambda_k} \neq \mathbf{0}$. This assertion also holds for V. Besides, for the origin, $Var\left(\sqrt{\frac{M\Delta_1\Delta_2}{(2\pi)^2}} U(\mathbf{0})\right) = f_\Delta(\mathbf{0})$. In terms of the discrete approximation (3.5), it implies that:

$$Var\left(\sqrt{\frac{M\Delta_1\Delta_2}{(2\pi)^2}}\, |J(\boldsymbol{\lambda_k})|\right) = f_\Delta(\boldsymbol{\lambda_k}).$$

(v) $U(\boldsymbol{\lambda_k})$ and $V(\boldsymbol{\lambda_k})$ are asymptotically Gaussian distributed.

Taking into account properties (i)-(v), it is possible to generate $Z(\mathbf{s}_j)$ values from equation (3.6), by simulation $U(\boldsymbol{\lambda_k})$ and $V(\boldsymbol{\lambda_k})$ variables from the (asymptotic) normal distribution. In this case, the variance could be approximated by:

$$\sigma^2_\mathbf{k} = Var\left(|J(\boldsymbol{\lambda_k})|\right) \approx \frac{(2\pi)^2}{M\Delta_1\Delta_2} f_\Delta(\boldsymbol{\lambda_k}). \tag{3.7}$$

From another point of view, we could consider (3.6) as the mechanism which generates the process. Therefore, we would have a circular process:

$$Z(\mathbf{s_{m-j}}) = Z(\mathbf{s_{-j}}); Z(\mathbf{s_j}) = Z(\mathbf{s_{m+j}}),$$

and, assuming that this process is also stationary, its covariogram satisfies:

$$C^*(\mathbf{u_{m-j}}) = C^*(\mathbf{u_{-j}}) = C^*(\mathbf{u_j}). \tag{3.8}$$

In this situation, it is easy to see (for instance, in Priestley (1981), pp. 258-261, for the one-dimensional case) that:

$$\sigma^2_{\mathbf{k}} = Var\left(|J(\boldsymbol{\lambda}_{\mathbf{k}})|\right) = \frac{1}{M}\sum_{j_1=0}^{m_1-1}\sum_{j_2=0}^{m_2-1} C^*(\mathbf{u_j})e^{-i\boldsymbol{\lambda}_{\mathbf{k}}^T\mathbf{u_j}}. \tag{3.9}$$

Note that, asymptotically, $C^*(\mathbf{u_j}) = C(\mathbf{u_j})$. Most spectral simulation algorithms are based on this result, aproximating $\sigma^2_{\mathbf{k}}$ by the Discrete Fourier Transform of the covariances (symmetrized in such a way that (3.8) holds). It may be also taken into account that the covariances of the original process may not be valid for a circular process. This fact may result in negative approximations of the variances $\sigma^2_{\mathbf{k}}$. In practice, negative estimations are normally set to zero, although better results may be expected when considering (3.7). Further comments on this problem are given at the end of this section.

In any of the spectral simulation methods based on (3.6), since the covariances verify (3.8), if we want to obtain a sample on a $n_1 \times n_2$ grid that reproduces a certain covariance structure, data must be generated on a $m_1 \times m_2$ grid with $m_l \geq 2n_l - 1, \quad l = 1, 2$. For simplicity, we consider $m_l = 2n_l - 1$, for $l = 1, 2$, although m_l may be preferably fixed to larger values (more details will be given at the end of the section).

3.1.2 Parametric Model.

As we pointed out in the introduction, sometimes one does not need to reproduce a complete set of data, but only its signal. Therefore, in some cases it would be enough with using a resampling technique for reproducing periodogram (or log-periodogram) values. Assume that the set of observations can be represented in the following way:

$$Z(\mathbf{s_j}) = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} a_{kl}\varepsilon(s_1 - \Delta_1 k, s_2 - \Delta_2 l), \tag{3.10}$$

where the innovation variables ε come from a white noise process and the $\{a_{kl}\}$ are an L^2-summable sequence. Then (as an extension of Theorem 10.3.1 in Brockwell and Davis (1991), pp. 346-347), the periodogram can be written as:

$$I(\boldsymbol{\lambda}_{\mathbf{k}}) = f_{\Delta}(\boldsymbol{\lambda}_{\mathbf{k}})W_{\mathbf{k}} + R_N^{\Delta}(\boldsymbol{\lambda}_{\mathbf{k}}) \tag{3.11}$$

where the $W_{\mathbf{k}}$'s are independent identically distributed random variables with standard exponetial distribution and $R_N^{\Delta}(\boldsymbol{\lambda}_{\mathbf{k}})$ is a residual term. The idea of a Bootstrap technique for resampling the periodogram in time series context (Franke and Härdle (1992)) comes from model (3.11). Ignoring the residual term $R_N^{\Delta}(\boldsymbol{\lambda}_{\mathbf{k}})$ leads to representing the periodogram as the response in a multiplicative regression model. Applying logarithms in (3.11), we have

$$Y_{\mathbf{k}} = m_{\Delta}(\boldsymbol{\lambda}_{\mathbf{k}}) + z_{\mathbf{k}} + r_{\mathbf{k}}^{\Delta} \tag{3.12}$$

where $m = \log f_{\Delta}$ denotes the log-spectral density, $Y_{\mathbf{k}}$ is the log-periodogram value at frequency $\boldsymbol{\lambda}_{\mathbf{k}}$ and

$$r_{\mathbf{k}}^{\Delta} = \log\left[1 + \frac{R_N^{\Delta}(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\Delta}(\boldsymbol{\lambda}_{\mathbf{k}})W_{\mathbf{k}}}\right]. \tag{3.13}$$

The variables $z_{\mathbf{k}}$ are independently and identically distributed with $Gumbel(0,1)$ distribution. The expected value for this variables is the Euler constant $E(z_{\mathbf{k}}) = -0.57721$ and the variance is $Var(z_{\mathbf{k}}) = \pi^2/6$.

Fan and Zhang (2004) propose a Bootstrap method for resampling log-periodogram values, based on model (3.12), in the discrete time series setting. The simulated log-periodogram values at the Fourier frequencies $\boldsymbol{\lambda}_{\mathbf{k}}$ are obtained as:

$$Y_{\mathbf{k}}^{*} = m_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}}) + z_{\mathbf{k}}^{*}, \tag{3.14}$$

where $m_{\hat{\theta}}$ is a parametric estimator of the log-spectral density and $z_{\mathbf{k}}^{*}$ are independent random realizations of a $Gumbel(0,1)$ distribution. This parametric estimator of the log-spectral density is obtained by maximizing the log-likelihood function associated with (3.12) when ignoring the residual term $r_{\mathbf{k}}^{\Delta}$. Proceeding in such a way, a source of variability in the periodogram scale is removed, given by $R_N^{\Delta}(\boldsymbol{\lambda}_{\mathbf{k}})$, and part of the uncertainty given by the $W_{\mathbf{k}}$ variables. In fact, the parametric estimator $\hat{\theta}$ is the Whittle estimator (see Section 2.5). Apart from the estimation problem, even when considering the theoretical spectral density, the results obtained with simulation methods based on (3.12), ignoring $r_{\mathbf{k}}^{\Delta}$, may not be satisfactory in the multidimensional case. Similar results are obtained from simulation methods based on (3.11) when ignoring $R_N^{\Delta}(\boldsymbol{\lambda}_{\mathbf{k}})$.

When introducing this straightforward extension of a time series simulation method, we are pursuing a double objective: first, we want to emphasize that the naive extension of one-dimensional techniques to the spatial context may not lead to satisfactory results. Secondly, we want to remark that ignoring this residual part $r_{\mathbf{k}}^{\Delta}$ causes a serious loss of variability in the estimation of the log-spectral density.

3.2 Fourier simulation methods.

In this section we will recall the Fourier Integral Method, proposed in Pardo-Igúzquiza and Chica-Olmo (1993) and comment some of its drawbacks. We will propose an extension of this method: the Modified Fourier Integral Method.

3.2.1 The Fourier Integral Method.

A spectral simulation algorithm, called the Fourier Integral Method (FIM), has been proposed for the simulation of stationary processes with a certain dependence structure. Originally introduced in Borgman *et al.* (1984), this algorithm was extended to higher dimensions in Pardo-Igúzquiza and Chica-Olmo (1993). Yao (1998) adapts this method for conditional simulation. Given a certain covariance structure (or a variogram model), the algorithm proposed by these authors is as follows:

1. Use the variogram or covariogram model to compute discrete covariances $C(\mathbf{u_j})$, for $j_1 = 0, \ldots, n_1 - 1$ and $j_2 = 0, \ldots, n_2 - 1$.
2. Compute the discrete Fourier transform of $\{C^*(\mathbf{u_j})\}$, defined by $C^*(\mathbf{u_j}) = C(\mathbf{u_j})$ if $j_l \leq n_l$ and $C^*(\mathbf{u_{m-j}}) = C(\mathbf{u_j})$ otherwise, and obtain the discrete density spectrum (3.9). If negative values are obtained, these values are often set to zero.
3. Draw random phases $\phi(\boldsymbol{\lambda}_\mathbf{k})$, from a uniform distribution in $[0, 2\pi]$. To obtain real values, phases must be symmetric: $\phi(\boldsymbol{\lambda}_\mathbf{k}) = -\phi(\boldsymbol{\lambda}_{-\mathbf{k}})$
4. Build the Fourier coefficients as $J(\boldsymbol{\lambda}_\mathbf{k}) = \sqrt{\sigma^2_\mathbf{k}} e^{-i\phi(\boldsymbol{\lambda}_\mathbf{k})}$, for $\mathbf{k} \neq \mathbf{0}$ and $J(\boldsymbol{\lambda}_\mathbf{0}) = \sqrt{2\sigma^2_\mathbf{0}} \cos(\boldsymbol{\lambda}_\mathbf{0})$.
5. Perform the Fast Fourier Transform (3.6) to get the simulated $Z(\mathbf{s_j})$ values.
6. Take a subgrid of $(n_1 \times n_2)$ observations (and compute the periodogram for these data, if that is the case).

Notice that, with this algorithm, the only source of variability in the simulated dependence structure comes from Step 6. For example, if one computes the periodogram with the complete set of data, no variations in the periodogram values will be found.

3.2.2 The Modified Fourier Integral Method.

We revise the FIM considering an additional source of variability in the frequency domain. We introduce in the amplitudes of the Fourier coefficients an exponential variable, as it is suggested

in the representation of the periodogram (3.11). The Modified Fourier Integral Method (MFIM) is as follows:

1. Compute the approximation of the spectral variances $\sigma^2_{\mathbf{k}}$. This could be done by different ways:
   (a) Proceed as in Steps 1 and 2 from FIM algorithm.
   (b) Use the asymptotic approximation (3.7).
   (c) Combine (a) and (b) (e.g. use (3.9) and if negative values are obtained, replace them by (3.7)).
2. Draw random phases $\phi(\boldsymbol{\lambda}_{\mathbf{k}})$, from a uniform distribution in $[0, 2\pi]$. To obtain real values, phases must be symmetric: $\phi(\boldsymbol{\lambda}_{\mathbf{k}}) = -\phi(\boldsymbol{\lambda}_{-\mathbf{k}})$
3. Build the Fourier coefficients as $J(\boldsymbol{\lambda}_{\mathbf{k}}) = \sqrt{\sigma^2_{\mathbf{k}} W_{\mathbf{k}}} e^{-i\phi(\boldsymbol{\lambda}_{\mathbf{k}})}$, for $\mathbf{k} \neq \mathbf{0}$ and $J(\boldsymbol{\lambda}_{\mathbf{0}}) = \sqrt{2\sigma^2_{\mathbf{0}} W_{\mathbf{0}}} \cos(\boldsymbol{\lambda}_{\mathbf{0}})$, where the variables $\{W_{\mathbf{k}}\}$ are independent and randomly sampled from a standard exponential distribution $W_{\mathbf{k}} \sim Exp(1)$.
4. Perform the Fast Fourier Transform (3.6) to get the simulated $Z(\mathbf{s_j})$ values.
5. Take a subgrid of $(n_1 \times n_2)$ observations (and compute the periodogram for these data, if that is the case).

By the following theorem, it is easy to show that the realizations of the spectral process $J(\boldsymbol{\lambda}_{\mathbf{k}})$ drawn fron the MFIM method verify the asymptotic conditions of independence, normal distribution, zero mean and variance given in (iv) on the real and imaginary parts.

Theorem. *(Box and Muller (1958)). Let A_1 and A_2 be independent random variables, $U(0,1)$ distributed. Consider the random variables:*

$$B_1 = (-2\log A_1)^{1/2}\cos(2\pi A_2), \quad B_2 = (-2\log A_1)^{1/2}\sin(2\pi A_2).$$

Then, B_1 and B_2 are independent random variables, $N(0,1)$ distributed.

Since $A_1 \sim U(0,1)$, the transformed variable $(-\log A_1)$ follows a standard exponential distribution, $Exp(1)$, which coincides with the distribution of the $W_{\mathbf{k}}$ variables involved in MFIM method. Taking random amplitudes $\sqrt{\sigma^2_{\mathbf{k}} W_{\mathbf{k}}}$, gives zero mean Gaussian variables with variance $\sigma^2_{\mathbf{k}}/2$. The computational efficiency in the generation of the Fourier coefficients can be improved, avoiding the computation of sines and cosines, by considering a similar approximation to that

given by Ross (1997), pp. 74-75.

This method provides realizations of a Gaussian process, which could be done by directly simulating Gaussian variables. For instance, in Chilès and Delfiner (1999) (pp.496-498) the algorithm for simulating a unidimensional spatial process (based on the approximation (3.9)) is thoroughly described. It is important to note that this algorithm is based on the approximation of a stationary circular process, which is equivalent to the circular embedding method proposed by Dietrich and Newsam (1997). The advantage of considering an algorithm based on the Box-Muller representation, makes easier the extension of this method to non-Gaussian cases (see Cressie (1993), p.205).

From Theorem 1.5.5 in Muirhead (1982), it can be seen that if J is spherically distributed, then the random phases are uniformly distributed on Π^2 and the distribution of J is characterized by the distribution of the amplitudes in the following way:

$$f_{|J|^2}(y) = Cy^{-1/2}h(y) \quad \text{and} \quad f_J(z) = Ch(z^2), \tag{3.15}$$

where f_J denotes the univariate density of the real and imaginary parts of $J(\boldsymbol{\lambda}_{\mathbf{k}})$. We could consider the generation of scaled Student's t random variables with $p > 2$ degrees of freedom, which corresponds with uniformly distributed random phases on Π^2 and squared random amplitudes $W^2_{\mathbf{k}}$ with density:

$$f_{|J|^2}(y) = \frac{\sqrt{p-2}\Gamma\left(\frac{p+1}{2}\right)}{p\Gamma\left(\frac{p}{2}\right)\sqrt{\pi}} \frac{y^{-1/2}}{\left(1+\frac{p-2}{p}y\right)^{\frac{p+1}{2}}}, \quad y > 0. \tag{3.16}$$

Both for the discrete and continuous cases, but more noticeable in this last situation, truncation errors from the computation in Step 2 may not be negligible. The discrete density spectrum is obtained from the Discrete Fourier Transform of $m_1 \times m_2$ covariances, and not from the whole set of covariances, which leads to a non finite sum. In the geostatistical context, if the range is large relative to the number of covariances, the discrete spectrum will be a poor approximation of the spectral density.

As we have already noticed, in order to simplify the description of the algorithms, we take $m_l = 2n_l - 1$, $l = 1, 2$, although m_l may be better fixed to other values. The discrete density spectrum (3.9) does not take into account the covariances for all possible lags. Thus, negative estimates for the spectral variances $\sigma^2_{\mathbf{k}}$ may be obtained. This may happen when the range of the spatial dependence is large, compared with the simulation grid size. If the covariogram has a finite range r, this truncation problem can be avoided by choosing $(m_l - 1)\Delta_l \geq 2r$. In case the covariogram has a non-finite range, the truncation problem persists no matter now large m_l are taken. In this case, it may be better to select $(m_l - 1)\Delta_l \geq 2r^*$, where r^* denotes the practical

range (see also Chilès and Delfiner (1999), pp.500-501 for different approaches).

From a computational point of view, it may be interesting to modify the m_l values, in order to take advantage of the FFT algorithm we chose, as it has been pointed out in Section 1.3.2. We could chose m_1 and m_2 as products of small prime factors. If m_l, $l = 1, 2$, satisfy this condition, then the computational effort required is proportional to $M \log(M)$. Therefore, m_l values should be approximated to the higher closest k-smooth number. Other authors consider $m_l = 2n_l$, $l = 1, 2$, which may be a good option if n_l are smooth numbers.

One must be careful in the construction of the Fourier coefficients. With an odd number of Fourier frequencies, the Fourier coefficient at the origin is real and in all the other frequencies we have complex coefficients. If m_l is even, then the frequency $\pm\frac{\pi}{\Delta_l}$ is reached. The algorithms described above should be adapted in a suitable way to these situations. The Fourier coefficients corresponding to frequencies with both components multiples of $\frac{\pi}{\Delta_l}$ must be handled in the same way as the origin, with real coefficients.

3.2.3 Aliasing correction.

Another problem of the spectral simulation method is related to the aliasing phenomena, which appears when the spectral density of a continuous spatial process presents significative side lobes outside Π^2_Δ.

In practice, in order to avoid inconvenients derived from the aliasing problem, the spacing in the simulation grid may be reduced, and consider $\Delta_l^* = \Delta_l/p_l$, $n_l^* = p_l n_l$, with integer $p_l > 1$. Proceeding in this way, the last subsampling step in the algorithms should be modified and one from each p_l simulated values should be taken in l dimension.

The aliasing problem has already been commented in Section 1.3.5 and, as we have already seen, the FIM method does not account for this feature when trying to recover the signal of a continuous spatial (geostatistical) processes from a discrete set of observations. In the Modified Fourier Integral Method, for the continuous process situation, we compute $\sigma^2_{\mathbf{k}}$ depending on Δ_1 and Δ_2, which can be seen as the aliased dicrete density spectrum.

Nevertheless, this approach may not be sufficient if there is a significant spectral mass beyond the frequency band Π^2_Δ. There are different options for reducing the aliased effect. Some modifi-

cations for aliasing reduction involve the specification of a functional form for the spectral density $f(\|\boldsymbol{\lambda}\|)$ out of the frequency band Π^2, in order to determine the aliasing term. The aliased version can be approximated, at each frequency, by truncation. That is:

$$f(\boldsymbol{\lambda}) \approx f_{\Delta}(\boldsymbol{\lambda}) - \sum_{m_1=-\tau_1}^{\tau_1} \sum_{m_2=-\tau_2}^{\tau_2} f(\|\boldsymbol{\lambda} + 2\pi\mathbf{m}\|).$$

The selection of the truncation terms τ_1 and τ_2 depends on the characteristics of the density, as well as on the size of the sample grid.

A more formal approach, in the time series setting, is given by Robinson (1976). In this work, the author investigates the biases that may occur in the estimation of the spectral density when some spectral mass lies beyond the bounded frequency band $[-\pi, \pi]$. Some modifications to reduce the biases are also proposed. Under isotropy, we can extend in a natural way this technique to reduce the effect of aliasing in a spatial spectrum. The aliasing term,

$$\sum_{m_1 \neq 0} \sum_{m_2 \neq 0} f(\|\boldsymbol{\lambda} + 2\pi\mathbf{m}\|) \tag{3.17}$$

can be approximated when the theoretical spectral density shows a certain behaviour for high frequencies. In particular, as an extension of (Robinson (1976)), the following cases are considered:

$$f(\|\boldsymbol{\lambda}\|) = C\|\boldsymbol{\lambda}\|^{-n}, \quad |\boldsymbol{\lambda}_1| > \pi \text{ or } |\boldsymbol{\lambda}_2| > \pi, \tag{3.18}$$

and when the exponent $n \in \mathbb{Z}$ is large enough, one could proceed to the limiting case

$$f(\|\boldsymbol{\lambda}\|) = Ce^{-\|\boldsymbol{\lambda}\|}, \quad |\boldsymbol{\lambda}_1| > \pi \text{ or } |\boldsymbol{\lambda}_2| > \pi. \tag{3.19}$$

For the Matérn spectral density (continuous spectrum), in Section 3.3, if we consider the modification for correcting aliasing proposed in (3.19), which makes sense for large smoothness parameter ν, the aliased term in the spectral density is bounded by (if f is even):

$$\begin{aligned}
&\sum_{|m_1|\geq 1} \sum_{|m_2|\geq 1} \phi e^{-\|\boldsymbol{\lambda}+2\pi\mathbf{m}\|} = \\
&\sum_{|m_1|\geq 1} \sum_{|m_2|\geq 1} \phi e^{-\sqrt{(\lambda_1+2\pi m_1)^2+(\lambda_2+2\pi m_2)^2}} \leq \\
&\phi \sum_{|m_1|\geq 1} \sum_{|m_2|\geq 1} e^{-(\lambda_1+2\pi m_1)^2} e^{-(\lambda_2+2\pi m_2)^2} = \\
&\phi \sum_{|m_1|\geq 1} e^{-(\lambda_1+2\pi m_1)^2} \sum_{|m_2|\geq 1} e^{-(\lambda_2+2\pi m_2)^2} = \\
&\frac{\phi \cosh\lambda_1 \cosh\lambda_2}{e^2 \sinh^2 1}.
\end{aligned}$$

If f is odd:

$$\sum_{|m_1|\geq 1}\sum_{|m_2|\geq 1}\phi e^{-\|\boldsymbol{\lambda}+2\pi\mathbf{m}\|}=\frac{\phi\sinh\lambda_1\sinh\lambda_2}{e^2\sinh^2 1}.$$

For the particular case of an exponential density ($\nu=1/2$), one could use a less conservative approach, given by (3.18); that is, considering $f(\boldsymbol{\lambda})=\phi\|\boldsymbol{\lambda}\|^{-2}$ outside the frequency band. The aliased term can be then bounded by:

$$\begin{aligned}&\sum_{|m_1|\geq 1}\sum_{|m_2|\geq 1}\phi\|(\lambda_1+2\pi m_1,\lambda_2+2\pi m_2)\|^{-2}=\\&\sum_{|m_1|\geq 1}\sum_{|m_2|\geq 1}\phi\frac{1}{(\lambda_1+2\pi m_1)^2+(\lambda_2+2\pi m_2)^2}\leq\\&\phi\sum_{|m_1|\geq 1}\frac{1}{(\lambda_1+2\pi m_1)^2}\sum_{|m_2|\geq 1}\frac{1}{(\lambda_2+2\pi m_2)^2}=\\&4\phi V_\nu\left(\frac{\boldsymbol{\lambda}_1}{\pi}\right)V_\nu\left(\frac{\boldsymbol{\lambda}_2}{\pi}\right),\end{aligned}$$

where

$$V_\nu(x)=\frac{\pi}{2}\left(\sin^{-2}\left(\frac{\pi x}{2}\right)+x^{-2}\right).$$

The bounds for the aliased term proposed above have a limited use in practice, when a dataset is given and the covariance structure is unknown. For a generation of a spatial process with a certain spatial-spectral density, the bounds above may help to calibrate the influence of the alised part. Nevertheless, we must say that, in the cases we have studied, the aliasing effect involves a lower error than that one caused by truncation in Step 2, both for FIM and MFIM.

3.3 Simulation Results.

In the lattice context, we consider the doubly-geometric process (see Martin (1979)), namely $BAR(1)$, with spectral density given by (2.92). 10000 generations of the process are drawn in a 20×20 regular grid. We compare the results obtained with those produced by methods for generating linear-by-linear processes (see Alonso *et al.* (1996)), the FIM, the MFIM and an extension of a simulation algorithm from time series.

Extending time series simulation techniques, we generate log-periodogram values from equation (3.14), with m_θ the theoretical log-spectral density, following Fan and Zhang (2004). We will call this procedure Log-Periodogram Simulation method (LPS). This procedure will be equivalent to generate periodogram values from representation (3.11), ingoring $R_N^{\Delta}(\boldsymbol{\lambda}_\mathbf{k})$.

As a represenative of geostatistical processes, we consider Gaussian spatial process with Matérn spectral density (see Stein (1999)). For certain smoothness and range parameters, the covariances of this model are not valid for a circular process. Therefore, we may obtain negative approximations for the spectral variances $\sigma^2_{\mathbf{k}}$. For the MFIM algorithm, we observe better results when considering option (c) in Step 1, although option (a) provide quite similar results. In this case, we take Cholesky's factorization method as reference (benchmark).

Mean Error, Mean Square Error and Whittle Error surfaces are computed in order to compare the performance of the periodogram as an estimator of the spectral density, when data are provided by different simulation procedures.

For B simulated samples, Mean, Mean Square and Whittle Errors of the log-periodogram as an estimate for the log-spectral density are given by (3.20), (3.21) and (3.22), respectively.

$$ME(m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}), Y_{\mathbf{k}}) = \tfrac{1}{B}\sum_{b=1}^{B}(m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - Y^b_{\mathbf{k}}); \tag{3.20}$$

$$MSE(m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}), Y_{\mathbf{k}}) = \tfrac{1}{B}\sum_{b=1}^{B}(m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - Y^b_{\mathbf{k}})^2, \tag{3.21}$$

$$WE(m_\theta, Y_{\mathbf{k}}) = \tfrac{1}{B}\sum_{b=1}^{B}(Y^b_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y^b_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}). \tag{3.22}$$

These three surfaces are compared when data are generated by traditional methods, FIM and MFIM. Besides, we also compare the results when log-periodogram values are obtained by LPS.

3.3.1 Bidimensional autoregresive process.

The $BAR(1)$ process has already been introduced in Section 2.6. It is the simplest case of the linear-by-linear processes introduced by Martin (1979). For the $BAR(1)$ model, realizations can be obtained with the following formula:

$$Z(i,j) = \beta_1 Z(i-1,j) + \beta_2 Z(i,j-1) - \beta_1\beta_2 Z(i-1,j-1) + \varepsilon(i,j) \tag{3.23}$$

where ε are independent, identically distributed Gaussian random variables, with zero-mean and variance σ^2. Parameters β_1 and β_2 must be in $[0,1)$ to guarantee stationarity. The log-spectral density for an $BAR(1)$ process (see equation (2.92)), with autoregression parameters $\beta_1 = \beta_2 = 0.5$ is shown in Figure 3.1.

A method for generating this process is proposed by Alonso *et al.* (1996). For $i = 2, \ldots, n_1$ and $j = 2, \ldots, n_2$, a realization from a bidimensional autoregresive process, in a regular grid $D = \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$ can be obtained by computing:

$$\begin{aligned} Z(i,j) &= \beta_1 Z(i-1,j) + \beta_2 Z(i,j-1) - \beta_1\beta_2 Z(i-1,j-1) + \varepsilon(i,j) \\ Z(i,1) &= \beta_1 Z(i-1,1) + \varepsilon(i,1) \\ Z(1,j) &= \beta_2 Z(1,j-1) + \varepsilon(1,j) \\ Z(1,1) &= \varepsilon(1,1) \end{aligned}$$

where

$$\begin{aligned} \varepsilon(i,j) &\sim N(0, \sigma^2) \\ \varepsilon(i,1) &\sim N(0, (1-\beta_1^2)^{-1}\sigma^2) \\ \varepsilon(1,j) &\sim N(0, (1-\beta_2^2)^{-1}\sigma^2) \\ \varepsilon(1,1) &\sim N(0, (1-\beta_1^2)^{-1}(1-\beta_2^2)^{-1}\sigma^2) \end{aligned}$$

and all the inputs are assumed to be jointly independent.

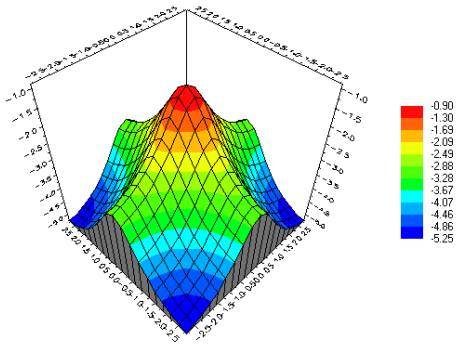

Figure 3.1: Log-spectral density for a $BAR(1)$ process, given by equation (2.92), with autoregression parameters $\beta_1 = \beta_2 = 0.5$.

Table 3.1 shows summary statistics (mean, median and standard deviation) for a 20×20 and a 50×50 regular grid simulation. These statistics are obtained from 10000 simulations. Results from LPS simulations are not affected by the sample size. The mean and median oscilate around the Euler constant c_0 and the Mean Square Error is about $(\pi^2/6 + c_0^2)$ (see remark at the end of the section for further explanation). Besides, the LPS method exhibits lower variation than the other methods. The MFIM shows a better performance than the FIM.

Table 3.1: Summary statistics.

Mean Error		Linear	MFIM	FIM	LPS
	Mean	-0.5100	-0.5120	-0.4823	-0.5762
20×20	Median	-0.5107	-0.5109	-0.4816	-0.5748
	St.dev.	0.0964	0.0967	0.0898	0.0682
	Mean	-0.5457	-0.5460	-0.5167	-0.5774
50×50	Median	-0.5458	-0.5461	-0.5165	-0.5775
	St.dev.	0.0369	0.0368	0.0345	0.0261
Mean Square Error					
	Mean	1.8533	1.8578	1.7698	1.9739
20×20	Median	1.8362	1.8414	1.7549	1.9599
	St.dev.	0.2827	0.2804	0.2726	0.2343
	Mean	1.8839	1.8836	1.7955	1.9789
50×50	Median	1.8829	1.8414	1.7931	1.9772
	St.dev.	0.1088	0.1099	0.1056	0.0899
Whittle Error					
	Mean	1.5754	1.5766	1.5474	1.5764
20×20	Median	1.5740	1.5748	1.5456	1.5757
	St.dev.	0.0589	0.0579	0.0554	0.0425
	Mean	1.5723	1.5720	1.5431	1.5774
50×50	Median	1.5719	1.5719	1.5427	1.5773
	St.dev.	0.0222	0.0226	0.0212	0.0164

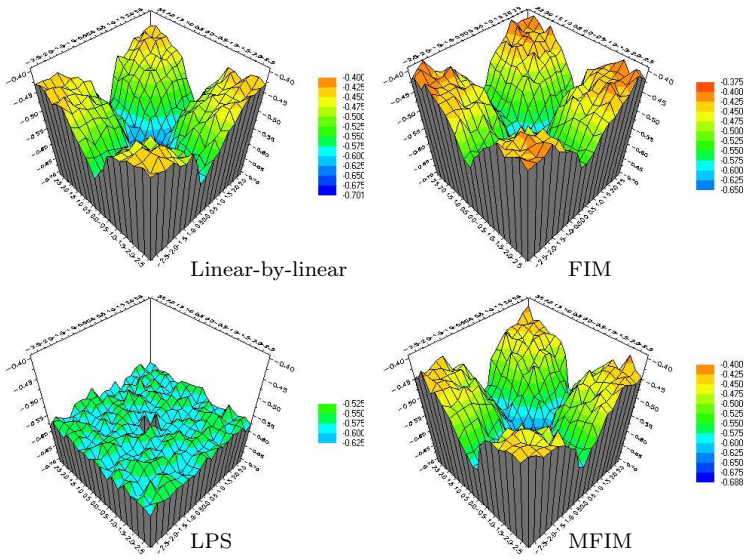

Figure 3.2: $BAR(1)$ process. Mean Error for the estimation of the log-spectral density. Linear-by-linear: simulations by Alonso *et al.* (1996). LPS: simulations by the extension of Fan's parametric Bootstrap. 20×20 grid.

For a 20×20 regular grid simulation, in terms of Mean Error (Figure 3.2), both FIM and the MFIM show a good behaviour, although MFIM is slightly better. The log-periodogram values from LPS do not capture all the variability. In Figure 3.2, the loss of variability in the LPS method is clear. Recall that this method ignores the term $r_{\mathbf{k}}^{\Delta}$ in the representation of the log-periodogram (3.12). This term can be considered *proportional* to the inverse of the spatial spectral density, represented in Figure 3.1. Removing the term $r_{\mathbf{k}}^{\Delta}$ provoques the lost of the lobes that appear in the Mean Error surfaces for the other methods. This behaviour is also shown in Figure 3.3, in Mean Square Error terms. The Mean Square Error surface obtained by LPS simulations shows an almost constant shape.

For Whittle's Error (Figure 3.4), LPS exhibits a good behaviour. This fact is not surprising because the log-periodogram values are computed from a regression model which also provides the log-likelihood (see remark at the end of the section). While results obtained from data generated by the FIM shows that it does not capture all the variability in terms of the log-periodogram, MFIM still shows a good behaviour.

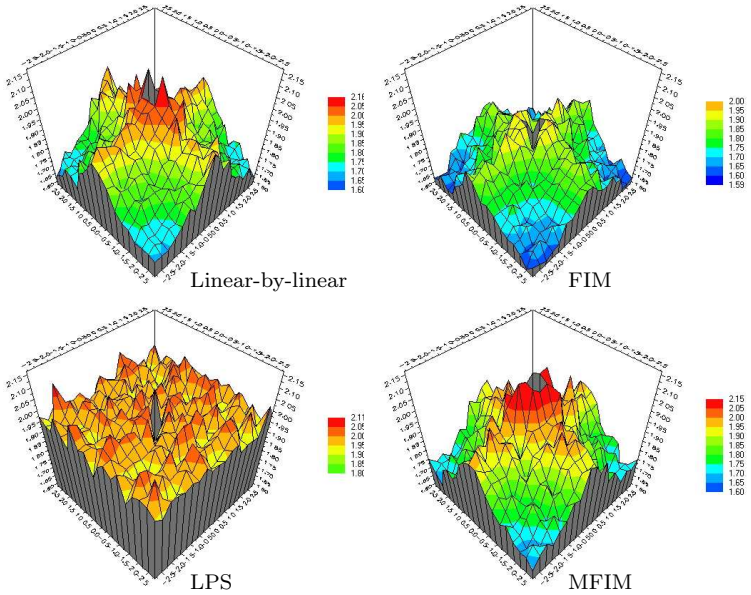

Figure 3.3: $BAR(1)$ process. Mean Square Error for the estimation of the log-spectral density. Linear-by-linear: simulations by Alonso *et al.* (1996). LPS: simulations by the extension of Fan's parametric Bootstrap. 20×20 grid.

It may be surprising the fact that LPS works badly for ME and MSE, but it does pretty well in terms of Whitle's Error. For $Y_{\mathbf{k}}$ generated by LPS method, ME, MSE and WE results are:

$$\frac{1}{B}\sum_{i=1}^{B}(Y_{\mathbf{k}}^i - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})) = \frac{1}{B}\sum_{i=1}^{B} z_{\mathbf{k}}^i \approx E(z_{\mathbf{k}}) = C_0 = -0.57721,$$

$$\frac{1}{B}\sum_{i=1}^{B}(Y_{\mathbf{k}}^i - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}))^2 = \frac{1}{B}\sum_{i=1}^{B}(z_{\mathbf{k}}^i)^2 \approx \frac{\pi^2}{6} + c_0^2 = 1.978,$$

$$\frac{1}{B}\sum_{i=1}^{B}(Y_{\mathbf{k}}^i - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y_{\mathbf{k}}^i - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}) = \frac{1}{B}\sum_{i=1}^{B}(z_{\mathbf{k}}^i - e^{z_{\mathbf{k}}^i}) \approx -1.57721.$$

3.3.2 Matérn spectral density family.

The Matérn class of spectral densities has been introduced in Section 1.3.6. Equations (1.67) and (1.68) correspond to the spatial spectral density and the corresponding covariance function. In this section, we have considered a Gaussian process with spatial spectral density belonging to the Matérn family.

In this context, we will confront two problems: the aliasing and the truncation errors. In Figure 3.5 we can see the error surfaces for the discrete density spectrum as an approximation

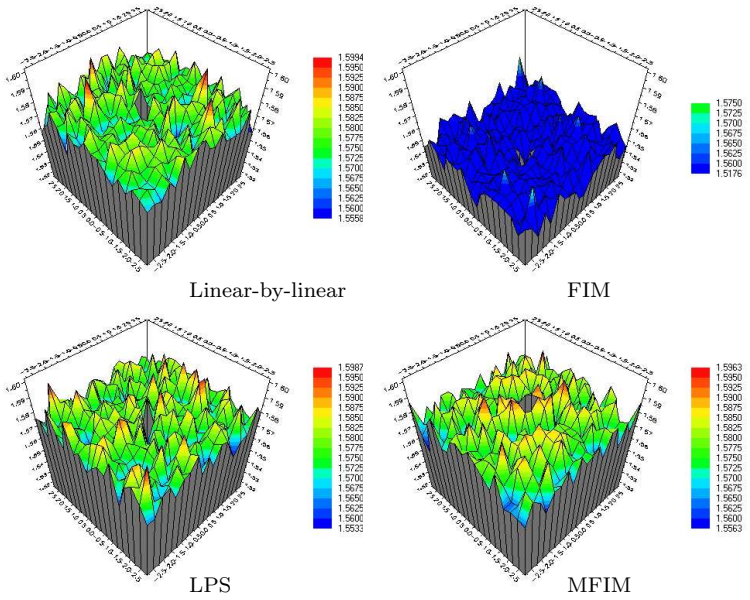

Figure 3.4: $BAR(1)$ process. Whittle's Error for the estimation of the log-spectral density. Linear-by-linear: simulations by Alonso *et al.* (1996). LPS: simulations by the extension of Fan's parametric Bootstrap. 20×20 grid.

to the spatial spectral density, in logarithmic scale. In order to point out that truncation errors become serious for large autocorrelation ranges, we have fixed the smoothness parameter $\nu = 0.5$ and, over a grid of size 41×41, different autocorrelation ranges ($a = \alpha^{-1}$) are explored: $a = 10$, $a = 20$ and $a = 40$. Autocorrelation ranges with values $a = 10$ and $a = 20$ correspond to a 25% and a 50% of the side-length of the grid. This effect is related to the construction of covariances, which are not valid for a circular process (see Section 3.1.2).

In Tables 3.2 to 3.4 we show summary statistics for the Mean, Mean Square and Whittle's errors for the log-periodogram as an estimator of the log-spectral density. Simulations were carried out considering a Matérn model, with smoothness parameter $\nu = 0.5$ and different autocorrelation ranges. In order to make results comparable, we have consider ranges of the 20%, 50% and 80% of the side-length of the grid. MFIM shows slightly better results for a 20×20 regular grid, and its performance improves for 50×50 simulations.

In Table 3.5, we show the results for $\nu = 0.05$ and different autocorrelation ranges, for 20×20 and 50×50 regular grids. In Table 3.6, the same summary statistics are shown for a larger smoothness parameter $\nu = 1.00$.

In Figures 3.6 to 3.8 we show the Mean Error, Mean Square Error and Whittle's Error surfaces for the estimation of the log-spectral density, by Cholesky factorization, FIM, MFIM and LPS methods. In these figures, the smoothness parameter is fixed to $\nu = 0.5$ and the autocorrelation range is large (80% of the side-length of the grid). We have already seen that truncation errors are serious in this situation, around frequencies with one zero component. The most relevant differences are found around the origin. The peaks near frequencies with components $\pm\pi$ appear because of the spectral density (see Figure 3.5).

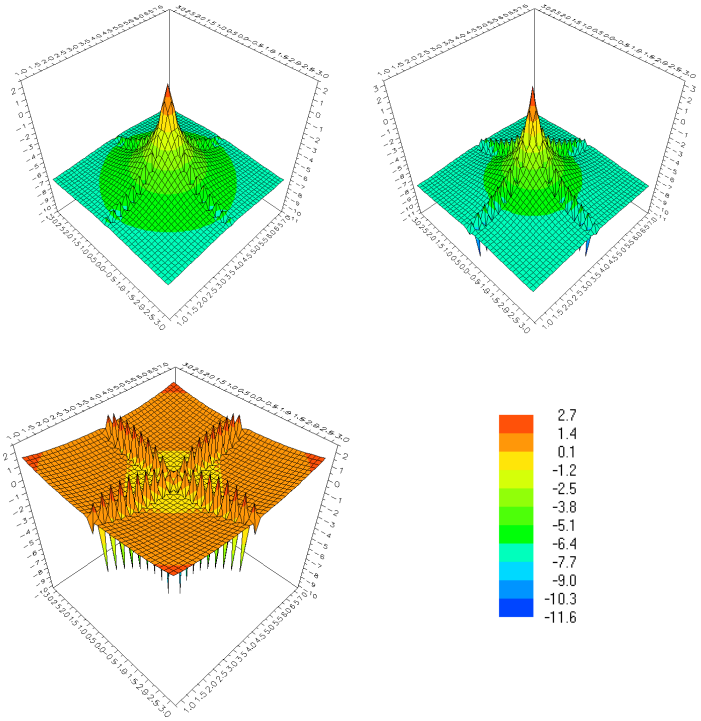

Figure 3.5: Error surfaces for the discrete approximation of the log-spectral density. Smoothness parameter $\nu = 0.5$. Grid size 41×41. From left to right and from top to bottom, autocorrelation range $a = \alpha^{-1}$: $a = 10$, $a = 20$ and $a = 40$.

Just a note about computational time: in a Pentium IV (2.6 Ghz), for the simulation a 50×50 regular grid using Cholesky factorization, it takes 28.19 seconds, approximately. The same simulation using MFIM takes 0.01 seconds.

Table 3.2: Summary statistics. Matern spectral density with parameters $\nu = 0.5$ and $\alpha = 80\%N$.

Mean Error		Cholesky	MFIM	FIM	LPS
20×20	Mean	0.3319	0.3445	0.3783	-0.5773
	Median	0.3226	0.3333	0.3729	-0.5759
	St.dev.	0.1453	0.1417	0.1262	0.0613
50×50	Mean	0.2685	0.2763	0.3052	-0.5771
	Median	0.2589	0.2671	0.2989	-0.5772
	St.dev.	0.0826	0.0784	0.0706	0.0251
Mean Square Error					
20×20	Mean	1.9102	1.9209	1.8983	1.9787
	Median	1.8863	1.8973	1.8846	1.0699
	St.dev.	0.2538	0.2511	0.2281	0.2110
50×50	Mean	1.7870	1.8016	1.7810	1.9785
	Median	1.7681	1.7858	1.7728	1.9767
	St.dev.	0.1516	0.1476	0.1325	0.0871
Whittle Error					
20×20	Mean	2.6621	2.6975	2.6691	1.5776
	Median	2.4958	2.5269	2.5757	1.5771
	St.dev.	0.6159	0.6217	0.4791	0.0384
50×50	Mean	2.6228	2.6649	2.6422	1.5772
	Median	2.4705	2.5169	2.5597	1.5771
	St.dev.	0.5522	0.5419	0.4273	0.0158

Table 3.3: Summary statistics. Matérn spectral density with parameters $\nu = 0.5$ and $\alpha = 20\%N$.

Mean Error		Cholesky	MFIM	FIM	LPS
20×20	Mean	0.1892	0.1904	0.2171	-0.5773
	Median	0.1883	0.1861	0.2166	-0.5759
	St.dev.	0.1106	0.1088	0.1008	0.0613
50×50	Mean	0.1539	0.1529	0.1806	-0.5771
	Median	0.1516	0.1503	0.1797	-0.5772
	St.dev.	0.0549	0.0535	0.0498	0.0251
Mean Square Error					
20×20	Mean	1.8207	1.8125	1.7741	1.9787
	Median	1.8073	1.8022	1.7670	1.9699
	St.dev.	0.2063	0.2010	0.1950	0.2110
50×50	Mean	1.7753	1.7768	1.7354	1.9785
	Median	1.7714	1.7741	1.7321	1.9767
	St.dev.	0.0869	0.0874	0.0827	0.0871
Whittle Error					
20×20	Mean	2.1712	2.1686	2.1426	1.5776
	Median	2.1474	2.1461	2.1283	1.5771
	St.dev.	0.1794	0.1794	0.1541	0.0384
50×50	Mean	2.1353	2.1338	2.1103	1.5772
	Median	2.1142	2.1142	2.0955	1.5771
	St.dev.	0.1160	0.1134	0.0991	0.0158

Table 3.4: Summary statistics. Matérn spectral density with parameters $\nu = 0.5$ and $\alpha = 50\%N$.

Mean Error		Cholesky	MFIM	FIM	LPS
20×20	Mean	0.2708	0.2728	0.3033	-0.5773
	Median	0.2652	0.2644	0.2999	-0.5759
	St.dev.	0.1313	0.1303	0.1178	0.0613
50×50	Mean	0.2330	0.2353	0.2643	-0.5771
	Median	0.2257	0.2281	0.2600	-0.5772
	St.dev.	0.0753	0.0726	0.0659	0.0251
Mean Square Error					
20×20	Mean	1.8706	1.8650	1.8334	1.9787
	Median	1.8561	1.8496	1.8231	1.9699
	St.dev.	0.2241	0.2194	0.2038	0.2110
50×50	Mean	1.7983	1.8028	1.7744	1.9785
	Median	1.7839	1.7902	1.7668	1.9767
	St.dev.	0.1282	0.1151	0.0726	0.0871
Whittle Error					
20×20	Mean	2.3919	2.3903	2.3629	1.5776
	Median	2.3126	2.3037	2.3152	1.5771
	St.dev.	0.3598	0.3616	0.2863	0.0384
50×50	Mean	2.4330	2.4412	2.4184	1.5772
	Median	2.3373	2.3483	2.3639	1.5771
	St.dev.	0.3700	0.3625	0.2906	0.0158

Table 3.5: Summary statistics. Matérn spectral density 20×20 regular grid, $\nu = 0.05$ with different ranges.

Mean Error		Cholesky	MFIM	FIM
80%	Mean	1.6390	1.6400	1.6705
	Median	1.6402	1.6401	1.6722
	St.dev.	0.0880	0.0870	0.0824
20%	Mean	1.6357	1.6367	1.6672
	Median	1.6370	1.6366	1.6687
	St.dev.	0.0871	0.0860	0.0818
50%	Mean	1.6360	1.6369	1.6673
	Median	1.6372	1.6369	1.6690
	St.dev.	0.0876	0.0863	0.0822
Mean Square Error				
80%	Mean	4.8680	4.8657	4.9074
	Median	4.8646	4.8617	4.9051
	St.dev.	0.2759	0.2787	0.2549
20%	Mean	4.8712	4.8701	4.9121
	Median	4.8668	4.8666	4.9085
	St.dev.	0.2757	0.2781	0.2548
50%	Mean	4.8657	4.8640	4.9060
	Median	4.8617	4.8613	4.9027
	St.dev.	0.2753	0.2546	0.0863
Whittle Error				
80%	Mean	9.8501	9.8458	9.88090
	Median	9.8136	9.8125	9.7882
	St.dev.	0.8455	0.8476	0.7368
20%	Mean	9.8572	9.8531	9.8159
	Median	9.8200	9.8245	9.7977
	St.dev.	0.8455	0.8473	0.7361
50%	Mean	9.8424	9.8382	9.8015
	Median	9.8066	9.8068	9.7838
	St.dev.	0.8442	0.8465	0.7358

Table 3.6: Summary statistics. Matérn spectral density: 20×20 regular grid, $\nu = 1.00$ with different ranges.

Mean Error		Cholesky	MFIM	FIM
80%	Mean	1.2255	1.3405	1.3742
	Median	1.1491	1.2604	1.3023
	St.dev.	0.3453	0.3149	0.2878
20%	Mean	0.3892	0.3918	0.4210
	Median	0.3635	0.3638	0.4025
	St.dev.	0.2138	0.2152	0.1930
50%	Mean	0.7654	0.8183	0.8537
	Median	0.6973	0.7506	0.7939
	St.dev.	0.3245	0.3117	0.2833
Mean Square Error				
80%	Mean	3.4625	3.8331	3.9306
	Median	3.2863	3.6326	3.7780
	St.dev.	1.1230	1.0634	0.9264
20%	Mean	2.0643	2.0647	2.0592
	Median	2.0204	2.0152	2.0353
	St.dev.	0.3742	0.3743	0.3337
50%	Mean	2.6223	2.7414	2.7947
	Median	2.5026	2.6135	2.7116
	St.dev.	0.7726	0.7604	0.6583

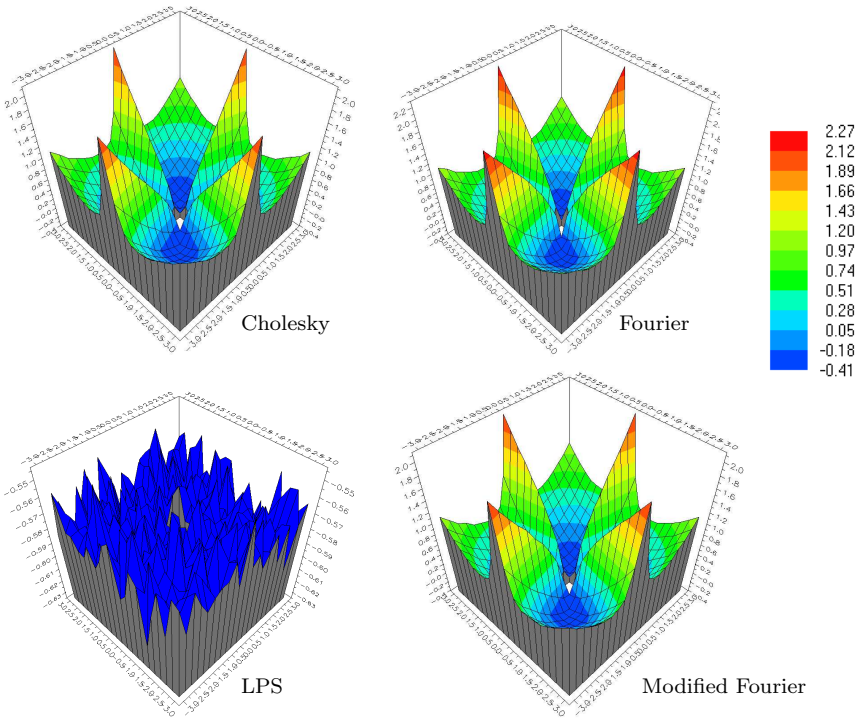

Figure 3.6: Matérn spectral density: $\nu = 0.5$, $\alpha = 80\%N$. Mean Error for the estimation of the log-spectral density. Linear-by-linear: simulations by Alonso *et al.* (1996). Fourier: simulations by FIM. Modified Fourier: simulations by MFIM. LPS: simulations by the extension of Fan's parametric Bootstrap. 20×20 grid.

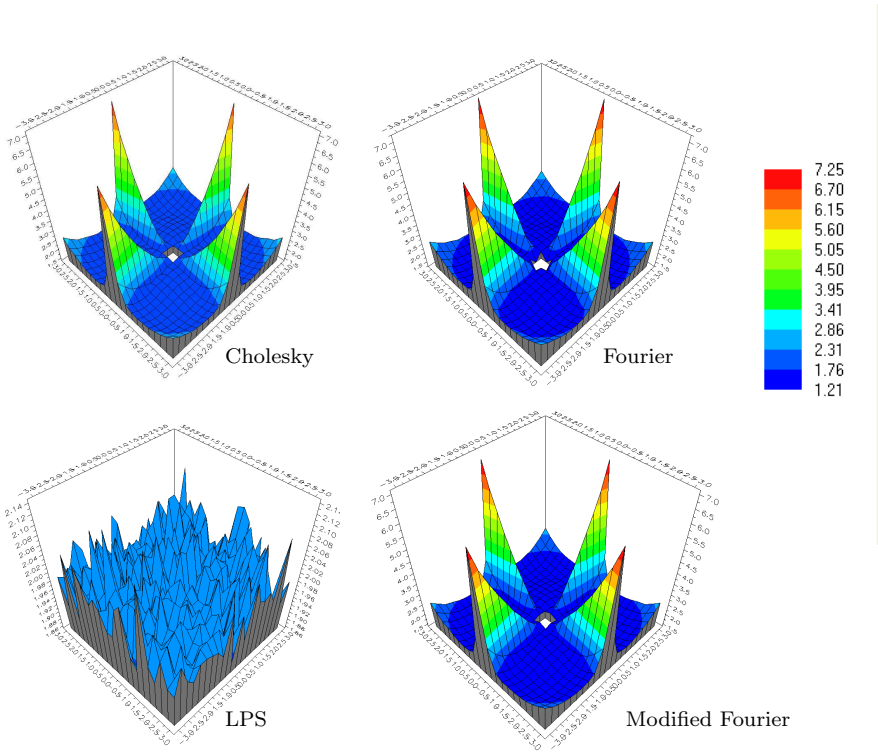

Figure 3.7: Matérn spectral density: $\nu = 0.5$, $\alpha = 80\%N$. Mean Square Error for the estimation of the log-spectral density. Linear-by-linear: simulations by Alonso *et al.* (1996). Fourier: simulations by FIM. Modified Fourier: simulations by MFIM. LPS: simulations by the extension of Fan's parametric Bootstrap. 20×20 grid.

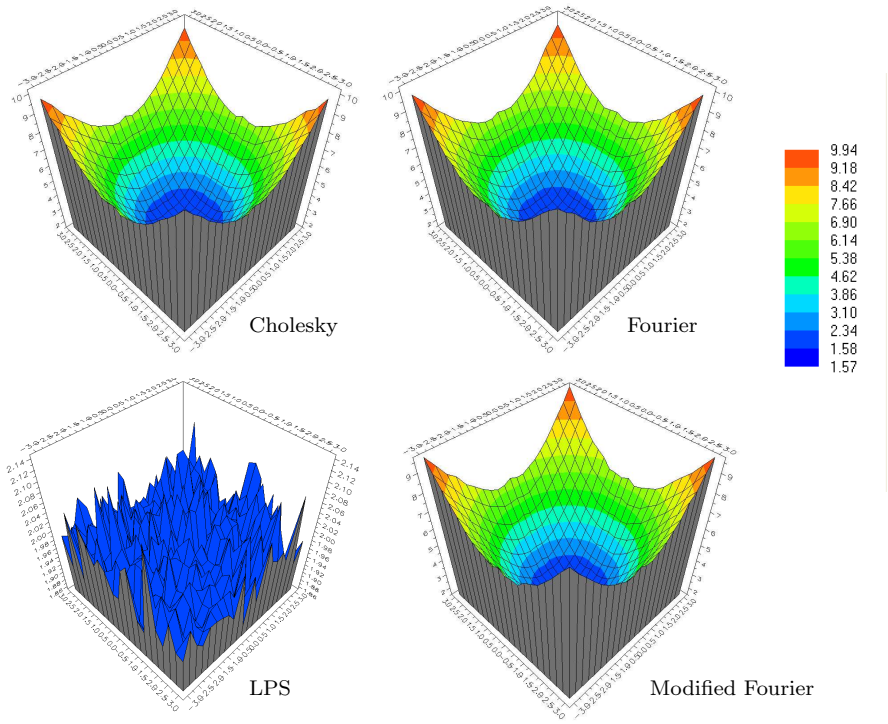

Figure 3.8: Matérn spectral density: $\nu = 0.5$, $\alpha = 80\%N$. Whittle's Error for the estimation of the log-spectral density. Linear-by-linear: simulations by Alonso *et al.* (1996). Fourier: simulations by FIM. Modified Fourier: simulations by MFIM. LPS: simulations by the extension of Fan's parametric Bootstrap. 20×20 grid.

Chapter 4

Goodness-of-fit tests for the spatial spectral density

Contents

This part of the dissertation is devoted to goodness-of-fit testing in spatial statistics. Before looking at the problem of goodness-of-fit testing in this context, we give a brief overview on goodness-of-fit tests for regression models. Nonparametric testing techniques in regression context have inspired both testing techniques for geostatistical and lattice data. In geostatistics, for instance, the variogram cloud can be thought as a dispersion plot from a certain regression model.

One could think about solving the crucial problem of modelling the spatial dependence in the spectral domain, instead of working with the variogram or the covariogram in the spatial setting (e.g. Fuentes (2002), for modelling non-stationary spatial dependence structures). The spatial spectral density is the Fourier Transform of the covariogram (see Section 1.3.4), so testing a certain covariance structure is equivalent to test a spatial spectral density model. From the theoretical point of view, as we have already commented along this manuscript, the main advantage of the spectral methodology is that the dependence between observations can be avoided, for a large enough sample. Therefore, traditional techniques on independent data may be applied to a suitable spectral transformation of the data, that is, the periodogram values at Fourier frequencies.

Besides, the periodogram can be obtained as the response variable in a multiplicative regression model, as in (2.14). In time series context, Paparoditis (2000) proposes a goodness-of-fit test based on a smoothed ratio between the periodogram and a parametric estimator of the spectral density, under the null hypothesis of an underlying parametric model. Equivalently, the log-periodogram can be seen as the exogenous variable in an additive regression model. This idea is considered in Fan and Zhang (2004), in time series context, where the authors apply a generalized likelihood ratio test for regression models (Fan *et al.* (2001)). In order to adapt a regression goodness-of-fit test to the spectral setting, other techniques could be considered. For instance, one could use tests based on the error distribution function, using the empirical process methodology (Stute (1997), Stute *et al.* (1998)). In time series case, Delgado *et al.* (2005) propose a goodness-of-fit test based on empirical processes. Other tests could be based on smoothed estimators of the regression function (Härdle and Mammen (1993), González Manteiga and Cao (1993) and Hart (1997), among others).

The main goal of this part is to show that one could take advantage of the goodness-of-fit test techniques for regression models and translate them into the spectral domain, in the sense that an estimator of the spectral density can be seen as the response variable in a regression model.

The first section of this part is devoted to a revision of goodness-of-fit tests for regression models, in order to provide an adequate background which allows the reader to understand the testing techniques for spatial models, both those tests we propose and other existing techniques. Goodness-of-fit tests for geostatistical data are also revised in Section 4.2. In Section 4.3, we

introduce two different testing techniques based on the spectral representation of the covariance structure, that is, the spectral density. We also provide a simulation study and real data applications, in Sections 4.4 and 4.5 respectively. Proofs of the theoretical results can be found in the Appendix of this chapter. The contents of this chapter can be found in Crujeiras *et al.* (2006b).

4.1 Goodness-of-fit tests for regression models

Nonparametric goodness-of-fit test for regression models may be mainly classified in two groups: tests based on the error distribution function, using the empirical process methodology (Stute (1997), Stute *et al.* (1998)) and those tests based on smoothed estimators of the regression function (Härdle and Mammen (1993), González Manteiga and Cao (1993) and Hart (1997), among others). For our brief overview, suppose that we have a sequence of independent observations $\{(X_i, Y_i), i = 1, \ldots, n\}$ from a population (X, Y). Consider, for the sake of simplicity, that both X and Y are one-dimensional random variables and assume that

$$m(x) = E(Y|X = x),$$

and consider the following model checking problem:

$$\begin{aligned} H_0 : &\quad m \in \mathcal{M}_\theta, \\ H_a : &\quad m \notin \mathcal{M}_\theta \end{aligned}$$

where $\mathcal{M}_\theta = \{m(\cdot, \theta), \theta \in \Theta\}$ is a given family of functions, with $\Theta \subset \mathbb{R}^p$.

A fundamental approach to this problem is the comparison between parametric and nonparametric models. Denote by $\hat{m}$ a nonparametric estimator of m, based on a linear smooth of $(Y_1, \ldots, Y_n)$ and let $\hat{\theta}$ denote a consistent estimate of θ, under the null hypothesis H_0.

For a local polynominal regression estimate $\hat{m}$, and considering $\hat{m}_{\hat{\theta}}(x)$ a local polynomial regression estimate of $m_{\hat{\theta}}(x)$, three major types of nonparametric regression tests can be considered (see Zhang and Dette (2004)):

- Härdle and Mammen (1993) propose the following test statistic

$$T_1 = \int (\hat{m}(x) - \hat{m}_{\hat{\theta}}(x))^2 dx, \tag{4.1}$$

  whereas González Manteiga and Cao (1993) simmultaneously studied a Riemann approximation to T_1

$$T_1^* = \sum_{i=1}^{n} (\hat{m}(X_i) - \hat{m}_{\hat{\theta}}(X_i))^2.$$

- Zheng (1996) introduces the test statistic

$$T_2 = \frac{1}{n}\sum_{i\neq j} K_h(X_i - X_j)(Y_i - m_{\hat{\theta}}(X_i))(Y_j - m_{\hat{\theta}}(X_j)), \tag{4.2}$$

  where K is a symmetric probability density function and $K_h(\cdot) = h^{-1}K(\cdot/h)$ in the rescaled kernel.

- Dette (1999) and Fan *et al.* (2001) consider

$$T_3 = \sum_{i=1}^{n}(Y_i - m_{\hat{\theta}}(X_i))^2 - \sum_{i=1}^{n}(Y_i - \hat{m}(X_i))^2. \tag{4.3}$$

Zhang and Dette (2004) provides a power comparison, under fixed and contiguous alternatives, of these three types of tests. These three tests statistics have some power detecting local alternatives, under null rate $(nh^{1/2})^{-1/2}$. But, from a local asymptotic point of view, ignoring the bias and using an identical bandwidth h for the three tests, then T_2 is more powerful than T_3, but it is less powerful than T_1, the L^2-distance test.

4.1.1 Empirical Process techniques.

Stute *et al.* (1998) introduce a testing technique which avoids smoothing the data. We briefly described this method.

Consider the integrated regression function defined as

$$I(x) = \int_{-\infty}^{x} m(u)dF(u), \quad x \in \mathbb{R} \tag{4.4}$$

where F denotes the unknown distribution function of the regressors X. The empirical analog of I becomes

$$I_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}} Y_i.$$

Consider a simple null hypothesis H_0: $m = m_{\theta_0}$. Then, a parametric estimator of the integrated regression function is given by:

$$I_0(x) = \int_{-\infty}^{x} m_{\theta_0}(y)\, dF_n(y) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}} m_{\theta_0}(X_i).$$

Thus, the difference between the empirical counterpart of I and the parametric estimator under H_0 is given by:

$$I_n(x) - I_0(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}} (Y_i - m_0(X_i)).$$

In this context, the empirical process marked by the regression errors:

$$\begin{aligned} R_n(x) &= n^{1/2}\left(I_n(x) - I_0(x)\right) \\ &= n^{-1/2}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}\left(Y_i - m_0(X_i)\right) \end{aligned}$$

has been studied by Stute (1997), who proved (under $E\left(Y^2\right) < \infty$) that

$$R_n \longrightarrow R_\infty,$$

in the Skorohod space $D[-\infty, \infty]$, where R_∞ is a Brownian motion with respect to time. In order to test whether $H_0 : m = m_\theta$ holds, we need to choose some functional (for instance the supremum that will lead to the Kolmogorov-Smirnov statistic). The critical value can be obtained from the distribution of such a functional computed from R_∞.

Under a composite null hypothesis H_0: $m \in \mathcal{M}_\theta$, consider $\hat{\theta}$ a consistent estimator of the true parameter θ_0. The goodness-of-fit test statistics is then based on the process

$$R_n^1(x) = n^{-1/2}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}\left(Y_i - m_{\hat{\theta}}(X_i)\right).$$

Under fairly general assumptions Stute (1997) proved that R_n^1 converges in distribution to a centered Gaussian limit, R_∞^1, with a quite complicate covariance structure. As a consequence, the principal components of R_∞^1 are difficult to obtain. This makes a real problem for full model checks, since optimal Neyman-Pearson tests for H_0 versus a given directional local alternative depend on these principal components. A solution to this problem is given by Bootstrap approximations.

Consider $\{(X_i^*, Y_i^*)\}_{i=1}^n$ a bootstrap resample of $\{(X_i, Y_i)\}_{i=1}^n$ and $\hat{\theta}^*$, the least squares estimator computed with this sample. The bootstrap version of R_n^1 is given by:

$$R_n^{1*}(x) = n^{-1/2}\sum_{i=1}^{n} 1_{\left\{X_i^* \leq x\right\}}\left(Y_i^* - m_{\hat{\theta}^*}(X_i^*)\right).$$

Name Ψ, a continuous functional to define the test statistic:

$$T_n = \Psi\left(R_n^1\right).$$

The null hypothesis H_0 is rejected if $T_n > c_\alpha^*$, for c_α^* satisfying

$$P^*\left(\Psi\left(R_n^{1*}\right) > c_\alpha^*\right) = \alpha.$$

Intuitively c_α^* is a reasonable estimator of the c_α satisfying

$$P\left(\Psi\left(R_n^1\right) > c_\alpha\right) = \alpha.$$

In practice we use

$$\hat{c}_\alpha^* = T_n^{*([B(1-\alpha)])}$$

the $[B(1-\alpha)]$-th order statistic of the bootstrap replications T_n^{*b}, $b = 1, 2, \ldots, B$.

If the Bootstrap sample is obtained by naive Bootstrap (based on the empirical distribution of the original sample), then this approach is inconsistent. González Manteiga and Cao (1993) proved that sampling from a bivariate distribution

$$\hat{F}_n(x, y) = n^{-1} \sum_{i=1}^{n} 1_{\{Y_i \le y\}} \int_{-\infty}^{x} K_h(t - X_i)\, dt$$

leads also to an inconsistent method. Stute *et al.* (1998) proved that a wild Bootstrap resampling is consistent and under H_0

$$R_n^{1*} \longrightarrow R_\infty^{1*},$$

with probability one in the space $D[-\infty, \infty]$, where R_∞^{1*} and R_∞^1 have the same distribution.

4.1.2 Likelihood ratio tests.

When testing a simple hypothesis vs. a simple alternative, it is shown that the most powerful test of a given size rejects the null hypothesis for small values of a likelihood ratio (by Neymann and Pearson's theorem). This fact led to the use of likelihood ratio tests in more general settings where the model is parametric and one or both of the hypothesis are composite.

Consider regression model

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\{\varepsilon_i\}$ are a sequence of i.i.d. random variables from $N(0, \sigma^2)$ and X_i has a density f with support $[0, 1]$. Suppose that the parameter space is

$$\mathcal{F}_k = \left\{ m \in L^2[0,1]; \int_0^1 m^{(k)}(x)^2 dx \le C \right\}$$

for a given constant C and consider the testing problem

$$\begin{aligned} H_0 &: \quad m(x) = \alpha_0 + \alpha_1 x, \\ H_a &: \quad m(x) \neq \alpha_0 + \alpha_1 x. \end{aligned} \tag{4.5}$$

The conditional log-likelihood function is given by

$$l_n(m) = -n\log(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - m(X_i))^2.$$

Denote by $(\hat{\alpha}_0, \hat{\alpha}_1)$ the maximum likelihood estimator under H_0 and by $\hat{m}$ the maximum likelihood estimator under the full model, that is:

$$\hat{m} = \arg\min_{m\in\mathcal{F}_k}\sum_{i=1}^{n}(Y_i - m(X_i))^2, .$$

In this case, the estimator obtained is a smoothing spline. The logarithm of the conditional maximum likelihood ratio statistic for the testing problem (4.5) is given by:

$$\Lambda = \frac{n}{2}\frac{RSS_0 - RSS_1}{RSS_1},$$

where

$$RSS_0 = \sum_{i=1}^{n}(Y_i - \hat{\alpha}_0 - \hat{\alpha}_1 X_i)^2, \quad \text{and} \quad RSS_1 = \sum_{i=1}^{n}(Y_i - \hat{m}(X_i))^2.$$

In an effort to derive a generally applicable testing procedure for multivariate nonparametric models, Fan *et al.* (2001) proposed a generalized likelihood ratio test. The authors point out that the maximum likelihood ratio test is not optimal due to its restrictive choice of smoothing parameters and, in general, maximum likelihood estimators under non parametric regression models are hard to obtain. In order to attenuate these difficulties, the maximum likelihood estimator under the alternative non parametric model may be replaced by a reasonable nonparametric estimator, leading to a generalized likelihood ratio:

$$\Lambda = l_n(H_1) - l_n(H_0),$$

where $l_n(H_1)$ denotes the log-likelihood with unknown regression function replaced by a non parametric regression estimator. In this case, the smoothing parameter can be selected to optimize the performance of the test.

An interesting feature of this type of tests is that, when Λ is based on a local linear estimator, the asymptotic distribution of the generalized likelihood ratio statistic exhibits a kind of Wilks phenomenon: under H_0, the asymptotic distribution of the ratio does not depend on nuisance parameters α_0, α_1 and σ^2 and the nuisance design density function. Fan *et al.* (2001) prove that this distribution is nearly a χ^2 with large degrees of freedom.

In time series context, Fan and Zhang (2004) consider a generalized likelihood ratio test in order to check whether a family of parametric time series models fits a set of data, without restrictions

on the forms of the alternative models. One of the tests statistic we introduce in this chapter is based in the ideas in Fan and Zhang (2004). Whereas the one-dimensional problem (time series) is a direct application of Fan *et al.* (2001), the results we obtain are based on extensions of the generalized likelihood ratio test .

4.2 Goodness-of-fit tests for spatial data.

Although we focus our attention on spectral techniques, the variogram is an important tool for the assessment of spatial variability and, of course, a crucial parameter for kriging. A first attempt to provide a testing technique for spatial variability, under stationarity and isotropy, is proposed by Diblasi and Bowman (2001). The authors propose a method for assessing the evidence for the presence of any spatial correlation, that is, a technique for checking for independence. The test statistic is based on a ratio of quadratic forms and the asymptotic distribution is studied, under the null hypothesis of independence. This null hypothesis (independence) implies that the variogram is constant.

The test statistic proposed in Diblasi and Bowman (2001) is given by:

$$T_{DB} = \frac{\sum_{i<j}(d_{ij} - \bar{d})^2 - \sum_{i<j}(d_{ij} - \hat{d}_{ij})^2}{\sum_{i<j}(d_{ij} - \hat{d}_{ij})^2}, \tag{4.6}$$

where d_{ij} is given by $|Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{1/2}$, $\bar{d}$ is the sample mean of the d_{ij} and $\hat{d}_{ij}$ represents a local linear approach of the variogram at distance $|s_i - s_j|$. The test statistic can be written as the ratio

$$T_{DB} = \frac{\mathbf{d}^T A \mathbf{d}}{\mathbf{d}^T B \mathbf{d}},$$

where $\mathbf{d}$ is the vector of the d_{ij}, $A = I - L$, N denotes the sample size, I is the identity matrix and L is a matrix filled with the value $2/(N(N-1))$; $B = (I - W)^T(I - W)$, where the rows of W consist of the smoothing weights used in the construction of the local linear estimator. The asymptotic null distribution for this test statistic is a shifted and scaled χ^2 although the authors proceed through the calculation of a p-value:

$$p = \mathbb{P}\{\mathbf{e}^T Q \mathbf{e} > 0\},$$

where $\mathbf{e}$ is zero-mean multivariate normal and matrix covariance Σ (see Diblasi and Bowman (2001) for more details) and $Q = A - tB$, where t denotes the observed value of the test statistic. The authors noticed the computational difficulties that may arise due to the sample size, as an initial sample of size n produces $N(N-1)/2$ differences d_{ij}. They propose binning to overcome this problem.

An extension of this test was considered by Maglione and Diblasi (2004) in order to assess a specific model for the variogram. The new test statistic is based on smoothed variables to reflect the correlation features, and when independence is the null hypothesis, the variance-covariance matrix of the variables involved in the statistic depends only on the distances between spatial locations. When a non-constant model for the variogram is considered in the null hypothesis, this variance-covariance matrix depends also on the shape of the variogram. The expression of the test statistic is:

$$T_{MD} = \frac{\sum_{k=1}^{\kappa}(S_k - a\gamma_0(h_k)^{1/4}\bar{S})^2 - \sum_{k=1}^{\kappa}(S_k - \hat{S}_k)^2}{\sum_{k=1}^{\kappa}(S_k - \hat{S}_k)^2},$$

where γ_0 denotes the variogram under the null hypothesis, $S_k = |R_k|^{1/2} - E(|R_k|^{1/2})$, with $k = 1, \ldots, \kappa = N(N-1)/2$ and $R_k = Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$. The variables $\hat{S}_k$ denote a smoothed approach of S_k and $a = 2^{1/2}\Gamma(3/4)\pi^{-1/2}$ is used to mathch $a\gamma_0(h_k)^{1/4}$ with the expected value of $|R_k|^{1/2}$, denoting by h_k the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$. As in the test for independence, the distribution is approximated by a shifted χ^2.

4.3 Testing the spatial spectral density

First of all, we will introduce the context where our study is carried out. Let Z be a zero-mean, second-order stationary process observed on a regular grid $D = \{0, \ldots, n_1 - 1\} \times \{0, \ldots, n_2 - 1\}$ and denote by $N = n_1 n_2$, number of observations and denote by C the covariance function of Z. Assuming that $\sum_{\mathbf{u}} |C(\mathbf{u})| d\mathbf{u} < \infty$, recall that by Khinchin's theorem (see Section 1.3), the covariance function of a stationary random process is the inverse Fourier Transform of the spectral density f.

We consider a spatial processes which can be represented as in (2.10). Then, the periodogram can be written as the response variable in a multiplicative regression model as in (2.14) and, after a logarithmic transform, the log-periodogram is the response variable in model (2.15).

In the next subsections we propose two different testing techniques, as the result of the extension to the multidimensional lattice data case of two tests: the first test is based on the ratio between the periodogram and the spectral density (see equation (2.14)). The second one consists on the extension of the generalized likelihood ratio test in regression models to a higher dimension particular case, that is, equation (2.15).

Our main goal is testing whether the spectral density for Z belongs to a parametric family $\mathcal{F}_\theta$,

with $\theta \in \Theta \subset \mathbb{R}^p$:

$$\begin{array}{lllll} H_0: & f & \in & \mathcal{F}_\theta = \{f_\theta; & \theta \in \Theta\}, \\ H_a: & f & \notin & \mathcal{F}_\theta = \{f_\theta; & \theta \in \Theta\}. \end{array} \tag{4.7}$$

Considering the log-spectral density, the problem can be written as

$$\begin{array}{lllll} \tilde{H}_0: & m & \in & \mathcal{M}_\theta = \{m_\theta; & \theta \in \Theta\}, \\ \tilde{H}_a: & m & \notin & \mathcal{M}_\theta = \{m_\theta; & \theta \in \Theta\}. \end{array} \tag{4.8}$$

The periodogram is written in (2.14) as the exogenous variable in a multiplicative regression model. From equation (2.15), the log-spectral density function m can be seen as a regression function in a model where the response is given by the log-periodogram (substracting a residual term $r_{\mathbf{k}}$) and the explanatory variables are the corresponding Fourier frequencies (fixed design case).

Provided that $n_1 \to \infty$, $n_2 \to \infty$ and $n_1/n_2 \to c$, for a constant c, the following assumptions on the process, spectral density and bidimensional kernel function, K, are needed.

Assumption 1. Assume the spatial process Z can be represented as in (2.10), and $\sum_{j,l} |j|^{1/2}|a_{j,l}| < \infty$, $\sum_{j,l} |l|^{1/2}|a_{j,l}| < \infty$ and $\sum_{j,l} |l|^4|j|^4|a_{j,l}| < \infty$. Assume also that the error process is such that $E(\varepsilon(\mathbf{s})) = 0$, $E(\varepsilon^2(\mathbf{s})) = \sigma^2$ and $E(\varepsilon^8(\mathbf{s})) < \infty$.

Assumption 2. The spectral density f is Lipschitz continuous and non vanishing, that is to say, $\inf_{\boldsymbol{\lambda} \in [-\pi,\pi]\times[-\pi,\pi]} f(\boldsymbol{\lambda}) > 0$.

Assumption 3. K is symmetric , bounded and non-negative bidimensional kernel with support $\Pi^2 = [-\pi, \pi] \times [-\pi, \pi]$, such that $\int_{\mathbb{R}^2} K(\mathbf{u})d\mathbf{u} = (2\pi)^2$ and $\int_{\mathbb{R}^2} K^2(\mathbf{u})d\mathbf{u} < \infty$. The rescaled kernel K_H is defined by $K_H(\mathbf{u}) = |H|^{-1/2}K(H^{-1/2}\mathbf{u})$, following (Ruppert and Wand (1994)). The sequence of bandwidth matrices is such that each entry of H tends to zero and $N|H|^{1/2} \to \infty$. Some further assumptions on the bandwidth matrix are needed in Theorem 4.

Assumption 4. The parameter space Θ is an open subset of $\mathbb{R}^p$ and the spectral density f_θ is twice differentiable w.r.t. θ with continuous second derivatives.

4.3.1 Using the periodogram for hypothesis testing.

It is known that, if assumption 1 holds,

$$E\left(\frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_\theta(\boldsymbol{\lambda}_\mathbf{k})}\right) = \frac{f(\boldsymbol{\lambda}_\mathbf{k})}{f_\theta(\boldsymbol{\lambda}_\mathbf{k})} + \mathcal{O}(N^{-1}\log N), \tag{4.9}$$

uniformly in $\mathbf{k}$. Condition (4.9) implies that, under H_0, the asymptotic expected value of this ratio equals one. We consider a squared deviation criterion on a kernel type estimator of the ratio between the periodogram and the spectral density (under H_0), as it is proposed in Paparoditis (2000) for the one-dimensional case.

Paparoditis (2000) gives procedures for testing both a simple and a composite hypothesis, based on a smoothed estimator of the ratio $I(\lambda)/f(\lambda)$. When testing a composite hypothesis $H_0 : f = f_\theta$ against $H_a : f \neq f_\theta$, a generalization of the test statistic proposed to the bidimensional case is given by:

$$T_P = N|H|^{1/4}\int_{\Pi^2}\left(\frac{1}{N|H|^{1/2}}\sum_{\mathbf{k}} K(H^{-1/2}(\boldsymbol{\lambda}-\boldsymbol{\lambda}_\mathbf{k}))\left(\frac{I(\boldsymbol{\lambda})}{f_{\hat\theta}(\boldsymbol{\lambda})}-1\right)\right)^2 d\boldsymbol{\lambda}, \tag{4.10}$$

where $\sum_{\mathbf{k}}$ extends over the Fourier frequencies and $\hat\theta$ is the Whittle estimator. Asymptotic normality of this statistic is also obtained.

Theorem 1. *Under assumptions (1)-(4) and under $H_0 : f_\theta \in \mathcal{F}_\theta$*

$$T_P - \mu_H \to N(0,\tau^2) \quad \textit{in distribution},$$

where μ_H and τ^2 are given by:

$$\mu_H = |H|^{-1/4}\int K^2(\mathbf{s})d\mathbf{s}, \tag{4.11}$$

$$\tau^2 = \frac{1}{2\pi^2}\int_{2\Pi^2}\left(\int_{\Pi^2} K(\mathbf{s})K(\mathbf{s}+\mathbf{u})d\mathbf{s}\right)^2 d\mathbf{u}, \quad 2\Pi^2 = [-2\pi,2\pi]\times[-2\pi,2\pi]. \tag{4.12}$$

We consider now the consistency properties of the test when testing a composite hypothesis, in the case that the true spectral density lies in $\mathcal{F}-\mathcal{F}_\theta$. Then, the Whittle estimator $\hat\theta$ is an efficient estimator of θ^*, where:

$$\theta^* = \arg\min_\theta L(\theta, f)$$

and

$$L(\theta,f) = \int_{\Pi^2}\left(\log f_\theta(\boldsymbol{\lambda}) - \frac{f(\boldsymbol{\lambda})}{f_\theta(\boldsymbol{\lambda})}\right)d\boldsymbol{\lambda}.$$

θ^* is not the true parameter, but it determines the best fit in $\mathcal{F}_\theta$. $L(\theta, f)$ is the Kullback-Leibler divergence between a Gaussian process and a Gaussian model. Following Dahlhaus and Wefelmeyer (1996), we can see that:

$$\sqrt{N}(\hat{\theta} - \theta^*) = \sqrt{N} \int_{\Pi^2} W(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda}))d\boldsymbol{\lambda} + o_{\mathbb{P}}(1) \tag{4.13}$$

where

$$W(\boldsymbol{\lambda}) = -\mathcal{H}^{-1} \nabla f_\theta^{-1}(\boldsymbol{\lambda})|_{\theta=\theta^*}, \qquad \mathcal{H} = \int_{\Pi^2} \nabla^2 D(\theta^*, f, \boldsymbol{\lambda}) d\boldsymbol{\lambda},$$

$$D(\theta^*, f, \boldsymbol{\lambda}) = \log f_{\theta^*}(\boldsymbol{\lambda}) + \frac{f(\boldsymbol{\lambda})}{f_{\theta^*}(\boldsymbol{\lambda})},$$

and ∇ and ∇^2 denote the first and second derivatives with respect to θ.

Some further assumptions must be made in order to guarantee that $\hat{\theta}$ is a consistent estimator of θ^*.

Assumption 5. $\Theta \subset \mathbb{R}^p$ is compact and f_θ is three times differentiable with respect to θ, with continuous derivatives. Besides, θ^* exists, is unique and lies in the interior of Θ.

Theorem 2. *Consider the problem of testing a composite hypothesis $H_0 : f \in \mathcal{F}_\theta$ vs. $H_a : f \in \mathcal{F} - \mathcal{F}_\theta$. Under assumptions (1)-(3) and (5), as $n_1, n_2 \to \infty$:*

$$N^{-1}|H|^{-1/4} T_P \to \int_{\Pi^2} \left(\frac{f(\boldsymbol{\lambda})}{f_{\theta^*}(\boldsymbol{\lambda})} - 1 \right)^2 d\boldsymbol{\lambda}$$

in probability.

This result is analogous to Theorem 3 in Paparoditis (2000) for time series context. As in the one-dimensional situation, this result implies the omnibus property of the T_P test, that is, T_P is consistent against any alternative such that $f \notin \mathcal{F}_\theta$. The power function of this test is a monotone increasing function. It is given in terms of the L^2 distance between f, the true spectral density, and f_{θ^*}, its best approximation in $\mathcal{F}_\theta$ (given the Kullback-Leibler discrepancy). An analogous result is obtained for the simple hypothesis case.

The problem of testing a simple hypothesis $H_0 : f_\theta = f_{\theta_0}$ vs. $H_a : f_\theta \neq f_{\theta_0}$ is solved using the test statistic T_P^0, which is obtained from T_P just replacing $\hat{\theta}$ for the parameter under H_0, θ_0. It is proved in the appendix that T_P^0 has the same limit behaviour.

Theorem 3. *Consider the simple hypothesis testing problem $H_0 : f_\theta = f_{\theta_0}$ vs. $H_a : f_\theta \neq f_{\theta_0}$ and let $f \in \mathcal{F} - \{f_{\theta_0}\}$ be the true spectral density. Under the assumptions of Theorem 2, as $n_1, n_2 \to \infty$:*

$$N^{-1}|H|^{-1/4}T_P^0 \to \int_{\Pi^2} \left(\frac{f(\boldsymbol{\lambda})}{f_{\theta_0}(\boldsymbol{\lambda})} - 1\right)^2 d\boldsymbol{\lambda}$$

in probability.

These consistency properties of the test guarantee that, for any level $\alpha \in (0, 1)$, the probability of rejecting the null hypothesis under the alternative approaches 1, as the sample size increases.

4.3.2 Using the log-periodogram for hypothesis testing.

In this part, we tackle the testing problem (4.8). Consider the following regression model:

$$Y_{\mathbf{k}}^{**} = m(\boldsymbol{\lambda}_{\mathbf{k}}) + z_{\mathbf{k}}^*, \tag{4.14}$$

where we denote by $Y_{\mathbf{k}}^{**} = Y_{\mathbf{k}}^* - r_{\mathbf{k}}$, $Y_{\mathbf{k}}^* = Y_{\mathbf{k}} - C_0$ and $z_{\mathbf{k}}^* = z_{\mathbf{k}} - C_0$. The $Y_{\mathbf{k}}^{**}$ variables are not observed, so we establish the testing procedure in terms of $Y_{\mathbf{k}}$, although the theoretical reasoning takes this fact into account.

Following Fan and Zhang (2004), we introduce the generalized likelihood ratio test statistic based on two likelihood approaches of equation (2.14). The first approach is given by the loglikelihood maximization under the null hypothesis. The second approach is purely non-parametric, obtained by a local loglikelihood function maximization. The loglikelihood function associated with (2.15), when $r_{\mathbf{k}}$ has been removed, is

$$\sum_{\mathbf{k}} \left[Y_{\mathbf{k}} - m(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y_{\mathbf{k}} - m(\boldsymbol{\lambda}_{\mathbf{k}})}\right], \tag{4.15}$$

as we have seen in Section 2.5.1. We will introduce two likelihood-based approaches to obtain the generalized likelihood ratio test statistic. Under the null hypothesis, the maximizer of the loglikelihood function of (2.15), when ignoring the residual part $r_{\mathbf{k}}$, is the Whittle estimate from equation (2.84).

From a nonparametric approach, we have seen in Section 2.5.1 that the log-spectral density function m can be approximated by a multidimensional local linear kernel estimator. The local maximum likelihood estimator $\hat{m}_{LK}(H, x) \equiv \hat{m}_{LK}(\mathbf{x})$ of $m(\mathbf{x})$ is $\hat{a}$ in the maximizer $(\hat{a}, \hat{b})$ of (2.89), where the rescaled kernel K_H satisfies assumption 3. Then, a generalized likelihood test statistic can be constructed as

$$T_{LK} = \sum_{\mathbf{k}} \left[e^{Y_{\mathbf{k}} - m_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}})} + m_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})\right]. \tag{4.16}$$

The local estimator $\hat{m}_{LK}$ contains biases even under the null hypothesis which affect the distribution under H_0. In the regression context, Härdle and Mammen (1993) in order to compare parametric vs. nonparametric regression fits, propose smoothing the residuals from both approaches. The bias correction technique consists on a reparametrization of the log-periodogram. Let θ denote the true parameter under H_0 and rewrite $m^{BC}(\boldsymbol{\lambda}) = m(\boldsymbol{\lambda}) - m_\theta(\boldsymbol{\lambda})$. Then, the hypothesis testing statement, in terms of m^{BC} is given by:

$$\begin{aligned} H_0 &: m^{BC} = 0, \\ H_a &: m^{BC} \neq 0. \end{aligned}$$

The expression for the test statistic is:

$$T_{LK,BC} = \sum_{\mathbf{k}} \left(e^{\tilde{Y}_{\mathbf{k}}} - e^{\tilde{Y}_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) \right),$$

where $\hat{\theta}$ is the Whittle estimator of θ and $\tilde{Y}_{\mathbf{k}} = Y_{\mathbf{k}} - m_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}})$ denote the synthetic data. m^*_{LK} is the local linear estimator of m^{BC}, considering these synthetic data. Although asymptotic distribution of the test statistic is also obtained, in practice, we approximate the null distribution of T_{LK} using Monte Carlo simulations. Consider the following decomposition of the test statistic.

$$T_{LK} = T_{LK,1} - T_{LK,2}$$

where

$$T_{LK,1} = \sum_{\mathbf{k}} \left[e^{(Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}))} + m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{(Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) \right]$$

$$T_{LK,2} = \sum_{\mathbf{k}} \left[e^{(Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}))} + m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{(Y_{\mathbf{k}} - m_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}}))} - m_{\hat{\theta}}(\boldsymbol{\lambda}\mathbf{k}) \right]$$

The test statistic $T_{LK,1}$ is the generalised likelihood ratio test statistic for testing between

$$\begin{aligned} \tilde{H}_0 : m &= m_\theta \\ \tilde{H}_a : m &\neq m_\theta \end{aligned}$$

while $T_{LK,2}$ is the maximum likelihood ratio test statistic for testing between

$$\begin{aligned} \bar{H}_0 : \theta &= \theta_0 \\ \bar{H}_a : \theta &\neq \theta_0 \end{aligned}$$

where θ_0 denotes the true parameter in the parametric family of models $\mathcal{M}_\theta$. For simplicity, we will denote the true parameter by θ, instead of θ_0 and the spectral density of Z will be denoted by f_θ. Under certain regularity conditions, the asymptotic null distribution of $T_{LK,2}$ is χ^2_p, where $p = dim(\theta)$. Hence, $T_{LK,2} = O_{\mathbb{P}}(1)$. Therefore, we can simplify the test statistic to $T_{LK,1}$ with a simple null hypothesis test:

$$T_{LK} = \sum_{\mathbf{k}} \left[e^{(Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}))} + m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{(Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) \right]. \tag{4.17}$$

In order to study the asymptotic properties of this statistic, we decompose T_{LK} in some addends. We consider T^*_{LK}, which is the same statistic as T_{LK} but replacing $Y_{\mathbf{k}}$ by $Y^{**}_{\mathbf{k}}$ given in equation (4.14) and $\hat{m}_{LK}$ by $\hat{m}^*_{LK}$. If the observed test statistic is larger than a selected critical value, then we reject the null hypothesis.

Define also the following quantities, related to the asymptotic distribution of the test statistic:

$$\mu_H = \frac{4\pi^2}{|H|^{1/2}} \left(K(\mathbf{0}) - \frac{1}{2} \int K^2(\mathbf{s}) d\mathbf{s} \right), \tag{4.18}$$

$$b_H = \sum_{\mathbf{k}} \frac{-|H|^2}{8 f_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} \int \int \mathbf{s}^T H_{m_\theta}(\boldsymbol{\lambda}_{\mathbf{k}}) \mathbf{s} \cdot (\mathbf{s}+\mathbf{u})^T H_{m_\theta}(\boldsymbol{\lambda}_{\mathbf{k}})(\mathbf{s}+\mathbf{u}) K(\mathbf{s}) K(\mathbf{s}+\mathbf{u}) d\mathbf{s} d\mathbf{u}, \tag{4.19}$$

$$\sigma^2 = \frac{2\pi^2}{|H|^{1/2}} \int \left(2K(\mathbf{s}) - K * K(\mathbf{s})\right)^2 d\mathbf{s}. \tag{4.20}$$

Where $H_{m_\theta}(\boldsymbol{\lambda}_{\mathbf{k}})$ is the Hessian matrix of m_θ.

Theorem 4. *Under assumptions (1)-(4), as $N^{(\zeta-1)/\zeta}|H|^{1/2} \geq c \log^\delta N$, for a constant c and some $\delta > (\zeta-1)/(\zeta-2))$, $\zeta > 2$ and provided that H_0 holds,*

$$\sigma^{-1}(T_{LK} - \mu_H + b_H) \to N(0,1),$$

where μ_H, b_H and σ^2 are given by (4.18), (4.19) and (4.20), respectively.

The former theorem extends Theorem 1 in Fan and Zhang (2004) to the multidimensional setting. Other goodness-of-fit testing techniques based on smoothed estimators $\hat{m}$ of the log-spectral density could be used. An L^2-approach could be considered:

$$T_C = \sum_{\mathbf{k}} (\hat{m}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}}))^2. \tag{4.21}$$

This test statistic was studied by González Manteiga and Cao (1993) (and simultaneously by Härdle and Mammen (1993), in a continuous form). For the test statistics (4.21), asymptotic normal distributions is obtained in the one-dimensional case. Also in the one-dimensional case, Zhang and Dette (2004) give a power comparison between nonparametric regression tests. Similarly, it would be possible to obtain the normal asymptotic distribution of the extensions of these tests.

The results in this section can be generalized for stationary random fields on $\mathbb{R}^d$, under a similar asymptotic framework. The d-variate kernel K (with support on $\Pi^d = [-\pi, \pi]^d$) and the $d \times d$ bandwidth matrix H must satisfy the corresponding assumption 3. For the T_P test, the expressions for the mean and the variance are given by:

$$\mu_H^{(d)} = \frac{1}{|H|^{1/4}} \int_{\Pi^d} K^2(\mathbf{s}) d\mathbf{s},$$

$$\tau^{2(d)} = \frac{1}{2^{d-1}\pi^d} \int_{2\Pi^d} \left(\int_{\Pi^d} K(\mathbf{s}) K(\mathbf{s}+\mathbf{u}) \right)^2 d\mathbf{u}, \quad 2\Pi^d = [-2\pi, 2\pi].$$

For the T_{LK} test:

$$\mu_H^{(d)} = \frac{(2\pi)^d}{|H|^{1/2}} \left(K(\mathbf{0}) - \frac{1}{2} \int_{\Pi^d} K^2(\mathbf{s}) d\mathbf{s} \right)$$

$$\sigma^{2(d)} = \frac{2^{d-1}\pi^d}{|H|^{1/2}} \int_{\Pi^d} \left(2K(\mathbf{s}) - K * K(\mathbf{s})\right)^2 d\mathbf{s}.$$

These expressions generalize the results in this section and those provided by Paparoditis (2000) and Fan and Zhang (2004).

4.3.3 Testing in practice.

Since the rate of convergence of the distributions of T_P and T_{LK} to their Gaussian limit is quite slow, we show an alternative way of estimating the distribution of the test statistic, under H_0, by a Monte Carlo approach. The performance of T_P and T_{LK} tests is shown in a simulation study. We propose the following algorithm, for computing the $p-$value of the test statistics T_P and T_{LK}:

Step 1. Obtain the parametric estimate $\hat{\theta}$.

Step 2. Compute the observed test statistic T^{obs}. For the T_P test:

$$T_P^{obs} = N|\hat{H}|^{1/4} \sum_{\mathbf{k}} \left(N^{-1}|\hat{H}|^{-1/2} \sum_{\mathbf{j}} K\left(H^{-1/2}(\boldsymbol{\lambda}_\mathbf{k} - \boldsymbol{\lambda}_\mathbf{j})\right) \left(\frac{I(\boldsymbol{\lambda}_\mathbf{j})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{j})} - 1 \right) \right)^2$$

and for the T_{LK}, obtain the non-parametric estimate $\hat{m}_{LK}(H, \cdot)$ and:

$$T_{LK}^{obs} = \sum_{\mathbf{k}} \left\{ e^{Y_\mathbf{k} - m_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{k})} - e^{Y_\mathbf{k} - \hat{m}_{LK}(\boldsymbol{\lambda}_\mathbf{k})} + m_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{k}) - \hat{m}_{LK}(\boldsymbol{\lambda}_\mathbf{k}) \right\},$$

Step 3. From $f_{\hat{\theta}}$, generate a random sample of size $N = n_1 \cdot n_2$.

Step 4. Using the generated random sample in *Step 3*, obtain the test statistic T^*.

Step 5. Repeat B times steps 3 and 4 and obtain the bootstrap test statistics $T_1^*, T_2^*, \ldots, T_B^*$.

Step 6. Compute the p-value of the test statistic as the percentage of the bootstrap replicates $\{T_1^*, T_2^*, \ldots, T_B^*\}$ that exceed T^{obs}.

Both for T_P and T_{LK} non-linear multidimensional optimization problems must be solved. Whittle estimates $\hat{\theta}$ are obtained in *Step 1*, using a discretized version of (2.85). Newton type methods can be used to solve this problem, although these methods are not suitable for situations

where local maximum values are found. In order to guarantee the convergence to a global maximum, genetic algorithms were implemented (e.g. Goldberg (1989)).

In the case of the algorithm for T_{LK}, the computational cost is highly increased in *Step 2* with the nonparametric estimation of the log-spectral density, obtained by local maximum loglikelihood. There is again a non-linear multidimensional optimization problem, which must be solved for every Fourier frequency. For each $\boldsymbol{\lambda}_{\mathbf{k}}$, we take $(Y_{\mathbf{k}}, \mathbf{0})$ as initial values of $(a, \mathbf{b})$ in (2.89). As it happens for solving *Step 1*, one could think of using genetic algorithms for avoiding convergence problems.

A key problem in nonparametric statistics is the selection of the bandwidth parameter. Optimal bandwidth selection for nonparametric testing in multidimensional problems is still an open question. Usually, in practice, the standard approach consists of examining a range of bandwidths.

Automatic bandwidth selection criteria is another alternative. For instance, the bandwidth matrix could be chosen by minimizing the Mean Integrated Square Error of the nonparametric estimator under the null hypothesis that $H_0 : f = f_{\theta_0}$:

$$\hat{H} = \arg\min_{H} E\left(\int_{\Pi^2} (\hat{m}_{LK}(H, \boldsymbol{\lambda}) - m_{\theta_0}(\boldsymbol{\lambda}))^2 d\boldsymbol{\lambda}\right). \tag{4.22}$$

Bandwidth estimation can be obtained using a Monte Carlo approach of the MISE error (4.22):

$$\hat{H} = \arg\min_{H} \frac{1}{M} \sum_{j=1}^{M} \int_{\Pi^2} \left(\hat{m}_{LK}(H, \boldsymbol{\lambda}) - m_{\theta_0}(\boldsymbol{\lambda})\right)^2 d\boldsymbol{\lambda}, \tag{4.23}$$

although in practice, the theoretical parameter θ_0 is replaced by an estimator $\hat{\theta}$. However, the computational cost of this approach can be really high in some cases (due to the computation of the local log-likelihood estimator). Since log-periodogram values are asymptotically independent, for a large enough sample, good approximations are expected using a traditional cross-validation criteria. That is, select $\hat{H}$ such that:

$$\hat{H} = \arg\min_{H} \sum_{\mathbf{k}} \left(\hat{m}_{LK}^{-k}(H, \boldsymbol{\lambda}_{\mathbf{k}}) - m_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}})\right)^2, \tag{4.24}$$

where $\hat{m}_{LK}^{-k}(H, \cdot)$ is the nonparametric estimator of the log-spectral density obtained by maximizing expression (2.89), deleting the frequency $\boldsymbol{\lambda}_{\mathbf{k}}$.

It is important to note that the bandwidth matrix H plays a different role in both test statistics. In the T_{LK} test, the bandwidth matrix is involved in the nonparametric estimation of

the log-spectral density. In the T_P test, the bandwidth matrix is not involved in the estimation procedure. Therefore, it may be expected that this test statistic will be less influenced by the bandwidth parameter.

The algorithm we propose for calibrating the $p-$value of the test statistics needs, in *Step 3*, the generation of a sample of size N, given a parametrically estimated spectral density function $f_{\hat{\theta}}$. For that purpose, we consider a spectral simulation procedure, as the MFIM, introduced in Chapter 2.

Remark. If Z is a continous process (geostatistical data), the summation in representation (2.10) is replaced by an integral (Priestley (1981)) and the spectrum of such a process is defined for all $\boldsymbol{\lambda}$ in $\mathbb{R}^2$. Although asymptotic theory has not been yet obtained in this case, the tests can be applied, with suitable modifications, when the observations are taken on a regular grid. In this case, the spectral density estimators should be modified in order to account for the spacing between data (e.g. Fuentes (2002)).

4.4 Simulation study.

In this section, we study the performance of the testing procedures in terms of size and power. For illustration purposes, we consider the bidimensional autoregressive process (the $BAR(1)$ from Chapter 3), defined as in (3.23). Parameters β_1 and β_2 in (3.23) belong to $[0, 1)$ to guarantee stationarity. The spectral density of this process can be factorized with respect to β_1 and β_2 as

$$f(\boldsymbol{\lambda}) = \frac{\sigma_\varepsilon^2}{(2\pi)^2} \cdot \frac{1}{1 + \beta_1^2 - 2\beta_1 \cos(\lambda_1)} \cdot \frac{1}{1 + \beta_2^2 - 2\beta_2 \cos(\lambda_2)}. \tag{4.25}$$

In order to study the size of the tests, we consider different values for the parameters β_1 and β_2 from 0.0 (which corresponds to the independence case) to 0.9. 1000 simulations of the process are generated on a 20×20 and 50×50 regular grid. Random sample generations of this process are obtained as in (Alonso *et al.* (1996)). Estimators for β_1 and β_2 are obtained from the periodogram of the generated data, using a discretized version of the Whittle log-likelihood (2.85).

We set the null hypothesis that Z is a doubly-geometric process, considering different parameters. A multiplicative Epanechnikov bidimensional kernel is used along the study. The bandwidth parameter has been chosen using the cross-validation criteria (4.24). In order to simplify the computations, we consider diagonal bandwidth matrices, with elements proportional to the spacing between frequencies:

$$H = r \cdot diag\left(\frac{2\pi}{n_1}, \frac{2\pi}{n_2}\right). \tag{4.26}$$

	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
(β_1, β_2)	T_P	T_{LK}	T_P	T_{LK}	T_P	T_{LK}
$(0.0, 0.0)$	0.014	0.009	0.043	0.054	0.090	0.105
$(0.1, 0.1)$	0.014	0.014	0.043	0.045	0.090	0.085
$(0.2, 0.2)$	0.018	0.011	0.051	0.049	0.024	0.088
$(0.3, 0.3)$	0.021	0.080	0.058	0.052	0.112	0.100
$(0.4, 0.4)$	0.020	0.090	0.058	0.053	0.099	0.099
$(0.5, 0.5)$	0.022	0.014	0.058	0.054	0.103	0.105
$(0.6, 0.6)$	0.023	0.015	0.067	0.059	0.113	0.117
$(0.7, 0.7)$	0.044	0.037	0.104	0.097	0.172	0.161
$(0.8, 0.8)$	0.096	0.067	0.210	0.171	0.289	0.225
$(0.9, 0.9)$	0.170	0.189	0.346	0.347	0.443	0.457
$(0.1, 0.9)$	0.088	0.092	0.195	0.186	0.287	0.264

Table 4.1: Size of the tests. 20×20 grid.

	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
(β_1, β_2)	T_P	T_{LK}	T_P	T_{LK}	T_P	T_{LK}
$(0.7, 0.7)$	0.018	0.023	0.060	0.056	0.125	0.105
$(0.9, 0.9)$	0.097	0.052	0.269	0.114	0.396	0.168

Table 4.2: Size of the tests, 50×50 grid.

	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
(β_1, β_2)	T_P	T_{LK}	T_P	T_{LK}	T_P	T_{LK}
$(0.7, 0.7)$	0.027	0.019	0.054	0.049	0.097	0.098
$(0.8, 0.8)$	0.034	0.030	0.075	0.070	0.119	0.119
$(0.9, 0.9)$	0.048	0.053	0.107	0.112	0.165	0.169
$(0.1, 0.9)$	0.028	0.031	0.072	0.069	0.131	0.117

Table 4.3: Size of the tests, 20×20 grid, with bias correction on the parameter estimates.

The behaviour of the test in size terms is shown in Table 4.1, at three different significance levels: 0.01, 0.05 and 0.10. The percentage of rejections of both test statistics are computed from 1000 simulations. The results are quite satisfactory for both test, when the autoregression parameters are smaller than 0.5. For autoregresion parameters near 1, the performance is not so good as in the previous cases. It may happens that, for high dependence parameters, this sample size is too small for hypothesis testing.

As an example, in Table 4.2, we show the results of applying T_{LK} and T_P for parameters $(0.7, 0.7)$ and $(0.9, 0.9)$, in a 50×50 regular grid. Despite increasing the sample size, the size of the test does not improve as it could be expected. In Figure 4.1 we observed that, for a 20×20 regular grid, large autoregression parameter estimates from Whittle's likelihood are seriously biased. It seems clear that the bias in the parametric estimation distorts the results in the approximation of the size of the tests.

As we have already commented, Whittle parameter estimates computed from the raw periodogram are biased. We propose a bootstrap correction technique, which can be included in the Bootstrap procedure for approximating the test statistic distribution. The modifications in the algorithm described in the section 3.3, in order to include the bias correction technique, are the following:

Step 1. Obtain the parametric estimate $\hat{\theta}$.

1.A. Generate B' random samples of size N from $f_{\hat{\theta}}$.

1.B. Estimate $\hat{\theta}_i^*$ for each sample.

1.C. $\hat{b}(\theta, \hat{\theta}) = \frac{1}{B'} \sum_i (\hat{\theta} - \hat{\theta}_i^*)$.

1.D. Replace $\hat{\theta}$ by the bias corrected version $\hat{\theta} + \hat{b}(\theta, \hat{\theta})$.

...

Step 5. Using the generated random sample in *Step 4*, obtain the test statistic T^*, correcting the parameter estimator $\hat{\theta}^*$ by $\hat{\theta}^* + \hat{b}(\theta, \hat{\theta})$, and repeat B times steps 3 and 4.

The percentage of rejections of both tests, in a 20×20 grid, when applying the Bootstrap bias correction on the parameter estimates, is shown in Table 4.3. Significative improvements are observed in all cases, although for parameters near one, the results are not still completely satisfactory.

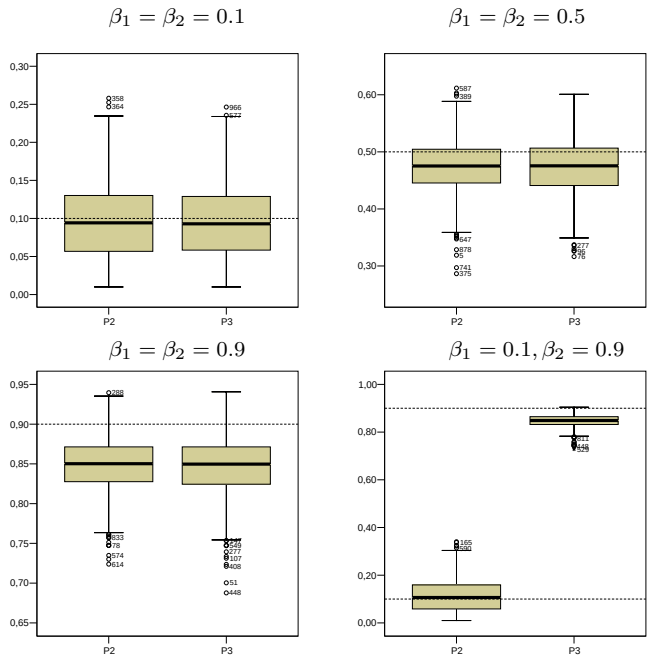

Figure 4.1: Parameter estimates.

Behaviour of the test in terms of power is shown in Table 4.4, when testing for independence, that is $H_0 : f = c$, for some positive constant c. We set as alternatives different parameters approaching the null hypothesis. It seems that T_{LK} performs better than T_P. This feature may be explained by the fact that the bandwidth matrix approximates the optimal bandwidth for the nonparametric estimation.

	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
(β_1, β_2)	T_P	T_{LK}	T_P	T_{LK}	T_P	T_{LK}
$(0.01, 0.01)$	0.017	0.009	0.062	0.047	0.110	0.088
$(0.05, 0.05)$	0.036	0.017	0.096	0.079	0.169	0.148
$(0.1, 0.1)$	0.085	0.097	0.192	0.254	0.307	0.374
$(0.2, 0.2)$	0.376	0.713	0.589	0.903	0.720	0.943
$(0.3, 0.3)$	0.882	0.993	0.952	1.000	0.980	1.000

Table 4.4: Power of the tests.Testing for independence.

4.5 Real data application.

The testing techniques we propose have been applied to different data sets. First, the tests are applied to Mercer and Hall wheat data, a classical example. We also use the tests in order to check spatial dependence between heavy metal depositions in mosses. In fact, this is the most appealing example, due to the ecological implications of the results.

4.5.1 Mercer and Hall wheat data.

Wheat data from Mercer and Hall experiment have been introduced in Section 1.4.1. Whittle (1954) fitted a zero-mean, first-order autoregressive model:

$$Z(\mathbf{s}) = \alpha_1(Z(s_1+1,s_2)+Z(s_1-1,s_2)) + \alpha_2(Z(s_1,s_2+1)+Z(s_1,s_2-1)) + \varepsilon(\mathbf{s}), \qquad (4.27)$$

where $\varepsilon(\mathbf{s})$ are zero-mean independent Gaussian random variables, with variance σ_ε^2. The corresponding spectral density is given by

$$f(\boldsymbol{\lambda}) = \frac{\sigma^2}{(2\pi)^2}\left(1-2\alpha_1\cos(\boldsymbol{\lambda}_1)-2\alpha_2\cos(\boldsymbol{\lambda}_2)\right)^{-2}. \qquad (4.28)$$

We will refer to model (4.27) as the spatial autoregressive model ($SAR(1)$ model).

As a first approach, we test for independence, using both T_{LK} and T_P test statistics. We examine a range of diagonal bandwidth matrix (4.26), with r varying from 2.0 to 20.0. In both cases, the hypothesis of independence is rejected (p-values lower than 0.001) along the whole bandwidth range.

Once the independence hypothesis is rejected, we apply T_{LK} and T_P in order to check that model (4.27) fits the data. We obtain as estimated parameters $\widehat{\alpha_1} = 0.23217$, $\widehat{\alpha_2} = 0.09267$ and variance 0.12452. The $p-$values for different bandwidths are shown in Figure 4.2. In the horizontal axis, we represent the parameter r from equation (4.26) varying from 2.0 to 20.0.

As it has been commented before, T_P test is less affected by the choice of the bandwidth, and the null hypothesis that the data admit a SAR model fit is accepted. T_{LK} test accepts the null hypothesis, for a significance level $\alpha = 0.05$, in most part of the bandwidth range, as it is shown in Figure 4.2. In particular, the null hypothesis is accepted for the cross-validation bandwidth.

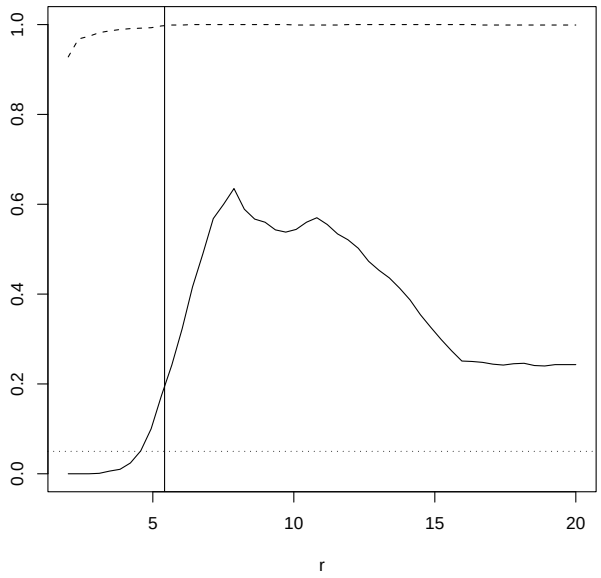

Figure 4.2: p-values for testing $SAR(1)$ model. Solid line: T_{LK} test $p-$values. Dashed line: T_P test $p-$values. Dotted line: significance level 0.05. Vertical solid line: cross-validation bandwidth.

4.5.2 Heavy metal concentrations.

We have also considered Selenium concentrations (from measurements taken in March, 2004). In this set, the independence hypothesis is rejected, by both T_{LK} and T_P along the complete bandwidth range. We have also applied these two tests in order to check whether model (3.23) or (4.27) fit the data. For BAR model, the following estimations for the parameters in (2.92) were obtained: $\hat{\sigma}^2 = 2658.20$, $\hat{\beta}_1 = 0.369$ and $\hat{\beta}_2 = 0.399$. For SAR model (4.27) we get: $\hat{\sigma}^2 = 2350.16$, $\hat{\alpha}_1 = 0.160$ and $\hat{\alpha}_2 = 0.175$.

In Figure 4.4, we show the p-values for T_{LK} and T_P when testing BAR and SAR models. Both hypothesis are accepted, so both spatial spectral densities (3.23) or (4.27) could explain the dependence structure of the data. In the horizontal axis, we represent the parameter r from equation (4.26) varying from 2.0 to 10.0. In Figure 4.3 we show the plots for the periodogram and the adjusted parametric models.

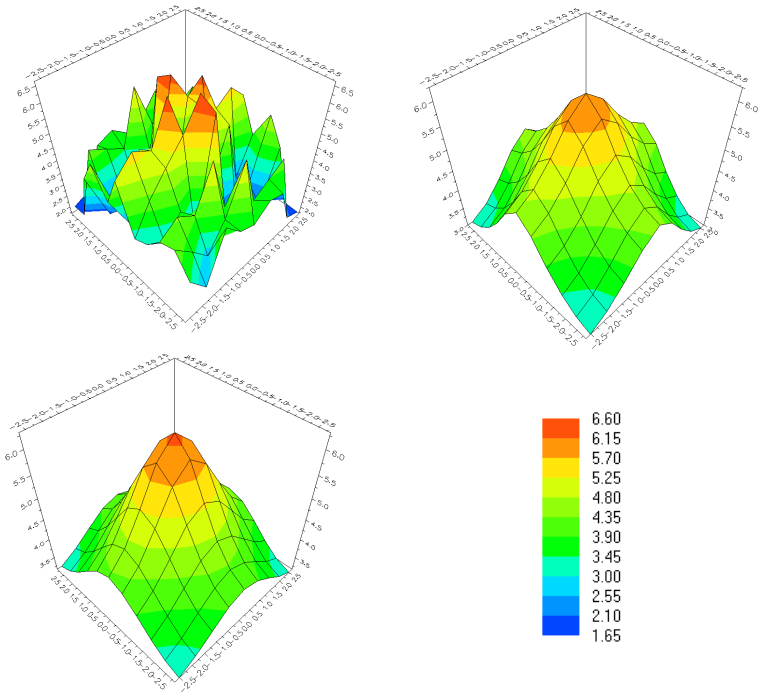

Figure 4.3: Spatial spectral density estimation for $\log(Se)$, March 2004. From top-left to right-bottom: periodogram, BAR and SAR models.

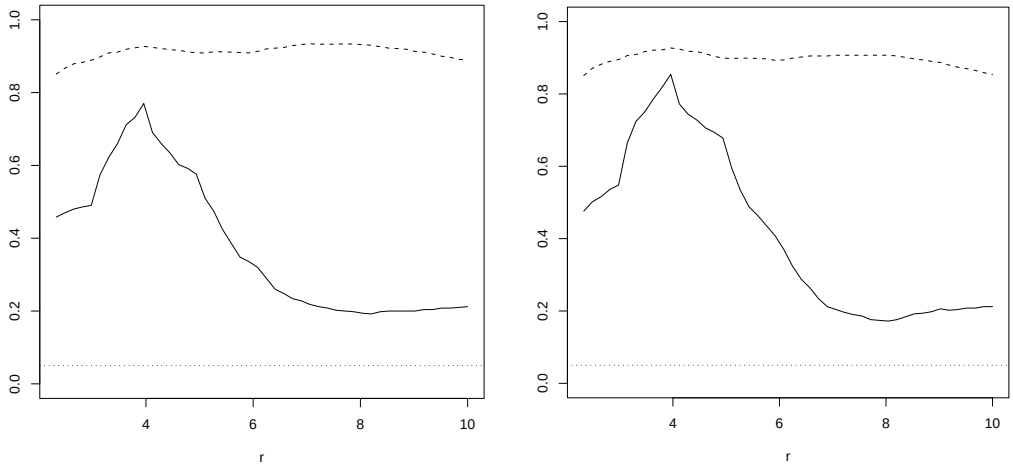

Figure 4.4: Left panel: p-values of T_{LK} and T_P for $\log(Se)$ concentrations in March 2004. Right panel: p-values. Solid line: T_{LK} test p-values. Dashed line: T_P test p-values. Dotted line: significance level 0.05. r denotes the scaling parameter in (4.26).

4.6 Appendix Chapter 4.

4.6.1 Proof of Theorem 1.

In order to prove Theorem 1, we must introduce some lemmas. Lemma 10 gives a decomposition of the T_P statistic as a sum of the test statistic when considering a simple null hypothesis plus a negligible term. Lemma 9 gives the asymptotic distribution of the T_P statistic, under $H_0 : \theta = \theta_0$. Lemmas 5 to 9 provide some tools which are needed in Lemma 10 and 11.

Lemma 5. *Assume that $\{\theta_N\}$ is a sequence of estimators of $\theta_0 \in \Theta \in \mathbb{R}^p$ such that $\sqrt{N}(\theta_N - \theta_0) = \mathcal{O}_{\mathbb{P}}(1)$. Assume that the spectral density f_{θ_0} is continuously differentiable w.r.t. θ with bounded derivatives in $\boldsymbol{\lambda} \in \Pi^2 = [-\pi, \pi] \times [-\pi, \pi]$. Then, under the assumptions in Theorem 1:*

$$\sup_{\boldsymbol{\lambda} \in \Pi^2} \left| \frac{f_{\theta_N}(\boldsymbol{\lambda}) - f_{\theta_0}(\boldsymbol{\lambda})}{f_{\theta_N}(\boldsymbol{\lambda})} \right| = \mathcal{O}_{\mathbb{P}}(N^{-1/2}). \tag{4.29}$$

Proof. Since for any θ_N, the estimated spectral density f_{θ_N} is continuous in Π^2, then

$$\sup_{\boldsymbol{\lambda} \in \Pi^2} \left| \frac{1}{f_{\theta_N}(\boldsymbol{\lambda})} \right| = \mathcal{O}_{\mathbb{P}}(1).$$

Besides, since $\sqrt{N}(\theta_N - \theta_0) = \mathcal{O}_{\mathbb{P}}(1)$, it implies that the difference between the estimator θ_N and the parameter θ_0 can be stochastically bounded by: $\theta_N - \theta_0 = \mathcal{O}_{\mathbb{P}}(N^{-1/2})$. For a fixed $\boldsymbol{\lambda}$, using a Taylor expansion of f_{θ_N} around f_θ and considering the Lagrange remainder, we have:

$$f_{\theta_N}(\boldsymbol{\lambda}) = f_{\theta_0}(\boldsymbol{\lambda}) + (\theta_N - \theta_0)^T \nabla f_{\tilde{\theta}}(\boldsymbol{\lambda}) \leq f_{\theta_0}(\boldsymbol{\lambda}) + \sum_{i=1}^{p} |\theta_N^i - \theta_0^i| \sup_{\boldsymbol{\lambda} \in \Pi^2} \left| \frac{\partial}{\partial \theta_i} f_{\tilde{\theta}}(\boldsymbol{\lambda}) \right|$$

for some $\tilde{\theta}$ with $\|\tilde{\theta} - \theta_0\| \leq \|\theta_N - \theta_0\|$. Therefore,

$$\sup_{\boldsymbol{\lambda} \in \Pi^2} |f_{\theta_N}(\boldsymbol{\lambda}) - f_{\theta_0}(\boldsymbol{\lambda})| \leq \sum_{i=1}^{p} |\theta_N^i - \theta_0^i| \sup_{\boldsymbol{\lambda} \in \Pi^2} \left| \frac{\partial}{\partial \theta_i} f_{\tilde{\theta}}(\boldsymbol{\lambda}) \right| = \mathcal{O}_{\mathbb{P}}(N^{-1/2}). \tag{4.30}$$

The result is proved combining equations (4.29) and (4.30). □

Lemma 6. *Consider Z a spatial process with representation (2.10) and suppose that assumption (1) holds. Then:*

$$\max_{\boldsymbol{\lambda} \in \Pi^2} E(R_n^4(\boldsymbol{\lambda})) = \mathcal{O}(N^{-2}), \tag{4.31}$$

$$\max_{\mathbf{k}} |R_n(\boldsymbol{\lambda}_{\mathbf{k}})| = \mathcal{O}_{\mathbb{P}}(N^{-1/2} \log N). \tag{4.32}$$

Proof. In order to prove (4.31), the residual term $R_n(\boldsymbol{\lambda})$ can be written as:

$$R_n(\boldsymbol{\lambda}) = A(\boldsymbol{\lambda})J_\varepsilon(\boldsymbol{\lambda})Y_n(-\boldsymbol{\lambda}) + A(-\boldsymbol{\lambda})J_\varepsilon(-\boldsymbol{\lambda})Y_n(\boldsymbol{\lambda}) + |Y_n(\boldsymbol{\lambda})|^2, \tag{4.33}$$

where,

$$A(\boldsymbol{\lambda}) = \sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} a_{jl}e^{-i(j,l)\boldsymbol{\lambda}}, \tag{4.34}$$

$$J_\varepsilon(\boldsymbol{\lambda}) = \frac{1}{2\pi\sqrt{N}}\sum_{s_1=0}^{n_1-1}\sum_{s_2=0}^{n_2-1}\varepsilon(\mathbf{s})e^{-i\mathbf{s}^T\boldsymbol{\lambda}}, \tag{4.35}$$

$$U_{n,j,l}(\boldsymbol{\lambda}) = \frac{1}{2\pi\sqrt{N}}\left\{\sum_{s_1=-j}^{n_1-1-j}\sum_{s_2=-l}^{n_2-1-l} e^{-i\mathbf{s}^T\boldsymbol{\lambda}}\varepsilon(\mathbf{s}) - \sum_{s_1=0}^{n_1-1}\sum_{s_2=0}^{n_2-1} e^{-i\mathbf{s}^T\boldsymbol{\lambda}}\varepsilon(\mathbf{s})\right\}, \tag{4.36}$$

and finally

$$Y_n(\boldsymbol{\lambda}) = \sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} a_{jl}e^{-i(j,l)\boldsymbol{\lambda}}U_{n,j,l}(\boldsymbol{\lambda}), \tag{4.37}$$

just following similar arguments to those in (Brockwell and Davis (1991)). Therefore, taking expectations on the fourth order moment:

$$\begin{aligned} E\left(R_n^4(\boldsymbol{\lambda})\right) &\leq k_1 E\left(|A(\boldsymbol{\lambda})J_\varepsilon(\boldsymbol{\lambda})Y_n(-\boldsymbol{\lambda})|^4\right) \\ &+k_2 E\left(|A(-\boldsymbol{\lambda})J_\varepsilon(-\boldsymbol{\lambda})Y_n(\boldsymbol{\lambda})|^4\right) + k_3 E(|Y_n(\boldsymbol{\lambda})|^8), \end{aligned}$$

for some positive constants k_1, $i = 1, 2, 3$. For the first term on the right hand side using Caucy-Schwarz inequality:

$$\begin{aligned} &E\left(|A(\boldsymbol{\lambda})J_\varepsilon(\boldsymbol{\lambda})Y_n(-\boldsymbol{\lambda})|^4\right) \leq |A(\boldsymbol{\lambda})|^4\left(E|J_\varepsilon(\boldsymbol{\lambda})|^8\right)^{1/2}\left(E|Y_n(-\boldsymbol{\lambda})|^8\right)^{1/2} \\ &= \mathcal{O}(1)\sqrt{E(|Y_n(-\boldsymbol{\lambda})|^8)}. \end{aligned}$$

For $E|Y_n(-\boldsymbol{\lambda})|^8$, we can get a bound taking into account that, if $|j| < n_1$ and $|l| < n_2$, $2\pi\sqrt{N}U_{n,j,l}$ is a sum of $4|j||l|$ independent indentically distributed (i.i.d.) random variables. For $|j| \geq n_1$, $|l| \geq n_2$, it is a sum of $4n_1n_2$ iid random variables. In the case $|j| < n_1$, $|l| \geq n_2$, it is a sum of $4|j|n_2$ iid random variables, whereas if $|j| \geq n_1$, $|l| < n_2$, it is a sum of $4|l|n_1$ iid random variables. Then, using the inequality:

$$\begin{aligned} &E\left(\sum_{j=1}^{n} Z_j\right)^8 \leq \\ &nEZ_1^8 + 28n^2EZ_1^6EZ_1^2 + 35n^2(EZ_1^4)^2 + 210n^3EZ_1^4(EZ_1^2)^2 + 105(EZ_1^2)^4 \end{aligned}$$

where Z_j are independent identically distributed random variables, with zero mean and finite eight-order moment, we have:

$$E\,|U_{n,j,l}(\boldsymbol{\lambda})|^8 \leq c_1|j||l|E(\varepsilon^8) + c_2|j|^2|l|^2E(\varepsilon^6)E(\varepsilon^2) + c_3|j|^2|l|^2E^2(\varepsilon^4)$$

$$+c_4|j|^3|l|^3E(\varepsilon^4)E^2(\varepsilon^2)+c_5|j|^4|l|^4E^4(\varepsilon^2).$$

By assumption (1), concercing the summability of $\{|j||l|a_{j,l}\}$ and Jensen's inequality, we get $E|Y_n(\boldsymbol{\lambda})|^8 \leq \mathcal{O}(N^{-4})$:

$$E(|Y_n(\boldsymbol{\lambda})|^8) = E\left(\left|\sum_{j,l} a_{j,l}e^{-i(j,l)\boldsymbol{\lambda}}U_{n,j,l}(\boldsymbol{\lambda})\right|^8\right) \leq$$

$$c_6\sum_{j,l} a_{j,l}\left(E\left(|U_{n,j,l}(\boldsymbol{\lambda})|^8\right)\right)$$

$$c_6\left(\frac{1}{N^4}\sum_{j,l}|a_{j,l}|c_1|j||l|E(\varepsilon(\mathbf{s})^8)+\frac{1}{N^4}\sum_{j,l}|a_{j,l}|c_2|j|^2|l|^2E(\varepsilon(\mathbf{s})^6)E(\varepsilon(\mathbf{s})^2)\right.$$

$$+\frac{1}{N^4}\sum_{j,l}|a_{j,l}|c_3|j|^2|l|^2E^2(\varepsilon(\mathbf{s})^4)+\frac{1}{N^4}\sum_{j,l}|a_{j,l}|c_4|j|^3|l|^3E(\varepsilon(\mathbf{s})^4)E^2(\varepsilon(\mathbf{s})^2)$$

$$\left.+\frac{1}{N^4}\sum_{j,l}|a_{j,l}|c_5|j|^4|l|^4E^4(\varepsilon(\mathbf{s})^2)\right) = \mathcal{O}(N^{-4}),$$

and from the expression above, we obtain that $E^{1/2}(|Y_n(\boldsymbol{\lambda})|^8) = \mathcal{O}(N^{-2})$.

The bound for (4.32) ca be obtained by a straightforward extension of the arguments in (Kooperberg *et al.* (1995)).

Let's prove now (4.32). Consider the expression of $J_\varepsilon(\lambda)$ given by (4.35) and split it in its real and imaginary parts. The real part of $J_\varepsilon(\boldsymbol{\lambda})$ is distributed as:

$$Re(J_\varepsilon(\boldsymbol{\lambda})) \sim N\left(0, \frac{A^TA\sigma^2}{(2\pi)^2N}\right),$$

where A is given by

$$A = \begin{pmatrix} 1 \\ \cos((1,0)\boldsymbol{\lambda}) \\ \vdots \\ \cos((1,n_2-1)\boldsymbol{\lambda}) \\ \cos((2,1)\boldsymbol{\lambda}) \\ \vdots \\ \cos((2,n_2-1)\boldsymbol{\lambda}) \\ \vdots \\ \cos((n_1-1,1)\boldsymbol{\lambda}) \\ \vdots \\ \cos((n_1-1,n_2-1)\boldsymbol{\lambda}) \end{pmatrix}.$$

We prove that the real part is $O_{\mathbb{P}}(\sqrt{\log N})$, where $N = n_1 \cdot n_2$. For that purpose, let $\nu \in \mathbb{R}$. We will prove that:

$$\mathbb{P}\left(\frac{1}{2\pi\sqrt{N}}\sum_{\mathbf{s}}\cos(\boldsymbol{\lambda}_{\mathbf{k}}^T\mathbf{s})\varepsilon(\mathbf{s}) \geq \nu\sqrt{\log N}\right) \to 0. \tag{4.38}$$

First, considering the distribution of $Re(J_\varepsilon(\boldsymbol{\lambda}))$, we can write:

$$\mathbb{P}\left(\frac{1}{2\pi\sqrt{N}}\sum_{\mathbf{s}}\cos(\boldsymbol{\lambda}_{\mathbf{k}}^T\mathbf{s})\varepsilon(\mathbf{s}) \geq \nu\sqrt{\log N}\right) = \sqrt{\frac{2\pi N}{A^TA\sigma^2}}\int_{\nu\sqrt{\log N}}^{\infty} e^{\left(-\frac{2\pi^2 N x^2}{A^TA\sigma^2}\right)}dx. \tag{4.39}$$

Applying a change of variable:

$$\frac{2\pi^2 N x^2}{A^TA\sigma^2} = \frac{y^2}{2},$$

we rewrite (4.39) as:

$$\mathbb{P}\left(\frac{1}{2\pi\sqrt{N}}\sum_{\mathbf{s}}\cos(\boldsymbol{\lambda}_{\mathbf{k}}^T\mathbf{s})\varepsilon(\mathbf{s}) \geq \nu\sqrt{\log N}\right) = \frac{1}{\sqrt{2\pi}}\int_{\frac{2\pi\nu}{\sigma\sqrt{A^TA}}\sqrt{N\log N}}^{\infty} e^{-\frac{y^2}{2}}dy. \tag{4.40}$$

Since the following exponential inequality holds:

$$\int_y^\infty \frac{e^{-x^2/2}}{e^{-y^2/2}}dx \leq \int_y^\infty \frac{x}{y}e^{(-x^2/2}dx = \frac{1}{y}\int_{-y^2/2}^\infty e^{-u}du = \frac{1}{y}e^{-y^2/2},$$

that is

$$\int_y^\infty e^{-x^2/2}dx \leq \frac{1}{y}e^{-y^2/2}, \quad y > 0,$$

expression (4.40) can be bounded by:

$$\frac{1}{\sqrt{2\pi}}\int_{\frac{2\pi\nu}{\sigma\sqrt{A^TA}}\sqrt{N\log N}}^{\infty} e^{-\frac{y^2}{2}}dy \leq \frac{\sigma\sqrt{A^TA}}{(2\pi)^{3/2}\nu\sqrt{N\log N}}e^{-\frac{1}{2}\left(\frac{2\pi\nu\sqrt{N\log N}}{\sigma\sqrt{A^TA}}\right)^2}, \tag{4.41}$$

and since $A^TA = \mathcal{O}(N)$, it is easy to see that the right hand side in (4.41) tends to zero. Then, (4.38) is proved. The same result hold for the imaginary part of $J_\varepsilon(\boldsymbol{\lambda})$.

We find a (uniform) bound for $Y_n(\boldsymbol{\lambda})$. We can write the expression as:

$$Y_n(\boldsymbol{\lambda}) = \sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} a_{jl}\exp(-i\boldsymbol{\lambda}^T(j,l))\frac{1}{2\pi\sqrt{N}}\left[\sum_{s_1=-j}^{n_1-1-j}\sum_{s_2=-l}^{n_2-1-l} e^{(-i(j,l)^T\boldsymbol{\lambda})\varepsilon(\mathbf{s})}\right.$$
$$\left. - \sum_{s_1=0}^{n_1-1}\sum_{s_2=0}^{n_2-1} e^{(-i(j,l)^T\boldsymbol{\lambda})\varepsilon(\mathbf{s})}\right]$$

Decomposing each addend in real and imaginary part and taking as an example just the one dealing with the cosines (since the same procedure can be applied to the other addends), we have:

$$\frac{1}{2\pi\sqrt{N}}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{s_1=-j}^{n_1-1-j}\sum_{s_2=-l}^{n_2-1-l}a_{jl}\cos(\boldsymbol{\lambda}^T(j+s_1,l+s_2))\varepsilon(\mathbf{s}) =$$
$$\frac{1}{2\pi\sqrt{N}}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{p_1=0}^{n_1-1}\sum_{p_2=0}^{n_2-1}a_{jl}\cos(\boldsymbol{\lambda}^T\mathbf{p})\varepsilon(\mathbf{p}-(j,l)),\quad \mathbf{p}^T=(p_1,p_2).$$

We will see that this term is an $O_{\mathbb{P}}\left(\sqrt{\frac{\log N}{N}}\right)$:

$$\mathbb{P}\left(\frac{1}{2\pi\sqrt{N}}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{p_1=0}^{n_1-1}\sum_{p_2=0}^{n_2-1}a_{jl}\cos(\boldsymbol{\lambda}^T\mathbf{p})\varepsilon(\mathbf{p}-(j,l))\geq\nu\sqrt{\frac{\log N}{N}}\right)=$$
$$\mathbb{P}\left(\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}a_{jl}N\left(0,\frac{\sigma^2}{(2\pi)^2N}\right)\geq\nu\sqrt{\frac{\log N}{N}}\right)\leq$$
$$\mathbb{P}\left(N\left(0,\frac{\sigma^2}{(2\pi)^2N}\right)\geq\frac{\nu}{S_A}\sqrt{\frac{\log N}{N}}\right)=\int_{\frac{\nu}{S_A}\sqrt{\frac{\log N}{N}}}^{\infty}\frac{\sqrt{(2\pi)^2N}}{\sigma}e^{\left(-\frac{(2\pi)^2Nx^2}{2\sigma^2}\right)}dx=$$
$$\int_{\frac{\nu 2\pi\sqrt{N}}{\nu S_A}\sqrt{\frac{\log N}{N}}}^{\infty}\frac{\sqrt{(2\pi)^2N}}{\sigma}\frac{\sigma}{2\pi\sqrt{N}}e^{-\frac{y^2}{2}}dy=\frac{1}{\sqrt{2\pi}}\int_{\frac{2\pi\nu\sqrt{\log N}}{S_A\sigma}}^{\sigma}e^{\left(-\frac{y^2}{2}\right)}dy\leq$$
$$\frac{1}{\sqrt{2\pi}}\frac{1}{\frac{2\pi\nu}{S_A\sigma\sqrt{\log N}}}e^{\left(-\frac{(2\pi)^2\nu^2\log n}{S_A^2\sigma^2}\right)}=\frac{S_A\sigma}{(2\pi)^{3/2}\nu\sqrt{\log N}}e^{\left(-\frac{(2\pi)^2\nu^2\log N}{S_A^1\sigma^2}\right)}=$$
$$\frac{S_A\sigma}{(2\pi)^{3/2}\nu\sqrt{\log N}}e^{\left(-\frac{(2\pi)^2\nu^2}{S_A^2\sigma^2}\log N\right)}=M\frac{e^{-p\log N}}{\sqrt{N}}\to 0$$

The constants involved in the proof are given by:

$$M=\frac{S_A\sigma}{(2\pi)^{3/2}\nu\sqrt{\log N}},\quad p=\frac{(2\pi)^2\nu^2}{2S_A^2\sigma^2}\quad\text{and}\quad S_A=\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}a_{jl}.$$

So, for all $\boldsymbol{\lambda}_{\mathbf{k}}$, we have the following stochastic convergence rates:

$$J_\varepsilon(\boldsymbol{\lambda}_{\mathbf{k}})=O_{\mathbb{P}}(\sqrt{\log N})\quad\text{and}\quad Y_n(\boldsymbol{\lambda}_{\mathbf{k}})=O_{\mathbb{P}}\left(\sqrt{\frac{\log N}{N}}\right).$$

Then, for the residual term $R_n(\boldsymbol{\lambda}_{\mathbf{k}})$, we obtain that:

$$R_n(\boldsymbol{\lambda}_{\mathbf{k}})=O_{\mathbb{P}}\left(\frac{\log N}{\sqrt{N}}\right).$$

□

Lemma 7. *Under assumptions (1) and (2), as* $N \to \infty$

$$\frac{|H|^{1/4}}{N} \int_{\Pi^2} \left(\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) \frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} \right)^2 d\boldsymbol{\lambda} \to 0 \text{ in probability.}$$

Proof. The proof of this lemma can be done by similar arguments to those in the proof of Lemma 5 in Paparoditis (2000), with bidimensional kernel function K and bandwidth matrix H.

We have that, from Lemma 6 and using the Cauchy-Schwarz inequality:

$$\begin{aligned}
&E(R_n(\boldsymbol{\lambda}_{\mathbf{k}}) R_n(\boldsymbol{\lambda}_{\mathbf{j}}) R_n(\boldsymbol{\lambda}_{\mathbf{i}}) R_n(\boldsymbol{\lambda}_{\mathbf{m}})) \leq \\
&\left\{ E(R_n^2(\boldsymbol{\lambda}_{\mathbf{k}}) R_n^2(\boldsymbol{\lambda}_{\mathbf{j}})) \right\}^{1/2} \left\{ E(R_n^2(\boldsymbol{\lambda}_{\mathbf{i}}) R_n^2(\boldsymbol{\lambda}_{\mathbf{m}})) \right\}^{1/2} \leq \\
&\left\{ E(R_n^4(\boldsymbol{\lambda}_{\mathbf{k}})) E(R_n^4(\boldsymbol{\lambda}_{\mathbf{j}})) E(R_n^4(\boldsymbol{\lambda}_{\mathbf{i}})) E(R_n^4(\boldsymbol{\lambda}_{\mathbf{m}})) \right\}^{1/4} \leq \\
&\mathcal{O}(N^{-2}).
\end{aligned}$$

We prove that

$$\frac{|H|^{1/4}}{N} \int_{\Pi^2} \left(\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) \frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} \right)^2 d\boldsymbol{\lambda}$$

tends to zero in L^2 norm.

$$\begin{aligned}
&0 \leq E \left(\frac{|H|^{1/4}}{N} \int_{\Pi^2} \left(\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) \frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} \right)^2 d\boldsymbol{\lambda} \right)^2 = \\
&\frac{|H|^{1/2}}{N^2} E \left(\int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) \frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})} d\boldsymbol{\lambda} \right)^2 = \\
&\frac{|H|^{1/2}}{N^2} E \left(\sum_{\mathbf{k}} \sum_{\mathbf{j}} \int_{\Pi^2} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) d\boldsymbol{\lambda} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})} \right)^2 = \\
&\frac{|H|^{1/2}}{N^2} E \left(\left[\sum_{\mathbf{k}} \sum_{\mathbf{j}} \int_{\Pi^2} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) d\boldsymbol{\lambda} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})} \right] \right. \\
&\left. \left[\sum_{\mathbf{l}} \sum_{\mathbf{m}} \int_{\Pi^2} K_H(\boldsymbol{\omega} - \boldsymbol{\lambda}_{\mathbf{l}}) K_H(\boldsymbol{\omega} - \boldsymbol{\lambda}_{\mathbf{m}}) d\boldsymbol{\omega} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{l}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{l}})} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{m}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{m}})} \right] \right) \leq \\
&\frac{|H|^{1/2}}{N^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} \int_{\Pi^2} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) d\boldsymbol{\lambda} \cdot \\
&\sum_{\mathbf{l}} \sum_{\mathbf{m}} \int_{\Pi^2} K_H(\boldsymbol{\omega} - \boldsymbol{\lambda}_{\mathbf{l}}) K_H(\boldsymbol{\omega} - \boldsymbol{\lambda}_{\mathbf{m}}) d\boldsymbol{\omega} \\
&\frac{1}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}}) f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}}) f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{l}}) f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{m}})} E |R_n(\boldsymbol{\lambda}_{\mathbf{k}}) R_n(\boldsymbol{\lambda}_{\mathbf{j}}) R_n(\boldsymbol{\lambda}_{\mathbf{l}}) R_n(\boldsymbol{\lambda}_{\mathbf{m}})| = \mathcal{O}(|H|^{1/2}),
\end{aligned}$$

where the last equality follows from the fact that

$$\int_{\Pi^2} K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})d\boldsymbol{\lambda} < \infty.$$

□

Lemma 8. *Consider Z a spatial process with spectral density f and denote $W_{\mathbf{k}} = V_{\mathbf{k}} - 1$, where $V_{\mathbf{k}} \sim Exp(1)$, independent random variables. Under assumptions (1)-(2),*

$$\frac{|H|^{1/4}}{N}\int_{\Pi^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}} K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})W_{\mathbf{k}}\frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})}d\boldsymbol{\lambda} \to 0$$

in probability.

Proof. Consider the following notation, in order to make the proof more brief:

$$K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}) = K_H^{\mathbf{k}}(\boldsymbol{\lambda}), \quad f(\boldsymbol{\lambda}_{\mathbf{k}}) = f_{\mathbf{k}} \quad \text{and} \quad R_n(\boldsymbol{\lambda}_{\mathbf{k}}) = R_n^{\mathbf{k}}.$$

We will prove L^2-consistency:

$$E\left(\frac{|H|^{1/4}}{N}\int_{\Pi^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}} K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})(V_{\mathbf{k}}-1)\frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})}d\boldsymbol{\lambda}\right)^2 \tag{4.42}$$

$$= \frac{|H|^{1/2}}{N^2}\int\int\sum_{\mathbf{k},\mathbf{j},\mathbf{l},\mathbf{m}} K_H^{\mathbf{k}}(\boldsymbol{\lambda})K_H^{\mathbf{j}}(\boldsymbol{\lambda})K_H^{\mathbf{l}}(\boldsymbol{\omega})K_H^{\mathbf{m}}(\boldsymbol{\omega})\frac{1}{f_{\mathbf{k}}f_{\mathbf{j}}}E\left(W_{\mathbf{k}}R_n^{\mathbf{j}}W_{\mathbf{l}}R_n^{\mathbf{m}}\right)d\boldsymbol{\lambda}d\boldsymbol{\omega} \tag{4.43}$$

In order to find a bound for this term, consider that $\mathbf{k}$, $\mathbf{j}$, $\mathbf{l}$ and $\mathbf{m}$ are all different indexes. From Theorem 2.3.2 in (Brillinger (1981)):

$$\begin{aligned}
&E(W_{\mathbf{k}}R_n^{\mathbf{j}}W_{\mathbf{l}}R_n^{\mathbf{m}}) = \\
&\text{cum}(W_{\mathbf{k}}R_n^{\mathbf{j}})\text{cum}(W_{\mathbf{l}}R_n^{\mathbf{m}}) + \text{cum}(W_{\mathbf{k}}R_n^{\mathbf{m}})\text{cum}(W_{\mathbf{l}}R_n^{\mathbf{j}}) \\
&+\text{cum}(R_n^{\mathbf{j}})\text{cum}(W_{\mathbf{k}}W_{\mathbf{l}}R_n^{\mathbf{m}}) + \text{cum}(R_n^{\mathbf{l}})\text{cum}(W_{\mathbf{k}}W_{\mathbf{l}}R_n^{\mathbf{j}}) \\
&+\text{cum}(W_{\mathbf{k}})\text{cum}(R_n^{\mathbf{j}}W_{\mathbf{l}}R_n^{\mathbf{m}}) + \text{cum}(W_{\mathbf{l}})\text{cum}(W_{\mathbf{k}}R_n^{\mathbf{j}}R_n^{\mathbf{m}}).
\end{aligned}$$

Since $\text{cum}(W_{\mathbf{k}}) = E(W_{\mathbf{k}}) = 0$ and $\text{cum}(W_{\mathbf{k}}W_{\mathbf{l}}) = E(W_{\mathbf{k}}W_{\mathbf{l}}) = 0$, and applying Theorem 2.3.2 of (Brillinger (1981)) on the three term cumulants, the expression above can be simplified:

$$E(W_{\mathbf{k}}R_n^{\mathbf{j}}W_{\mathbf{l}}R_n^{\mathbf{m}}) = E(W_{\mathbf{k}}R_n^{\mathbf{j}})E(W_{\mathbf{k}}R_n^{\mathbf{m}}) + E(W_{\mathbf{l}}R_n^{\mathbf{j}})E(W_{\mathbf{l}}R_n^{\mathbf{m}}) = \mathcal{O}(N^{-2}),$$

where the last equality is obtained recalling the expression for $R_n^{\mathbf{j}}$ in (4.33), and from a straightforward extension of Lemma 2 in Paparoditis (2000). Then, (4.43) is $\mathcal{O}(|H|^{1/2})$.
Consider the case $\mathbf{k} = \mathbf{j} \neq \mathbf{l} = \mathbf{m}$. By the Cauchy-Schwarz inequality and Lemma 6,

$$|E(W_{\mathbf{k}}R_n^{\mathbf{j}}W_{\mathbf{l}}R_n^{\mathbf{m}})| \leq \sqrt{E(W_{\mathbf{k}}R_n^{\mathbf{j}})^2E(W_{\mathbf{l}}R_n^{\mathbf{m}})^2} \leq \mathcal{O}(N^{-1}).$$

Then, (4.43) is $\mathcal{O}(N^{-1}|H|^{1/2})$. For the case $\mathbf{k} \neq \mathbf{j} \neq \mathbf{l} = \mathbf{m}$, using the same arguments, (4.43) is also $\mathcal{O}(N^{-1}|H|^{1/2})$. □

Lemma 9. *Assume that assumption (2) is fulfilled and consider $U_{\mathbf{k}}$ independent identically distributed random variables with $E(U_{\mathbf{k}}) = 1$, $Var(U_{\mathbf{k}}) = 1$ and $E(U_{\mathbf{k}}^4) < \infty$. Then,*

$$\frac{|H|^{1/4}}{N} \int_{\Pi^2} \left(\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})(U_{\mathbf{k}} - 1) \right)^2 d\boldsymbol{\lambda} - \mu_H \to N(0, \tau^2),$$

where μ_H and τ^2 are given in (4.11) and (4.12), respectively and the sum $\sum_{\mathbf{k}}$ extends over the set of Fourier frequencies.

Proof. Let $Z_{\mathbf{k}} = U_{\mathbf{k}} - 1$.

$$\begin{aligned}
&\frac{|H|^{1/4}}{N} \int_{\Pi^2} \left(\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})(U_{\mathbf{k}} - 1) \right)^2 d\boldsymbol{\lambda} - \mu_H = \\
&\frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k}} K_H^2(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) Z_{\mathbf{k}}^2 d\boldsymbol{\lambda} \\
&-|H|^{1/4} \int_{\Pi^2} K^2(\mathbf{u}) d\mathbf{u} + \frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k} \neq \mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) Z_{\mathbf{k}} Z_{\mathbf{j}} d\boldsymbol{\lambda} \\
&= T_1 - \mu_H + T_2.
\end{aligned}$$

Note that, as $N \to \infty$:

$$\begin{aligned}
&|E(T_1) - \mu_H| = \\
&|H|^{-1/4} \left| \int_{\Pi^2} \frac{1}{N|H|^{1/2}} \sum_{\mathbf{k}} K^2(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})) d\boldsymbol{\lambda} - \int_{\Pi^2} K^2(\mathbf{u}) d\mathbf{u} \right| \to 0.
\end{aligned}$$

For the variance of this first term T_1, since the $Z_{\mathbf{k}}$ are independent zero-mean variables:

$$Var(T_1) = Var\left(|H|^{1/4} \int_{\Pi^2} \frac{1}{N|H|^{1/2}} \sum_{\mathbf{k}} K^2(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})) Z_{\mathbf{k}} d\boldsymbol{\lambda} \right)$$

$$= \frac{1}{N^2|H|^{3/2}} \left(\sum_{\mathbf{k}} \int_{\Pi^2} K^2(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})) d\boldsymbol{\lambda} \right)^2 Var(Z_{\mathbf{0}}^2) = \mathcal{O}\left(N^{-1}|H|^{-1/2} \right).$$

using the same arguments as above. Let's analyze T_2. Define, for $\mathbf{j} \neq \mathbf{k}$

$$a(\mathbf{k}, \mathbf{j}) = a(k_1, k_2, j_1, j_2) = \frac{|H|^{1/4}}{N} \int_{\Pi^2} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) d\boldsymbol{\lambda}$$

and define $a(\mathbf{k}, \mathbf{k}) = 0$; then, T_2 can be decomposed as follows:

$$T_2 = \frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k} \neq \mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) Z_{\mathbf{k}} Z_{\mathbf{j}} d\boldsymbol{\lambda} = \sum_{\mathbf{k}} \sum_{\mathbf{j}} a(\mathbf{k}, \mathbf{j}) Z_{\mathbf{k}} Z_{\mathbf{j}}.$$

Define $b(\mathbf{k},\mathbf{j}) = b(k_1,k_2,j_1,j_2)$ as:

$$b(k_1,k_2,j_1,j_2) = a(k_1,k_2,j_1,j_2) + a(k_1,-k_2,j_1,j_2) + a(k_1,k_2,-j_1,j_2) + a(k_1,k_2,j_1,-j_2) +$$
$$a(-k_1,k_2,j_1,j_2) + a(-k_1,-k_2,j_1,j_2) + a(-k_1,k_2,-j_1,j_2) + a(-k_1,k_2,j_1,-j_2) +$$
$$a(-k_1,-k_2,-j_1,j_2) + a(-k_1,-k_2,j_1,-j_2) + a(-k_1,-k_2,-j_1,j_2) + a(k_1,-k_2,-j_1,-j_2) +$$
$$a(k_1,-k_2,-j_1,-j_2) + a(k_1,k_2,-j_1,-j_2) + a(-k_1,k_2,-j_1,-j_2) + a(-k_1,-k_2,-j_1,-j_2).$$

Then, T_2 can be written as $T_2 = Q_N + T_3$ where

$$Q_N = \sum_{k_1=1}^{m_1}\sum_{k_2=1}^{m_2}\sum_{j_1=1}^{m_1}\sum_{j_2=1}^{m_2} b(\mathbf{k},\mathbf{j}) Z_{\mathbf{k}} Z_{\mathbf{j}}$$

and

$$T_3 = \sum_{k_2=-m_2}^{m_2}\sum_{j_1=-m_1}^{m_1}\sum_{j_2=-m_2}^{m_2} a(k_1,k_2,j_1,j_2) Z_{\mathbf{k}} Z_{\mathbf{j}} \delta_{\mathbf{kj}},$$

where the function

$$\delta_{\mathbf{kj}} = \begin{cases} 1 & \text{if} \quad 0 < \mathbf{1}_{(k_1=0)} + \mathbf{1}_{(k_2=0)} + \mathbf{1}_{(j_1=0)} + \mathbf{1}_{(j_2=0)}, \\ 0 & \text{otherwise,} \end{cases}$$

and $\mathbf{1}$ is the indicator function. Consider any of the addends in the expression of T_3, for instance:

$$\sum_{k_2=-m_2}^{m_2}\sum_{j_1=-m_1}^{m_1}\sum_{j_2=-m_2}^{m_2} a(0,k_2,j_1,j_2) Z_{\mathbf{k}} Z_{\mathbf{j}}.$$

Since the $Z_{\mathbf{k}}$ are independent zero-mean random variables, in order to obtain a non-null expectation term, $\mathbf{k}$ must equal to $\mathbf{j}$. For $\mathbf{k}_0 = (0,k_2)$:

$$E\left(\sum_{k_2=-m_2}^{m_2}\sum_{j_1=-m_1}^{m_1}\sum_{j_2=-m_2}^{m_2} a(0,k_2,j_1,j_2) Z_{\mathbf{k}} Z_{\mathbf{j}}\right) =$$

$$\frac{|H|^{1/4}}{N}\sum_{k_2=-m_2}^{m_2}\int_{\Pi^2} K_H^2(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}_0}) d\boldsymbol{\lambda} E(Z_{\mathbf{k}_0}^2) = \mathcal{O}(n_1^{-1}|H|^{-1/4}).$$

Besides:

$$E\left(\sum_{k_2=-m_2}^{m_2}\sum_{j_1=-m_1}^{m_1}\sum_{j_2=-m_2}^{m_2} a(0,k_2,j_1,j_2) Z_{\mathbf{k}} Z_{\mathbf{j}}\right)^2 =$$

$$\frac{|H|^{1/2}}{N^2}\sum_{k_2=-m_2}^{m_2}\sum_{j_1=-m_1}^{m_1}\sum_{j_2=-m_2}^{m_2}\int_{\Pi^2}\int_{\Pi^2} (K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}_0}) K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})$$
$$\cdot K_H(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{k}_0}) K_H(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{j}}) d\boldsymbol{\lambda} d\boldsymbol{\omega})\, E\left(Z_{\mathbf{k}_0} Z_{\mathbf{j}}\right)^2$$

and then

$$E\left(\sum_{k_2=-m_2}^{m_2}\sum_{j_1=-m_1}^{m_1}\sum_{j_2=-m_2}^{m_2} a(0,k_2,j_1,j_2)Z_{\mathbf{k}}Z_{\mathbf{j}}\right)^2 = \mathcal{O}(n_1^{-1}).$$

Analogous expressions are obtained for the other addends. Therefore

$$T_2 = \sum_{k_1=1}^{m_1}\sum_{k_2=1}^{m_2}\sum_{j_1=1}^{m_1}\sum_{j_2=1}^{m_2} b(\mathbf{k},\mathbf{j})Z_{\mathbf{k}}Z_{\mathbf{j}} + o_{\mathbb{P}}(1) = Q_N + o_{\mathbb{P}}(1).$$

In order to prove the asymptotic normal distribution of Q_N, we will apply Theorem 5.2 in de Jong (1987). For that purpose, we must write Q_N as a quadratic form, namely $Q_N = \sum_{i,j} c_{i,j}Z_iZ_j$, where i and j are one-dimensional indexes and Z_i are i.i.d. random variables with zero mean and unit variance.

First, define a new subindex for the Fourier frequencies $\boldsymbol{\lambda}_{\mathbf{k}}$, with $\mathbf{k} = (k_1, k_2)$ and $k_l = 0, \pm 1, \ldots, \pm m_l$, for $l = 1, 2$. Consider $\boldsymbol{\lambda}_{\mathbf{k}} = \boldsymbol{\lambda}_{\mathbf{k}'}$ where $\mathbf{k}' = (k_1', k_2')$, with $k_l' = 1, \ldots, m_l' = 2m_l + 1$, in such a way that $k_l' = k_l + m_l + 1$ for $l = 1, 2$. Let $M = m_1' \times m_2'$ and denote by $\mathcal{M}_{M\times M}$ the space of square matrices with size M, that is, with M rows and M columns.

The new coefficients, with one dimensional indexes, are given by the following matrix:

$$A = (c_{ij}), \quad A \in \mathcal{M}_{M\times M},$$

and each entry of this matrix is defined by $a_{ij} = b_{\mathbf{ij}}$ and $a_{ii} = 0$, where the bidimensional indexes $\mathbf{i}$ determine unidimensional indexes i such that:

$$\mathbf{i} = (i_1, i_2), \quad \text{if } (i_1 - 1)m_2' \leq i \leq i_1 m_2' \quad \text{and } i = (i_1 - 1)m_2' + i_2, \tag{4.44}$$

Now, define the variables:

$$Z_i = Z_{\mathbf{i}}, \quad \text{where } i = (i_1 - 1)m_2' + i_2, \quad i = 1, \ldots, M.$$

With this definitions, Q_N can be written as a quadratic form with one-dimensional indexes:

$$Q_N = \sum_{i,j} c_{i,j}Z_iZ_j.$$

In order to apply Theorem 5.2 (de Jong (1987)) on the quadratic form Q_N, we must prove that, as $N \to \infty$:

1. There exists a sequence $k(n_1, n_2) \to \infty$ such that

$$k(n_1, n_2)^4 \frac{1}{Var(Q_N)} \max_i \sum_j c_{ij}^2 \to 0.$$

Taking into account that n_1 and n_2 tend to infinity at the same rate, it holds that $E(T_3) = \mathcal{O}(n_1^{-1}|H|^{-1/4})$ and $Var(T_3) \leq E(T_3^2) = \mathcal{O}(n_1^{-1})$. Then, applying that $Var(Q_N) = Var(T_2) + Var(T_3) - 2Cov(T_2, T_3)$, the variance of the quadratic form can be approximated by $Var(T_2)$, since $Var(T_3) \leq \mathcal{O}(n_1^{-1})$ and $|Cov(T_2, T_3)| \leq \sqrt{Var(T_2)Var(T_3)} = \mathcal{O}(n_1^{-1})$, using the Cauchy-Schwarz inequality.

We prove that

$$Var(T_2) = Var\left(\sum_{\mathbf{k}}\sum_{\mathbf{j}} a(\mathbf{k},\mathbf{j})Z_{\mathbf{k}}Z_{\mathbf{j}}\right) = \tag{4.45}$$

$$\sum_{\mathbf{k}}\sum_{\mathbf{j}}\sum_{\mathbf{l}}\sum_{\mathbf{m}} a(\mathbf{k},\mathbf{j})a(\mathbf{l},\mathbf{m})E(Z_{\mathbf{k}}Z_{\mathbf{j}}Z_{\mathbf{l}}Z_{\mathbf{m}}) \tag{4.46}$$

The non-vanishing terms correspond to $\mathbf{k} \neq \mathbf{j}$ and $\mathbf{l} \neq \mathbf{m}$.Then, (4.45) can be written as:

$$\begin{aligned}
&Var(T_2) = \\
&\frac{|H|^{1/2}}{N^2}\sum_{\mathbf{k}\neq\mathbf{j}}\sum_{\mathbf{l}\neq\mathbf{m}}\int_{\Pi^2} K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})d\boldsymbol{\lambda} \\
&\qquad\qquad \cdot\int_{\Pi^2} K_H(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{l}})K_H(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{m}})d\boldsymbol{\omega}E(Z_{\mathbf{k}}Z_{\mathbf{j}}Z_{\mathbf{l}}Z_{\mathbf{m}}) = \\
&\frac{2|H|^{1/2}}{N^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}}\int_{\Pi^2} K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})d\boldsymbol{\lambda} \\
&\qquad\qquad \cdot\int_{\Pi^2} K_H(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{j}})K_H(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{j}})d\boldsymbol{\omega} \\
&-\frac{2|H|^{1/2}}{N^2}\sum_{\mathbf{k}}\int_{\Pi^2} K_H^2(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})d\boldsymbol{\lambda}\int_{\Pi^2} K_H^2(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{k}})d\boldsymbol{\omega},
\end{aligned}$$

where $E(Z_{\mathbf{k}}Z_{\mathbf{j}}Z_{\mathbf{l}}Z_{\mathbf{m}}) = 1$ if an only if $\mathbf{k} = \mathbf{j} \neq \mathbf{l} = \mathbf{m}$ or $\mathbf{k} = \mathbf{m} \neq \mathbf{j} = \mathbf{l}$. The second addend is $\mathcal{O}(N^{-1}|H|^{-1/2})$:

$$\begin{aligned}
&\frac{2|H|^{1/2}}{N^2}\sum_{\mathbf{k}}\int_{\Pi^2}\int_{\Pi^2} K_H^2(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H^2(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{k}})d\boldsymbol{\lambda}d\boldsymbol{\omega} \\
&= \frac{2}{N^2|H|^{3/2}}\sum_{\mathbf{k}}\left(\int_{\Pi^2} K^2(H^{-1/2}(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}))d\boldsymbol{\lambda}\right)^2 \\
&= \frac{2}{N^2|H|^{1/2}}\sum_{\mathbf{k}}\left(\int_{\Pi^2} K^2(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{k}})d\boldsymbol{\omega}\right)^2 = \mathcal{O}(N^{-1}|H|^{-1/2}),
\end{aligned}$$

whereas, for the first term we obtain:

$$\frac{2|H|^{1/2}}{N^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}}\int\int K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})K_H(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{j}})d\boldsymbol{\lambda}d\boldsymbol{\omega}$$
$$=\frac{2|H|^{1/2}}{N^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}}\left(\int K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})d\boldsymbol{\lambda}\right)^2$$
$$=\frac{2}{N^2|H|^{1/2}}\sum_{\mathbf{k}}\sum_{\mathbf{j}}\left(\int K(\mathbf{u})K\left(\mathbf{u}+H^{-1/2}2\pi\left(\frac{k_1-j_1}{n_1},\frac{k_2-j_2}{n_2}\right)^T\right)d\mathbf{u}\right)^2$$
$$=\frac{2}{N|H|^{1/2}}\sum_{k_1=-2m_1}^{2m_1}\sum_{k_2=-2m_2}^{2m_2}c(\mathbf{k},N)\left(\int K(\mathbf{u})K\left(\mathbf{u}+H^{-1/2}2\pi\left(\frac{k_1}{n_1},\frac{k_2}{n_2}\right)^T\right)d\mathbf{u}\right)^2$$
$$\to\frac{1}{2\pi^2}\int_{2\Pi^2}\left(\int_{\Pi^2}K(\mathbf{u})K(\mathbf{u}+\mathbf{x})d\mathbf{u}\right)^2d\mathbf{x},$$

where $2\Pi^2=[-2\pi,2\pi]\times[-2\pi,2\pi]$ and $c(\mathbf{k},N)=\frac{2m_1+1-|k_1|}{n_1}\frac{2m_2+1-|k_2|}{n_2}$. Therefore, in order to prove the required condition, since c_{ij}^2 is a squared sum of $a(\mathbf{i},\mathbf{j})$ terms, we prove the condition for one of the addends, that is, for $a^2(\mathbf{i},\mathbf{j})$. Besides, using that $K_H(\cdot)\leq|H|^{-1/2}C$, for $0<C<\infty$:

$$k^4(n_1,n_2)\max_{\mathbf{k}}\sum_{\mathbf{j}}a^2(\mathbf{k},\mathbf{j})=$$
$$k^4(n_1,n_2)\max_{\mathbf{k}}\sum_{\mathbf{j}}\left(\frac{|H|^{1/4}}{N}\int K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})d\boldsymbol{\lambda}\right)^2$$
$$\frac{k^4(n_1,n_2)C^2}{N^2|H|^{1/2}}\sum_{\mathbf{j}}\int\int K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})K_H(\boldsymbol{\omega}-\boldsymbol{\lambda}_{\mathbf{j}})d\boldsymbol{\omega}d\boldsymbol{\lambda}=$$
$$\mathcal{O}(k^4(n_1,n_2)N^{-1}|H|^{-1/2}).$$

So, this condition is satisfied for all $k(n_1,n_2)\to\infty$ such that

$$\frac{k^4(n_1,n_2)}{n_1n_2|H|^{1/2}}\to 0.$$

2. We also have to check that $\max_{\mathbf{k}}E(Z_{\mathbf{k}}^2)\mathbf{1}_{\{|Z_{\mathbf{k}}|>k(n_1,n_2)\}}\to 0$, but this assertion follows just taking into account that $Z_{\mathbf{k}}$ are identically distributed with $E(Z_{\mathbf{k}}^2)=1$.

3. It remains to show that $\dfrac{\max_i\mu_i^2}{Var(Q_N)}\to 0$ where μ_i, $i=1,2,\ldots,M$ are the eigenvalues of the matrix $A=(c_{ij})$ define above.

The matrix A is symmetric, because the c_{ij} entries are defined in terms of the $a(\mathbf{i},\mathbf{j})$ terms defined above. Besides, the $a(\mathbf{i},\mathbf{j})$ satisfy that $a(\mathbf{i},\mathbf{j})=a(\mathbf{j},\mathbf{i})$, and

$$\sum_{\mathbf{j}}|a(\mathbf{i},\mathbf{j})|=\mathcal{O}(|H|^{1/4}).$$

Thus, the same condition applies on the c_{ij} terms.

Now, to prove the required condition, since A is a symmetric $M \times M$ matrix, there exists an ortogonal matrix U such that $U^{-1}AU$ is diagonal. This result implies that B is diagonalizable with real eigenvalues, $\{\mu_i\}$, with $i = 1, \ldots, M$, with $M = m_1' \times m_2'$. The $\|\cdot\|_\infty$ norm of the matrix B is given by:

$$\|A\|_\infty = \max_i \sum_j |c_{ij}|,$$

where the maximum is taken over $i \in \{1, \ldots, M\}$ and consider the spectral ratio of the matrix:

$$\rho(A) = \max_i |\mu_i|.$$

The spectral ratio of the matrix can be bounded by any norm in the matrix space $\mathcal{M}_{M\times M}$; therefore, for the particular case of the supremun norm $\|\cdot\|_\infty$:

$$\max_i |\mu_i| \leq \max_i \sum_j |c_{ij}|.$$

To prove the result, just take into account that:

$$\max_i \mu_i^2 \leq \left(\max_i |\mu_i|\right)^2 \leq \left(\max_i \sum_j |c_{ij}|\right)^2.$$

Then, since $|H|^{1/2} \to 0$ and $Var(Q_N) \to \tau^2$:

$$\frac{\max_i \mu_i^2}{Var(Q_N)} \leq \frac{\left(\max_i \sum_j |c_{ij}|\right)^2}{Var(Q_N)} = \frac{\mathcal{O}(|H|^{1/2})}{Var(Q_N)} \to 0.$$

□

Lemma 10. *Let T_P^0 denote the test statistic in (4.10) assuming that the true parameter is given by θ_0. Then, under assumptions in Theorem 1:*

$$T_P = T_P^0 + o_{\mathbb{P}}(1).$$

Proof. The test statistic T_P^0 is given by

$$T_P^0 = N|H|^{1/4} \int_{\Pi^2} \left(\frac{1}{N|H|^{1/2}} \sum_{\mathbf{k}} K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})) \left(\frac{I(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} - 1 \right) \right)^2 d\boldsymbol{\lambda}. \tag{4.47}$$

Note that:

$$\frac{I(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}})} - 1 = \left(\frac{I(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} - 1 \right) - \left(\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}}) - f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}})} \right) \frac{I(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})}. \tag{4.48}$$

Therefore,

$$T_P = T_P^0 + N|H|^{1/4} \int_{\Pi^2} \left(\frac{1}{N} \sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}) \frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} \left(\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{k}) - f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} \right) \right)^2 d\boldsymbol{\lambda}$$

$$- \frac{2|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{j}) \left(\frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} - 1 \right) \frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{j}) - f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{j})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{j})} \frac{I(\boldsymbol{\lambda}_\mathbf{j})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{j})} d\boldsymbol{\lambda}$$

For the second addend, using Lemma 5 and the fact that

$$\int_{\Pi^2} \left(\frac{1}{N} \sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}) \frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} \right)^2 d\boldsymbol{\lambda} = \mathcal{O}_{\mathbb{P}}(1)$$

we have

$$N|H|^{1/4} \int_{\Pi^2} \left(\frac{1}{N} \sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}) \frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} \left(\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{k}) - f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} \right) \right)^2 d\boldsymbol{\lambda}$$

$$\leq N|H|^{1/4} \left(\sup_{\mathbf{k}} \left| \frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{k}) - f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} \right| \right)^2 \int_{\Pi^2} \left(\frac{1}{N} \sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}) \frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} \right)^2 d\boldsymbol{\lambda} =$$

$\mathcal{O}_{\mathbb{P}}(|H|^{1/4})$.

For the last addend:

$$\frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{j}) \cdot$$

$$\left(\frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} - 1 \right) \frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{j}) - f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{j})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{j})} \frac{I(\boldsymbol{\lambda}_\mathbf{j})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{j})} d\boldsymbol{\lambda} =$$

$M_1 + M_2$,

where

$$M_1 = \frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{j}) \left(\frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} - 1 \right) \cdot$$

$$\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{j}) - f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{j})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{j})} \left(\frac{I(\boldsymbol{\lambda}_\mathbf{j})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{j})} - 1 \right) d\boldsymbol{\lambda}$$

and

$$M_2 = \frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{j}) \left(\frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} - 1 \right) \frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{j}) - f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{j})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_\mathbf{j})} d\boldsymbol{\lambda}.$$

We will prove that $M_1 = o_{\mathbb{P}}(1)$. Recall that

$$\frac{I(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})} - 1 = W_\mathbf{k} + \frac{R_n(\boldsymbol{\lambda}_\mathbf{k})}{f_{\theta_0}(\boldsymbol{\lambda}_\mathbf{k})},$$

where $W_{\mathbf{k}} = V_{\mathbf{k}} - 1$, and the $V_{\mathbf{k}}$ are independent identically distributed $Exp(1)$. Then,

$$M_1 = \frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) d\boldsymbol{\lambda}$$
$$\times \left\{ W_{\mathbf{k}} W_{\mathbf{j}} \left(\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}}) - f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}})} \right) + W_{\mathbf{k}} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})} \left(\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}}) - f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}})} \right) \right.$$
$$+ W_{\mathbf{j}} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} \left(\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}}) - f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}})} \right)$$
$$\left. + \frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})} \left(\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}}) - f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}})} \right) \right\} = C_1 + C_2 + C_3 + C_4.$$

In order to prove the bounds for C_j, $j = 1, 2, 3, 4$, we have to consider the Taylor expansion of $f_{\hat{\theta}}(\boldsymbol{\lambda})$ around $f_{\theta_0}(\boldsymbol{\lambda})$, for a fixed $\boldsymbol{\lambda}$:

$$f_{\hat{\theta}}(\boldsymbol{\lambda}) = f_{\theta_0}(\boldsymbol{\lambda}) + (\hat{\theta} - \theta_0)^T \nabla f_{\theta_0}(\boldsymbol{\lambda}) + \frac{1}{2} (\hat{\theta} - \theta_0)^T \nabla^2 f_{\tilde{\theta}}(\boldsymbol{\lambda}) (\hat{\theta} - \theta_0),$$

where $\|\tilde{\theta} - \theta_0\| \leq \|\hat{\theta} - \theta_0\|$. By similar arguments to those in Lemma 5, for $\boldsymbol{\lambda}_{\mathbf{j}}$

$$\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_j) - f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}})} = \mathcal{O}_{\mathbb{P}}(1) \left((\hat{\theta} - \theta_0)^T \nabla f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}}) + \frac{1}{2} (\hat{\theta} - \theta_0)^T \nabla^2 f_{\tilde{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}}) (\hat{\theta} - \theta_0) \right), \tag{4.49}$$

and the $\mathcal{O}_{\mathbb{P}}(1)$ factor is uniform in $\mathbf{j}$. We will see that $C_1 = \mathcal{O}_{\mathbb{P}}(N^{-1/2}) + \mathcal{O}_{\mathbb{P}}(|H|^{1/4})$. Taking into account (4.49), C_1 can be written as:

$$C_1 = \mathcal{O}_{\mathbb{P}}(1)(\hat{\theta} - \theta_0)^T \frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) W_{\mathbf{k}} W_{\mathbf{j}} \nabla f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}}) d\boldsymbol{\lambda}$$

$$+ \mathcal{O}_{\mathbb{P}}(1)(\hat{\theta} - \theta_0)^T \frac{|H|^{1/4}}{2N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) W_{\mathbf{k}} W_{\mathbf{j}} \nabla^2 f_{\tilde{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}}) d\boldsymbol{\lambda} (\hat{\theta} - \theta_0).$$

Since

$$\frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) W_{\mathbf{k}} W_{\mathbf{j}} d\boldsymbol{\lambda} = \mathcal{O}_{\mathbb{P}}(1)$$

and the derivatives of f_θ are uniformly bounded, the first addend in C_1 is $\mathcal{O}_{\mathbb{P}}(N^{-1/2})$. Taking into account that $(\hat{\theta} - \theta_0) = \mathcal{O}_{\mathbb{P}}(N^{-1/2})$, the second addend is $\mathcal{O}_{\mathbb{P}}(|H|^{1/4})$. In order to obtain a bound for C_2, one should consider the results in Lemma 8. From Taylor expansion (4.49), C_2 can be written as:

$$C_2 = \mathcal{O}_{\mathbb{P}}(1)(\hat{\theta} - \theta_0)^T \frac{|H|^{1/4}}{N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) W_{\mathbf{k}} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})} \nabla f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}}) d\boldsymbol{\lambda}$$

$$+ \mathcal{O}_{\mathbb{P}}(1)(\hat{\theta} - \theta_0)^T \frac{|H|^{1/4}}{2N} \int_{\Pi^2} \sum_{\mathbf{k}} \sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{j}}) W_{\mathbf{k}} \frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})} \nabla^2 f_{\tilde{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}}) (\hat{\theta} - \theta_0).$$

From Lemma 8, we have that:

$$\frac{|H|^{1/4}}{2N}\int_{\Pi^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})W_{\mathbf{k}}\frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}d\boldsymbol{\lambda}=o_{\mathbb{P}}(1).$$

Then, the first addend in C_2 is $\mathcal{O}_{\mathbb{P}}(N^{-1/2})o_{\mathbb{P}}(1)$. For the second addend, one should note that $|R_n(\boldsymbol{\lambda}_{\mathbf{j}})|=\mathcal{O}_{\mathbb{P}}(N^{-1/2})$, from Lemma 6. Then:

$$\mathcal{O}_{\mathbb{P}}(1)(\hat{\theta}-\theta_0)^T\frac{|H|^{1/4}}{2N}\int_{\Pi^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})W_{\mathbf{k}}\frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}\nabla^2 f_{\tilde{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}})(\hat{\theta}-\theta_0)d\boldsymbol{\lambda}=$$

$$=\mathcal{O}_{\mathbb{P}}(N^{-1/2}|H|^{1/4}).$$

The third addend C_3 can be bounded using the same arguments as in the proof for C_2. For the last addend C_4, and taking also into account Lemma 6:

$$|C_4|\leq\sum_{\mathbf{j}}\left|\frac{f_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{j}})-f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}\right|\frac{|H|^{1/4}}{N}\int_{\Pi^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})\frac{|R_n(\boldsymbol{\lambda}_{\mathbf{j}})|}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}\frac{|R_n(\boldsymbol{\lambda}_{\mathbf{k}})|}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})}d\boldsymbol{\lambda}$$

$$=\mathcal{O}_{\mathbb{P}}(N^{-1/2})\mathcal{O}_{\mathbb{P}}(|H|^{1/4}N)\mathcal{O}_{\mathbb{P}}(N^{-1})=\mathcal{O}_{\mathbb{P}}(|H|^{1/4}N^{-1/2}).$$

$M_2=o_{\mathbb{P}}(1)$ can be proved using similar arguments. □

Lemma 11. *If $\theta=\theta_0$ is the true parameter, under assumptions (1)-(4):*

$$T_P^0-\mu_H\to N(0,\tau^2),$$

as $N\to\infty$, where μ_H and τ^2 are given in (4.11) and (4.12), respectively and T_P^0 is given in (4.47).

Proof. Recall the expression for the periodogram

$$I(\boldsymbol{\lambda}_{\mathbf{k}})=f(\boldsymbol{\lambda}_{\mathbf{k}})V_{\mathbf{k}}+R_n(\boldsymbol{\lambda}_{\mathbf{k}}),\tag{4.50}$$

where $\{\boldsymbol{\lambda}_{\mathbf{k}}\}$ denote the Fourier frequencies and recall the notation $W_{\mathbf{k}}=1-V_{\mathbf{k}}$ (where $V_{\mathbf{k}}$ are independent identically distributed random variables with $Exp(1)$ distribution) introduced in Lemma 10. Then:

$$\frac{I(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})}-1=W_{\mathbf{k}}+\frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})}.$$

The statistic T_P^0 can be decomposed in three addends in the following way:

$$T_P^0-\mu_H=\frac{|H|^{1/4}}{N}\int_{\Pi^2}\left(\sum_{\mathbf{k}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})W_{\mathbf{k}}\right)^2d\boldsymbol{\lambda}-\mu_H\tag{4.51}$$

$$+\frac{|H|^{1/4}}{N}\int_{\Pi^2}\left(\sum_{\mathbf{k}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})\frac{R_n(\boldsymbol{\lambda}_{\mathbf{k}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{k}})}\right)^2d\boldsymbol{\lambda}\tag{4.52}$$

$$+\frac{2|H|^{1/4}}{N}\int_{\Pi^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{j}})W_{\mathbf{k}}\frac{R_n(\boldsymbol{\lambda}_{\mathbf{j}})}{f_{\theta_0}(\boldsymbol{\lambda}_{\mathbf{j}})}d\boldsymbol{\lambda}\tag{4.53}$$

From Lemma 7, (4.52) tends to zero in probability. Also, from Lemma 8, (4.53) tends to zero in probability. The theorem is proved by Lemma 9. □

Proof of Theorem 1. Theorem 1 is proved combining the results in Lemma 10 and Lemma 11. □

4.6.2 Proofs of Theorems 2 and 3.

Before proving the theorem, we must verify that (4.13) holds. The prove of the following lemma is obtained generalizing Theorem 3.2 in Dahlhaus and Wefelmeyer (1996).

Lemma 12. *Under assumptions (2) and (5), if f is bounded and bounded away from cero, then*

$$\sqrt{N}(\hat{\theta} - \theta^*) - \sqrt{N} \int_{\Pi^2} W(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda}))d\boldsymbol{\lambda} \to 0$$

in probability, where

$$W(\boldsymbol{\lambda}) = -\mathcal{H}^{-1}\nabla f_\theta^{-1}(\boldsymbol{\lambda})|_{\theta=\theta^*}, \qquad \mathcal{H} = \int_{\Pi^2} \nabla^2 G(\theta^*, f, \boldsymbol{\lambda})d\boldsymbol{\lambda},$$

$$G(\theta, f, \boldsymbol{\lambda}) = \log f_\theta(\boldsymbol{\lambda}) + \frac{f(\boldsymbol{\lambda})}{f_\theta(\boldsymbol{\lambda})}.$$

Proof. Let's write the Kullback-Lebiler discrepancy between f and f_θ

$$L(\theta, f) = \int_{\Pi^2} \left(\log f_\theta(\boldsymbol{\lambda}) + \frac{f(\boldsymbol{\lambda})}{f_\theta(\boldsymbol{\lambda})} \right) d\boldsymbol{\lambda}, \tag{4.54}$$

in a more general form as:

$$L(\theta, f) = \int_{\Pi^2} G(\theta, f, \boldsymbol{\lambda})d\boldsymbol{\lambda}.$$

In our particular case, the function in the integrand is given by:

$$G(\theta, f, \boldsymbol{\lambda}) = a_\theta(\boldsymbol{\lambda}) + b_\theta(\boldsymbol{\lambda})f(\boldsymbol{\lambda}) \quad \text{where } a_\theta(\boldsymbol{\lambda}) = \log f_\theta(\boldsymbol{\lambda}), \quad b_\theta(\boldsymbol{\lambda}) = f_\theta^{-1}(\boldsymbol{\lambda}).$$

Then, the $\mathcal{H}$ matrix can be written as:

$$\mathcal{H} = \int_{\Pi^2} \left(\nabla^2 a_{\theta^*}(\boldsymbol{\lambda}) + \nabla^2 b_{\theta^*}(\boldsymbol{\lambda}) f(\boldsymbol{\lambda}) \right) d\boldsymbol{\lambda},$$

where θ^* gives the best fit in $\mathcal{F}_\theta$. Considering $L(\theta, I)$ the analogous expression to (4.54), but replacing f by the periodogram I, it is straightforward to see that $L(\hat{\theta}, I) \leq L(\theta^*, I)$ and $L(\theta^*, f) \leq L(\hat{\theta}, f)$, only recalling the definitions of $\hat{\theta}$ and θ^*:

$$\hat{\theta} = \arg\min_\theta L(\theta, I) \text{ and } \theta^* = \arg\min_\theta L(\theta, f).$$

Since

$$\sup_{\theta} |L(\theta, I) - L(\theta, f)| \to 0 \tag{4.55}$$

in probability (see Dahlhaus and Wefelmeyer (1996), Lemma A.7), then the Kullback-Leibler discrepancy $L(\hat{\theta}, f)$ converges to $L(\theta^*, f)$ in probability. This result is proved by the convergence of Cesaro sums of the Fourier transform of $f_{\theta}^{-1}(\boldsymbol{\lambda})$.

Therefore, $\hat{\theta}$ tends to θ^* in probability. The result follows from a Taylor expansion of $\nabla L(\hat{\theta}, I)$ around $\nabla L(\theta^*, I)$. Note that $\nabla L(\hat{\theta}, I) = 0$, then:

$$0 = \nabla L(\theta^*, I) + \nabla^2 L(\tilde{\theta}, I)(\hat{\theta} - \theta^*). \tag{4.56}$$

for $\tilde{\theta}$ such that $\|\tilde{\theta} - \theta^*\| \leq \|\hat{\theta} - \theta^*\|$. For the first addend:

$$\begin{aligned}
\nabla L(\theta^*, I) &= \int_{\Pi^2} \nabla G(\theta^*, I, \boldsymbol{\lambda}) d\boldsymbol{\lambda} = \int_{\Pi^2} \left(\nabla a_{\theta^*}(\boldsymbol{\lambda}) + \nabla b_{\theta^*}(\boldsymbol{\lambda}) I(\boldsymbol{\lambda})\right) d\boldsymbol{\lambda} \\
&= \int_{\Pi^2} \nabla G(\theta^*, f, \boldsymbol{\lambda}) d\boldsymbol{\lambda} + \int_{\Pi^2} \nabla b_{\theta^*}(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda})) d\boldsymbol{\lambda} \\
&= \int_{\Pi^2} \nabla b_{\theta^*}(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda})) d\boldsymbol{\lambda},
\end{aligned}$$

since the first term is zero. For the second addend in (4.56), it can be seen that, for $\tilde{\theta}$ such that $\|\tilde{\theta} - \theta^*\| \leq \|\hat{\theta} - \theta^*\|$:

$$\nabla^2 L(\tilde{\theta}, I) = \int_{\Pi^2} \nabla^2 G(\tilde{\theta}, f, \boldsymbol{\lambda}) d\boldsymbol{\lambda} + \int_{\Pi^2} \nabla^2 b_{\tilde{\theta}}(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda})) d\boldsymbol{\lambda}.$$

Then, (4.56) can be written as:

$$-\int_{\Pi^2} \nabla b_{\theta^*}(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda})) d\boldsymbol{\lambda} = \int_{\Pi^2} \nabla^2 G(\tilde{\theta}, f, \boldsymbol{\lambda}) d\boldsymbol{\lambda} + \int_{\Pi^2} \nabla^2 b_{\tilde{\theta}}(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda})) d\boldsymbol{\lambda}.$$

Provided that $\mathcal{H}$ is non-singular:

$$-\sqrt{N}\mathcal{H}^{-1} \int_{\Pi^2} \nabla b_{\theta^*}(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda})) d\boldsymbol{\lambda} = \tag{4.57}$$

$$\sqrt{N}\mathcal{H}^{-1} \left(\int_{\Pi^2} \nabla^2 G(\tilde{\theta}, f, \boldsymbol{\lambda}) d\boldsymbol{\lambda} + \int_{\Pi^2} \nabla^2 b_{\tilde{\theta}}(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda})) d\boldsymbol{\lambda} \right) (\hat{\theta} - \theta^*). \tag{4.58}$$

By the smoothness of G,

$$\int_{\Pi^2} \nabla^2 G(\tilde{\theta}, f, \boldsymbol{\lambda}) d\boldsymbol{\lambda} \to \mathcal{H} \tag{4.59}$$

in probability, and by Lemma A.7 in (Dahlhaus and Wefelmeyer (1996)):

$$\int_{\Pi^2} \nabla^2 b_{\tilde{\theta}}(\boldsymbol{\lambda})(I(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda})) d\boldsymbol{\lambda} \to 0 \tag{4.60}$$

also in probability.

The result is proved replacing (4.59) and (4.60) in (4.57)-(4.58). □

Proof of Theorem 2. Once we have obtained the $\sqrt{N}$-consistency of $\hat{\theta}$ as an estimator of θ^*, the proof of the theorem is analogous as the proof of Theorem 3 in Paparoditis (2000). Note that:

$$\frac{I(\boldsymbol{\lambda_k})}{f_{\hat{\theta}}(\boldsymbol{\lambda_k})} - 1 = \left(\frac{I(\boldsymbol{\lambda_k})}{f(\boldsymbol{\lambda_k})} - 1\right) + \frac{I(\boldsymbol{\lambda_k})}{f_{\theta^*}(\boldsymbol{\lambda_k})}\left(\frac{f_{\theta^*}(\boldsymbol{\lambda_k})}{f_{\hat{\theta}}(\boldsymbol{\lambda_k})} - 1\right) + \frac{I(\boldsymbol{\lambda_k})}{f(\boldsymbol{\lambda_k})}\left(\frac{f(\boldsymbol{\lambda_k})}{f_{\theta^*}(\boldsymbol{\lambda_k})} - 1\right),$$

and recall that $I(\boldsymbol{\lambda_k})/f(\boldsymbol{\lambda_k}) = V_{\mathbf{k}} + R_n(\boldsymbol{\lambda_k})/f(\boldsymbol{\lambda_k})$. Then,

$$N^{-1}|H|^{-1/4}T_P = \int_{\Pi^2}\left(\frac{1}{N|H|^{1/2}}\sum_{\mathbf{k}} K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda_k}))\frac{I(\boldsymbol{\lambda_k})}{f(\boldsymbol{\lambda_k})}\left(\frac{f(\boldsymbol{\lambda_k})}{f_{\theta^*}(\boldsymbol{\lambda_k})} - 1\right)\right)^2 d\boldsymbol{\lambda} + o_{\mathbb{P}}(1).$$

The first addend can be decomposed in two terms:

$$\int_{\Pi^2}\left(N^{-1}\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda_k})\left(\frac{f(\boldsymbol{\lambda_k})}{f_{\theta^*}(\boldsymbol{\lambda_k})} - 1\right)\right)^2 d\boldsymbol{\lambda}$$
$$+ \int_{\Pi^2}\left(N^{-1}\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda_k})\left(W_{\mathbf{k}} + \frac{R_n(\boldsymbol{\lambda_k})}{f(\boldsymbol{\lambda_k})}\right)\left(\frac{f(\boldsymbol{\lambda_k})}{f_{\theta^*}(\boldsymbol{\lambda_k})} - 1\right)\right)^2 d\boldsymbol{\lambda},$$

where $W_{\mathbf{k}} = V_{\mathbf{k}} - 1$. From Lemma 8:

$$N^{-2}\int_{\Pi^2}\sum_{\mathbf{k}}\sum_{\mathbf{j}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda_k})K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda_j})f(\boldsymbol{\lambda_k})/f_{\theta^*}(\boldsymbol{\lambda_k})f(\boldsymbol{\lambda_j})/f_{\theta^*}(\boldsymbol{\lambda_j})W_{\mathbf{k}}\frac{R_n(\boldsymbol{\lambda_j})}{f(\boldsymbol{\lambda_j})}d\boldsymbol{\lambda} = o_{\mathbb{P}}(1).$$

From Lemma 7, we have:

$$N^{-2}\int_{\Pi^2}\left(\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda_k})f(\boldsymbol{\lambda_k})/f_{\theta^*}(\boldsymbol{\lambda_k})\frac{R_n(\boldsymbol{\lambda_k})}{f(\boldsymbol{\lambda_k})}\right)^2 d\boldsymbol{\lambda} = o_{\mathbb{P}}(1).$$

Besides,

$$N^{-2}\int_{\Pi^2}\left(\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda_k})W_{\mathbf{k}}\right)^2 d\boldsymbol{\lambda} \to 0.$$

Then,

$$N^{-1}|H|^{-1/4}T_P = \int_{\Pi^2}\left(N^{-1}\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda_k})\left(\frac{f(\boldsymbol{\lambda_k})}{f_{\theta^*}(\boldsymbol{\lambda_k})} - 1\right)\right)^2 d\boldsymbol{\lambda} + o_{\mathbb{P}}(1). \tag{4.61}$$

Besides, this uniform convergence holds:

$$N^{-1}\sum_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda_k})\frac{f(\boldsymbol{\lambda_k})}{f_{\theta^*}(\boldsymbol{\lambda_k})} \to \frac{f(\boldsymbol{\lambda})}{f_{\theta^*}(\boldsymbol{\lambda})} \tag{4.62}$$

The result is concluded from (5.28) and (4.62).

□

Proof of Theorem 3. The result is obtained using:

$$\frac{I(\boldsymbol{\lambda_k})}{f_{\theta_0}(\boldsymbol{\lambda_k})} - 1 = \left(\frac{I(\boldsymbol{\lambda_k})}{f(\boldsymbol{\lambda_k})}\right) + \left(W_{\mathbf{k}} + \frac{R_n(\boldsymbol{\lambda_k})}{f(\boldsymbol{\lambda_k})}\right)\left(\frac{f(\boldsymbol{\lambda_k})}{f_{\theta_0}(\boldsymbol{\lambda_k})} - 1\right) + \left(\frac{f(\boldsymbol{\lambda_k})}{f_{\theta_0}(\boldsymbol{\lambda_k})} - 1\right)$$

and similar arguments to those in the proof of Theorem 2. □

4.6.3 Proof of Theorem 4.

From now on, note that $N = n_1 n_2$ denotes the number of data points whereas n denotes the number of Fourier frequencies and $\|\cdot\|$ is the L^2-norm. We will drop the subindex 0 and denote by θ the true parameter under the null hypothesis. Define

$$q_1(m(\boldsymbol{\lambda}_\mathbf{k}), Y_\mathbf{k}) = +e^{Y_\mathbf{k} - m(\boldsymbol{\lambda}_\mathbf{k})} - 1, \quad q_2(m(\boldsymbol{\lambda}_\mathbf{k}), Y_\mathbf{k}) = -e^{Y_\mathbf{k} - m(\boldsymbol{\lambda}_\mathbf{k})},$$

where m can be replaced by m_θ or by $\hat{m}_{LK}$. Assume m_θ is the log-likelihood under the null hypothesis and denote

$$\varepsilon_\mathbf{k} = q_1(m_\theta(\boldsymbol{\lambda}_\mathbf{k}), Y_\mathbf{k}^{**}) = e^{Y_\mathbf{k}^{**} - m_\theta(\boldsymbol{\lambda}_\mathbf{k})} - 1, \quad q_2^\mathbf{k} = q_2(m_\theta(\boldsymbol{\lambda}_\mathbf{k}), Y_\mathbf{k}^{**}) = -e^{Y_\mathbf{k}^{**} - m_\theta(\boldsymbol{\lambda}_\mathbf{k})},$$

where $Y_\mathbf{k}^{**}$ is given by (4.14). Define, also

$$\Gamma(\boldsymbol{\lambda}_\mathbf{k}) = -E\left(q_2(m_\theta(\boldsymbol{\lambda}), Y_\mathbf{k}^{**})\right) \frac{1}{4\pi^2}.$$

Some other constants and vectors that will appear in our computations are

$$\boldsymbol{\beta}^T = \boldsymbol{\beta}(\boldsymbol{\lambda})^T = \Big(m_\theta(\boldsymbol{\lambda}), |H|^{1/2} \nabla^T m_\theta(\boldsymbol{\lambda})\Big) \in \mathbb{R}^3,$$

where $\nabla m_\theta(\boldsymbol{\lambda})$ denotes the gradient vector $\nabla m_\theta(\boldsymbol{\lambda}) = \Big(\frac{\partial m_\theta}{\partial x}(\boldsymbol{\lambda}), \frac{\partial m_\theta}{\partial y}(\boldsymbol{\lambda})\Big)$ and $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$ denote the derivatives with respect to the first and second components. Besides, define

$$\boldsymbol{\beta}_2^T = \boldsymbol{\beta}_2(\boldsymbol{\lambda})^T = \sqrt{n|H|^{1/2}}(a - m(\boldsymbol{\lambda}), |H|^{1/2}(\mathbf{b} - \nabla m(\boldsymbol{\lambda}))^T) \in \mathbb{R}^3,$$

where $(a, \mathbf{b})$ are the parameters in the non-parametric model (2.89),

$$W_\mathbf{k} = W_\mathbf{k}(\boldsymbol{\lambda}) = (1, |H|^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k})) \in \mathbb{R}^3,$$

$$r_N^2 = \frac{1}{N|H|^{1/2}}, \quad \text{and } \bar{m}_\mathbf{k} = \bar{m}_\mathbf{k}(\boldsymbol{\lambda}) = m_\theta(\boldsymbol{\lambda}) + \nabla^T m_\theta(\boldsymbol{\lambda})(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}).$$

Besides,

$$\Phi_{n,j} = \sup_{\boldsymbol{\lambda} \in \Pi^2, \|\alpha\| = c_1 r_N} \left| q_2(\boldsymbol{\beta}_*^T W_\mathbf{k} + \alpha^T W_\mathbf{k}, Y_\mathbf{k}) |H|^{(j-1)/2} \|(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k})\|^{(j-1)} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}) \right|,$$

where $\boldsymbol{\beta}_*$ denotes $\boldsymbol{\beta}$ or $\boldsymbol{\beta}_2$ and assume that

$$E(\Phi_{n,j})^\zeta = \mathcal{O}(1), j = 1, 2, 3.$$

Lemma 13. *The Generalized Likelihood Ratio Test statistic*

$$T_{LK} = \sum_\mathbf{k} \left[e^{Y_\mathbf{k} - m_\theta(\boldsymbol{\lambda}_\mathbf{k})} + m_\theta(\boldsymbol{\lambda}_\mathbf{k}) - e^{Y_\mathbf{k} - \hat{m}_{LK}(\boldsymbol{\lambda}_\mathbf{k})} - \hat{m}_{LK}(\boldsymbol{\lambda}_\mathbf{k}) \right] \tag{4.63}$$

admits the following decomposition

$$T_{LK} = T^*_{LK} + B_1 + B_2 - B_3 \tag{4.64}$$

where T^*_{LK} *is the same as* T_{LK} *but replacing* $Y_{\mathbf{k}}$ *by* $Y^{**}_{\mathbf{k}}$ *and* $\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})$ *by* $\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})$ *and,*

$$B_1 = \sum_{\mathbf{k}} \left\{1 - e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}\right\} \left(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})\right),$$

$$B_2 = \sum_{\mathbf{k}} e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} \left(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})\right)^2,$$

$$B_3 = \sum_{\mathbf{k}} \frac{R_N(\boldsymbol{\lambda}_{\mathbf{k}})}{f_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} \left\{e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - 1\right\}.$$

Proof.

$$\begin{aligned}
&T_{LK} - T^*_{LK} = \\
&\sum_{\mathbf{k}} \Big[\hat{m}^*_{LK}(\boldsymbol{\lambda}_k) - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) + e^{Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} \\
&+ e^{Y^{**}_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}\Big] = \\
&\sum_{\mathbf{k}} \Big[\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) + e^{Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} + e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} \\
&- e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} + e^{Y^{**}_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}\Big]
\end{aligned}$$

By Taylor's expansion of $h_{\mathbf{k}}(x) = e^{Y_{\mathbf{k}} - x}$ evaluated at $\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})$, and doing the expansion around $\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})$:

$$e^{Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} = e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}$$

$$+\frac{1}{2}(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^2 e^{Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - \frac{1}{3!}(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^3 e^{Y_{\mathbf{k}} - z_{\mathbf{k}}}$$

where $z_{\mathbf{k}}$ is such that $|\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})| \geq |z_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})|$. The last addend is given in Lagrange's remainder form, and it can be bounded by:

$$\begin{aligned}
&(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^3 e^{Y_{\mathbf{k}} - z_{\mathbf{k}}} \\
&= (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^3 \left(1 + (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - z_{\mathbf{k}}) + \frac{1}{2}(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - z_{\mathbf{k}})^2 + \dots\right) \\
&= \mathcal{O}_{\mathbb{P}}(N^{-3/2} \log^3 N),
\end{aligned}$$

applying Lemma 6:

$$|Y_{\mathbf{k}} - z_{\mathbf{k}}| \leq |\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - z_{\mathbf{k}}| \leq \mathcal{O}_{\mathbb{P}}(N^{-1/2} \log N).$$

Then, $T_{LK} - T^*_{LK}$ can be written as:

$$\begin{aligned}
&\sum_{\mathbf{k}} \Big[e^{Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} + e^{Y^{**}_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} + \\
&\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) + e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} (\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&+ \frac{1}{2} e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^2 + \mathcal{O}_{\mathbb{P}}(N^{-3/2} \log^3 N) \Big] \\
&= \sum_{\mathbf{k}} \Big[e^{Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} + e^{Y^{**}_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} \\
&+ (1 - e^{Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}) (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) \\
&+ \frac{1}{2} e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^2 + \mathcal{O}_{\mathbb{P}}(N^{-3/2} \log^3 N) \Big] = \\
&\sum_{\mathbf{k}} \Big[(1 - e^{Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}) (\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) \Big] + \\
&\sum_{\mathbf{k}} \Big[\frac{1}{2} e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} (\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^2 \Big] + \\
&\sum_{\mathbf{k}} \Big[e^{Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} + e^{Y^{**}_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} \Big] \\
&+ \mathcal{O}_{\mathbb{P}}(N^{-3/2} \log^3 N)
\end{aligned}$$

The first two addends correspond to B_1 and B_2. To obtain the third part, B_3, we should recall the following relations:

$$I(\boldsymbol{\lambda}_{\mathbf{k}}) = f_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) V_{\mathbf{k}} + R_N(\boldsymbol{\lambda}_{\mathbf{k}}), \quad Y_{\mathbf{k}} = m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) + z_{\mathbf{k}} + r_{\mathbf{k}} \tag{4.65}$$

$$e^{Y_{\mathbf{k}}} = f_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) V_{\mathbf{k}} e^{r_{\mathbf{k}}}, \quad e^{Y^{**}_{\mathbf{k}}} = f_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) V_{\mathbf{k}}, \quad e^{Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}}} = f_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) V_{\mathbf{k}} (e^{r_{\mathbf{k}}} - 1). \tag{4.66}$$

And recall also that $e^{m_\theta(\boldsymbol{\lambda})} = f_\theta(\boldsymbol{\lambda})$. In order to derive the final expression for $B3$, we must consider the following Taylor's expansion:

$$e^{Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} = e^{Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} + (Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}}) e^{Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} + \frac{1}{2} (Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}})^2 e^{c_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}, \tag{4.67}$$

where $c_{\mathbf{k}}$ is such that $|Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}}| \geq |c_{\mathbf{k}} - Y^{**}_{\mathbf{k}}|$. Besides, the difference between $Y_{\mathbf{k}}$ and $Y^{**}_{\mathbf{k}}$ is bounded by:

$$|Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}}| = |r_{\mathbf{k}} - C_0|, \quad \text{where } C_0 \text{ is the Euler constant.}$$

From the expression for $r_{\mathbf{k}}$:

$$r_{\mathbf{k}} = \log \left(1 + \frac{R_N(\boldsymbol{\lambda}_{\mathbf{k}})}{f(\boldsymbol{\lambda}_{\mathbf{k}}) V_{\mathbf{k}}} \right),$$

and taking into account that $R_N(\boldsymbol{\lambda}_{\mathbf{k}})$ is uniformly bounded by:

$$\max_{\mathbf{k}} |R_N(\boldsymbol{\lambda}_{\mathbf{k}})| = \mathcal{O}_{\mathbb{P}}(N^{-1/2} \log N), \tag{4.68}$$

the remainder in Taylor's expansion can be bounded by:

$$|Y_{\mathbf{k}} - Y_{\mathbf{k}}^{**}| = \mathcal{O}_{\mathbb{P}}(\log N^{-1/2} \log\log N),$$

and this bound is uniform in $\boldsymbol{\lambda}_{\mathbf{k}}$. Using similar arguments, we have:

$$e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} = e^{Y^{**}_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} + (Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}}) e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} + \mathcal{O}_{\mathbb{P}}(\log N^{-1/2} + \log\log N). \tag{4.69}$$

Applying Taylor's expansion in $B3$ (for two groups of addends) we have

$$B3 = \sum_{\mathbf{k}} \left(e^{Y_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}(Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}}) - e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}(Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}}) + \mathcal{O}_{\mathbb{P}}(\log N^{-1/2} + \log\log N) \right) =$$

$$\sum_{\mathbf{k}} e^{Y_{\mathbf{k}}}(Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}})(e^{-m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} - e^{-\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}) + \mathcal{O}_{\mathbb{P}}(N \log N^{-1/2} \log\log N) =$$

$$\sum_{\mathbf{k}} e^{Y_{\mathbf{k}}}(Y_{\mathbf{k}} - Y^{**}_{\mathbf{k}}) \frac{1}{f_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} \left(1 - e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} \right) + \mathcal{O}_{\mathbb{P}}(N \log N^{-1/2} + N \log\log N)$$

And, with another Taylor's expansion on $e^{Y_{\mathbf{k}}}$ around $e^{Y^{**}_{\mathbf{k}}}$:

$$B3 = \sum_{\mathbf{k}} (e^{Y_{\mathbf{k}}} - e^{Y^{**}_{\mathbf{k}}}) \frac{1}{f_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} (1 - e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}) + \mathcal{O}_{\mathbb{P}}(N \log N^{-1/2} + N \log\log N) =$$

$$\sum_{\mathbf{k}} r_N(\boldsymbol{\lambda}_{\mathbf{k}}) \frac{1}{f_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} (1 - e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}) + \mathcal{O}_{\mathbb{P}}(N \log N^{-1/2} + N \log\log N).$$

□

We prove now that T^*_{LK} follows an asymptotically normal distribution.

Proof of Theorem 4. .
The regresion model (4.14) under the null hypothesis

$$Y^{**}_{\mathbf{k}} = m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) + z_k \tag{4.70}$$

can be seen regresion model with non-Gaussian error variables with density (2.17). The asymptotic distribution of T^*_{LK} is obtained as a particular case of Theorem 10 in (Fan *et al.* (2001)), extended to the bidimensional situation. The loglikelihood associated with model (5.10):

$$f(Y^{**}_{\mathbf{k}}, m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})) = Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} \tag{4.71}$$

and the generalized likelihood ratio test statistic is given by

$$T^*_{LK} = \sum_{\mathbf{k}} \left\{ e^{Y^{**}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} + m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y^{**}_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) \right\}.$$

Using Taylor's expansion of the loglikelihood function, with the notation introduced above:

$$T^*_{LK} = \sum_{\mathbf{k}} \left\{ \varepsilon_{\mathbf{k}} (\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})) + q_2^{\mathbf{k}} \frac{1}{2} (\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}))^2 + \mathcal{O}_{\mathbb{P}}(N^{-3/2} \log^3 N) \right\} \quad (4.72)$$

For the sake of simplicity, we will drop the residual part. Now, using the asymptotic representation for the nonparametric estimator given in Lemma 15, and the expression for $H_N(\boldsymbol{\lambda})$ in (4.73), the non-negligible part of (4.72) can be written as:

$$\begin{aligned}
T^*_{LK} &= \sum_{\mathbf{k}} \left\{ \varepsilon_{\mathbf{k}} r_N^2 \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} \sum_{\mathbf{i}} \varepsilon_{\mathbf{i}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}}))(1 + o_{\mathbb{P}}(1)) + \varepsilon_{\mathbf{k}} H_N(\boldsymbol{\lambda}_{\mathbf{k}}) \right. \\
&\left. + \frac{1}{2} q_2^{\mathbf{k}} \left[r_N^2 \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} \sum_{\mathbf{i}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}}))(1 + o_{\mathbb{P}}(1)) + H_N(\boldsymbol{\lambda}_{\mathbf{k}}) \right]^2 \right\} = \\
&r_N^2 \sum_{\mathbf{k}} \sum_{i} \varepsilon_{\mathbf{k}} \varepsilon_i \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}}))(1 + o_{\mathbb{P}}(1)) + \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} H_N(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&+ \frac{1}{2} \sum_{\mathbf{k}} q_2^{\mathbf{k}} \left[r_N^4 \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-2} \sum_{\mathbf{i}} \sum_{\mathbf{j}} \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{j}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{j}} - \boldsymbol{\lambda}_{\mathbf{k}}))(1 + o_{\mathbb{P}}(1))^2 \right. \\
&\left. + H_N^2(\boldsymbol{\lambda}_{\mathbf{k}}) + 2 r_N^2 \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} \sum_{\mathbf{i}} \varepsilon_{\mathbf{i}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}}))(1 + o_{\mathbb{P}}(1)) H_N(\boldsymbol{\lambda}_{\mathbf{k}}) \right] = \\
&r_N^2 \sum_{\mathbf{k}} \sum_{\mathbf{i}} \varepsilon_{\mathbf{k}} \varepsilon_i \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}}))(1 + o_{\mathbb{P}}(1)) + \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} H_N(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&+ \frac{r_N^4}{2} \sum_{\mathbf{k}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-2} \sum_{\mathbf{i}} \sum_{\mathbf{j}} \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{j}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{j}} - \boldsymbol{\lambda}_{\mathbf{k}}))(1 + o_{\mathbb{P}}(1))^2 \\
&+ \frac{1}{2} \sum_{\mathbf{k}} q_2^{\mathbf{k}} H_N^2(\boldsymbol{\lambda}_{\mathbf{k}}) - r_N^2 \sum_{\mathbf{k}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} \sum_{\mathbf{i}} \varepsilon_{\mathbf{i}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}}))(1 + o_{\mathbb{P}}(1)) H_N(\boldsymbol{\lambda}_{\mathbf{k}}) = \\
&S_{1N} + S_{2N} + R_{1N} + R_{2N} + R_{3N},
\end{aligned}$$

where

$$H_N(\boldsymbol{\lambda}) = r_N^2 \Gamma(\boldsymbol{\lambda})^{-1} \sum_{\mathbf{k}} \left[q_1(\beta(\boldsymbol{\lambda})^T W_{\mathbf{k}}, Y^{**}_{\mathbf{k}}) - \varepsilon_{\mathbf{k}} \right] K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}))(1 + o_{\mathbb{P}}(1)). \quad (4.73)$$

The residual terms are

$$R_{1N} = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} H_N(\boldsymbol{\lambda}_{\mathbf{k}}), \quad (4.74)$$

$$R_{2N} = \frac{-1}{2} \sum_{\mathbf{k}} q_2^{\mathbf{k}} H_N^2(\boldsymbol{\lambda}_{\mathbf{k}}), \quad (4.75)$$

$$R_{3N} = -r_N^2 \sum_{\mathbf{k}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} \sum_{\mathbf{i}} \varepsilon_{\mathbf{i}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) H_N(\boldsymbol{\lambda}_{\mathbf{k}}), \quad (4.76)$$

and the leading terms:

$$S_{1N} = r_N^2 \sum_{\mathbf{k}} \sum_{\mathbf{i}} \varepsilon_{\mathbf{k}} \varepsilon_{\mathbf{i}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) \tag{4.77}$$

and

$$S_{2N} = \frac{r_N^4}{2} \sum_{\mathbf{k}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-2} \sum_{\mathbf{i}} \sum_{\mathbf{j}} \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{j}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{j}} - \boldsymbol{\lambda}_{\mathbf{k}})). \tag{4.78}$$

The addend S_{1N} given by (4.77) can be decomposed as:

$$\begin{aligned}
&S_{1N} = r_N^2 \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}}^2 \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{k}})) \\
&+ r_N^2 \sum_{\mathbf{k} \neq \mathbf{i}} \varepsilon_{\mathbf{k}} \varepsilon_i \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) = \\
&= r_N^2 \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}}^2 \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} K(H^{-1/2}\mathbf{0}) \\
&+ r_N^2 \sum_{\mathbf{k} \neq \mathbf{i}} \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) = S_{1N}^1 + S_{1N}^2.
\end{aligned}$$

For the first addend,

$$\begin{aligned}
&S_{1N}^1 = r_N^2 \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}}^2 \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} K(H^{-1/2}\mathbf{0}) = \\
&= \frac{1}{N|H|^{1/2}} \sum_{\mathbf{k}} 4\pi^2 E^{-1}(q_2(m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}), Y_{\mathbf{k}}^{**})) K(H^{-1/2}\mathbf{0}) \to_{\mathbb{P}} \frac{4\pi^2}{|H|^{1/2}} K(\mathbf{0}).
\end{aligned}$$

Therefore:

$$S_{1N} \approx \frac{4\pi^2}{|H|^{1/2}} K(\mathbf{0}) + S_{1N}^2,$$

with

$$S_{1N}^2 = r_N^2 \sum_{\mathbf{k} \neq \mathbf{i}} \varepsilon_{\mathbf{k}} \varepsilon_{\mathbf{i}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})).$$

Note that:

$$E(\varepsilon_{\mathbf{k}}^2 / \boldsymbol{\lambda} = \boldsymbol{\lambda}_{\mathbf{k}}) = -E(q_2(m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}), Y_{\mathbf{k}}^{**}),$$

and consider the following decomposition for (4.78):

$$\begin{aligned}
&S_{2N} = \frac{r_N^4}{2} \sum_{\mathbf{k}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-2} \sum_{\mathbf{i}=\mathbf{j}} \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{j}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{j}} - \boldsymbol{\lambda}_{\mathbf{k}})) \\
&+ \frac{r_N^4}{2} \sum_{\mathbf{k}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-2} \sum_{\mathbf{i} \neq \mathbf{j}} \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{j}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{j}} - \boldsymbol{\lambda}_{\mathbf{k}})) = \\
&S_{2N}^1 + S_{2N}^2.
\end{aligned}$$

The first part S_{2N}^1 converges in probability to:

$$S_{2N}^1 = \frac{r_N^4}{2} \sum_{\mathbf{k}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-1} \sum_{\mathbf{i}} \varepsilon_{\mathbf{i}}^2 K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) \to_{\mathbb{P}} \frac{4\pi^2}{2|H|^{1/2}} \int K^2(\mathbf{u}) d\mathbf{u}.$$

The addend S_{2N}^2 can be decomposed in two parts:

$$\begin{aligned} S_{2N}^2 &= \frac{r_N^4}{2} \sum_{\mathbf{k}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-2} \sum_{\mathbf{i} \neq \mathbf{j}} \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{j}} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{j}} - \boldsymbol{\lambda}_{\mathbf{k}})) \\ &= \frac{r_N^4}{2} K(\mathbf{0}) \sum_{\mathbf{k}} \sum_{\mathbf{j}} \varepsilon_{\mathbf{k}} \varepsilon_{\mathbf{j}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-2} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{j}} - \boldsymbol{\lambda}_{\mathbf{k}})) + \\ &\frac{r_N^4}{2} \sum_{\mathbf{i},\mathbf{j}} \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{j}} \sum_{\mathbf{k} \neq \mathbf{i}, \mathbf{k} \neq \mathbf{j}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-2} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{k}})) K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{j}} - \boldsymbol{\lambda}_{\mathbf{k}})) \\ &= S_{2N}^{21} + S_{2N}^{22}. \end{aligned}$$

The variance of the first addend can be bounded by:

$$\begin{aligned} Var(S_{2N}^{21}) &= Var\left(\frac{r_N^4}{2} K(\mathbf{0}) \sum_{\mathbf{k}} \sum_{\mathbf{j}} \varepsilon_{\mathbf{k}} \varepsilon_{\mathbf{j}} q_2^{\mathbf{k}} \Gamma(\boldsymbol{\lambda}_{\mathbf{k}})^{-2} K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{j}} - \boldsymbol{\lambda}_{\mathbf{k}})) \right) \\ &= \mathcal{O}(N^{-2}|H|^{-3/2}) = o(|H|^{-1/2}), \end{aligned}$$

therefore, this addend is

$$S_{2N}^{21} = o_{\mathbb{P}}(|H|^{-1/4}).$$

Then, in the expression of T_{LK}^* we have:

$$T_{LK}^* \approx \mu_H + R_{1N} + R_{2N} + R_{3N} + S_{1N}^2 + S_{2N}^{22} = \mu_H + R_{1N} + R_{2N} + R_{3N} + \frac{1}{2} W_N |H|^{-1/4},$$

where

$$W_N = \frac{|H|^{1/4}}{N} \sum_{\mathbf{i} \neq \mathbf{j}} \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{j}} \Gamma(\boldsymbol{\lambda}_{\mathbf{j}})^{-1} \left(2K_H(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{j}}) - K_H * K_H(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{j}}) \right).$$

Besides, if we define, for $\mathbf{i} \neq \mathbf{j}$:

$$b(\mathbf{i}, \mathbf{j}) = \frac{|H|^{1/4}}{N} (2K_H(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{j}}) - K_H * K_H(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{j}})) \Gamma(\boldsymbol{\lambda}_{\mathbf{j}})^{-1},$$

and $b(\mathbf{i}, \mathbf{i}) = 0$. Then, W_N can be written as:

$$W_N = \sum_{\mathbf{i}} \sum_{\mathbf{j}} b(\mathbf{i}, \mathbf{j}) \varepsilon_{\mathbf{i}} \varepsilon_{\mathbf{j}}.$$

In order to prove the asymptotic normal distribution of W_N, we will apply Proposition 3.2 in de Jong (1987). For that purpose, we must write W_N as a quadratic form of indpendent random

variables, namely $W_N = \sum_{i<j} c_{i,j}\varepsilon_i\varepsilon_j$, where i and j are one-dimensional indexes.

As it is done in the proof of Theorem 1, define a new subindex for the Fourier frequencies $\boldsymbol{\lambda}_{\mathbf{k}}$, with $\mathbf{k} = (k_1, k_2)$ and $k_l = 0, \pm 1, \ldots, \pm m_l$, for $l = 1, 2$. Consider $\boldsymbol{\lambda}_{\mathbf{k}} = \boldsymbol{\lambda}_{\mathbf{k}'}$ where $\mathbf{k}' = (k_1', k_2')$, with $k_l' = 1, \ldots, m_l' = 2m_l + 1$, in such a way that $k_l' = k_l + m_l + 1$ for $l = 1, 2$. Let $M = m_1' \times m_2'$. The new coefficients, with one dimensional indexes, are given by the following matrix:

$$A = (a_{ij}), \quad A \in \mathcal{M}_{M\times M},$$

and each entry of this matrix is defined by $a_{ij} = b_{\mathbf{ij}}$ and $a_{ii} = 0$, where the bidimensional indexes $\mathbf{i}$ determine unidimensional indexes i such that:

$$\mathbf{i} = (i_1, i_2), \quad \text{if } (i_1 - 1)m_2' \leq i \leq i_1 m_2' \quad \text{and } i = (i_1 - 1)m_2' + i_2, \tag{4.79}$$

With this definitions, W_N can be written as a quadratic form with one-dimensional indexes:

$$W_N = \sum_{i,j} a_{i,j}\varepsilon_i\varepsilon_j.$$

For a_{ij} we have that:

$$a_{ij} = \frac{|H|^{1/4}}{N} 2K_H(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_j)\Gamma(\boldsymbol{\lambda}_j)^{-1} - \frac{|H|^{1/4}}{N} K_H * K_H(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_j)\Gamma(\boldsymbol{\lambda}_j)^{-1} = a_{i,j}^1 - a_{i,j}^2,$$

where

$$a_{ij}^1 = \frac{|H|^{1/4}}{N} 2K_H(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_j)\Gamma(\boldsymbol{\lambda}_j)^{-1}, \quad a_{ij}^2 = \frac{|H|^{1/4}}{N} K_H * K_H(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_j)\Gamma(\boldsymbol{\lambda}_j)^{-1}.$$

Now, if we define:

$$c_{ij}^1 = a_{ij}^1, \quad c_{ij}^2 = c_{ji}^1,$$
$$c_{ij}^3 = a_{ij}^2, \quad c_{ij}^4 = c_{ji}^3.$$

Define also, $W_{ij} = \left(c_{ij}^1 + c_{ij}^2 - c_{ij}^3 - c_{ij}^4\right)\varepsilon_i\varepsilon_j$. Then, W_N can be wrriten as:

$$W_N = \sum_{i<j} W_{ij}.$$

The variance of this form is given by (4.20). In order to apply Proposition 3.2 in (de Jong (1987)), we must check some conditions on W_N. The first one is the W_N is clean, but this is clear, by definition (see definition 2.1 in de Jong (1987)). Consider:

$$G_I = \sum_{i<j} W_{ij}^4,$$
$$G_{II} = \sum_{i<j<k} \left\{E\left(W_{ij}^2 W_{ik}^2\right) + E\left(W_{ji}^2 W_{jk}^2\right) + E\left(W_{ki}^2 W_{kj}^2\right)\right\},$$
$$G_{III} = \sum_{i<j<k<l} \left\{E\left(W_{ij}W_{ik}W_{lj}W_{lk}\right) + E\left(W_{ij}W_{il}W_{kj}W_{kl}\right) + E\left(W_{ik}W_{il}W_{jk}W_{jl}\right)\right\}.$$

We must check that G_I, G_{II} and G_{III} are of smaller order than $Var(W_N)$, which is given by (4.20). It is easy to see that $G_I = \mathcal{O}(N^{-2}|H|^{-1/2})$, just taking into account that:

$$E\left(a_{ij}^1\varepsilon_i\varepsilon_j\right)^4 = \mathcal{O}\left(\frac{|H|}{N^4}16K_H^4(\boldsymbol{\lambda_i}-\boldsymbol{\lambda_j})\Gamma(\boldsymbol{\lambda_j})^{-4}\right) = \mathcal{O}\left(\frac{1}{N^4|H|^{1/2}}\right),$$

and $E(a_{ij}^3\varepsilon_i\varepsilon_j)^4 = \mathcal{O}(N^{-4})$. G_{II} is $\mathcal{O}(N^{-1}|H|^{-1/2})$, since:

$$E\left((a_{ij}^1)^2(a_{ij}^2)^2\right) = \mathcal{O}\left(\frac{1}{N^4|H|^{1/2}}\right).$$

Similar computations lead to $G_{III} = \mathcal{O}\left(|H|^{1/2}\right)$. Then, we have that $W_N \to N(0,\sigma^2)$. Finally, we must find a bound for the residual terms R_{1N}, R_{2N} and R_{3N} in (4.75), (4.76) and (4.76), respectively and $H_N(\boldsymbol{\lambda})$ is given by (4.73). We can see that both R_{1N} and R_{3N} are stochastically bounded. In fact $R_{1N} = N^{1/2}|H|R_{1N0}$, where

$$R_{1N0} = \frac{1}{\sqrt{N}}\sum_{\mathbf{k}}\varepsilon_{\mathbf{k}}\int \mathbf{u}^T H_{m_\theta}(\boldsymbol{\lambda_k})\mathbf{u}K(\mathbf{u})d\mathbf{u}$$

and $R_{3N} = N^{1/2}|H|R_{3N0}$, where

$$R_{3N0} = \frac{-1}{\sqrt{N}}\sum_{\mathbf{k}}\varepsilon_{\mathbf{k}}\Gamma(\boldsymbol{\lambda_k})^{-1}\int\int \mathbf{s}^T H_{m_\theta}(\boldsymbol{\lambda_k})\mathbf{s}\mathbf{u}^T H_{m_\theta}(\boldsymbol{\lambda_k})\mathbf{u}K(\mathbf{s})K(\mathbf{u})d\mathbf{s}d\mathbf{u}.$$

Both R_{1N0} and R_{3N0} are asymptotically normal, and therefore, stochastically bounded. The remaining residual term, R_{2N} admits the following asymptotic expression:

$$R_{2N} = \frac{-|H|^2}{8}\sum_{\mathbf{k}}\frac{1}{f_\theta(\boldsymbol{\lambda_k})}\int\int \mathbf{s}^T H_{m_\theta}(\boldsymbol{\lambda_k})\mathbf{s}(\mathbf{s}+\mathbf{u})^T H_{m_\theta}(\boldsymbol{\lambda_k})(\mathbf{s}+\mathbf{u})K(\mathbf{s})K(\mathbf{u})d\mathbf{s}d\mathbf{u}.$$

An additional term of the bias, b_H is obtained from R_{2N}, as $N^{1/2}|H| \to \infty$. □

The following lemmas are needed for bounding B_1, B_2 and B_3 in Lemma 13.

Lemma 14. *Define*

$$\Psi_N(\boldsymbol{\lambda}) = |H|^{1/2}(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(e^{Y_{\mathbf{k}}^{**}-\bar{m}_{\mathbf{k}}}-1)W_{\mathbf{k}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda_k}) \tag{4.80}$$

By Taylor's expansion and conditions in Theorem 3, the following hold also uniformly in $\boldsymbol{\lambda}$:

$$|H|^{1/2}l(\boldsymbol{\beta}) = \Psi_N(\boldsymbol{\lambda})^T\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}^T A\boldsymbol{\beta} + \Delta_1(\boldsymbol{\beta}),$$

where $l(\boldsymbol{\beta})$ is given by:

$$l(\boldsymbol{\beta}) = \sum_{\mathbf{k}}\left[-(N|H|^{1/2})^{-1/2}\boldsymbol{\beta}^T W_{\mathbf{k}} - e^{Y_{\mathbf{k}}^{**}-\bar{m}_{\mathbf{k}}-(N|H|^{1/2})^{-1/2}\boldsymbol{\beta}^T W_{\mathbf{k}}} + e^{Y_{\mathbf{k}}^{**}-\bar{m}_{\mathbf{k}}}\right]K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda_k}).$$

Proof. The expression for $l(\beta)$ is given by:

$$|H|^{1/2} l(\boldsymbol{\beta}) = \\ |H|^{1/2} \sum_{\mathbf{k}} \left\{ -(N|H|^{1/2})^{-1/2} \boldsymbol{\beta}^T W_{\mathbf{k}} - e^{Y_{\mathbf{k}}^{**} - (N|H|^{1/2})^{-1/2} \boldsymbol{\beta}^T W_{\mathbf{k}}} + e^{Y_{\mathbf{k}}^{**} - \bar{m}_k} \right\} K_H(\boldsymbol{\lambda}_k - \boldsymbol{\lambda}),$$

and by Taylor's expansion, we can write:

$$e^{Y_{\mathbf{k}}^{**} - (N|H|^{-1/2})^{-1/2} \boldsymbol{\beta}^T W_{\mathbf{k}}} = e^{Y_{\mathbf{k}}^{**} - \bar{m}_{\mathbf{k}}} + (N|H|^{-1/2})^{-1/2} \boldsymbol{\beta}^T W_{\mathbf{k}} e^{Y_{\mathbf{k}}^{**} - \bar{m}_{\mathbf{k}}} + \\ \frac{1}{2} (N|H|^{-1/2})^{-1} \boldsymbol{\beta}^T W_{\mathbf{k}} W_{\mathbf{k}}^T \boldsymbol{\beta} e^{Y_{\mathbf{k}}^{**} - c_{\mathbf{k}}},$$

where $c_{\mathbf{k}}$ is such that $|(N|H|^{-1/2})^{-1/2} \boldsymbol{\beta}^T W_{\mathbf{k}} - \bar{m}_{\mathbf{k}}| \geq |c_{\mathbf{k}} - \bar{m}_{\mathbf{k}}|$. Then, $|H|^{1/2} l(\boldsymbol{\beta})$ is given by:

$$\begin{aligned} & |H|^{1/2} \sum_{\mathbf{k}} \left\{ -(N|H|^{1/2})^{-1/2} \boldsymbol{\beta}^T W_{\mathbf{k}} - (N|H|^{-1/2})^{-1/2} \boldsymbol{\beta}^T W_{\mathbf{k}} e^{Y_{\mathbf{k}}^{**} - \bar{m}_k} \right. \\ & \left. -\frac{1}{2} (N|H|^{1/2})^{-1} \boldsymbol{\beta}^T W_{\mathbf{k}} W_{\mathbf{k}}^T \boldsymbol{\beta} (e^{Y_{\mathbf{k}}^{**} - c_{\mathbf{k}}} - 1 + 1) \right\} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) \\ & = |H|^{1/2} (N|H|^{1/2})^{-1/2} \sum_{\mathbf{k}} (e^{Y_{\mathbf{k}}^{**} - \bar{m}_k} - 1) W_{\mathbf{k}}^T K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) \boldsymbol{\beta} \\ & -\frac{1}{2} |H|^{1/2} (N|H|^{1/2})^{-1} \sum_{\mathbf{k}} \boldsymbol{\beta}^T W_{\mathbf{k}} W_{\mathbf{k}}^T \boldsymbol{\beta} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) \\ & -\frac{1}{2} |H|^{1/2} (N|H|^{1/2})^{-1} \sum_{\mathbf{k}} \boldsymbol{\beta}^T W_{\mathbf{k}} W_{\mathbf{k}}^T \boldsymbol{\beta} (e^{Y_{\mathbf{k}}^{**} - c_{\mathbf{k}}} - 1) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) \\ & = \Psi_N(\boldsymbol{\lambda})^T \boldsymbol{\beta} + \frac{1}{2N} \sum_{\mathbf{k}} \boldsymbol{\beta}^T W_{\mathbf{k}} W_{\mathbf{k}}^T \boldsymbol{\beta} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) + \Delta_1(\boldsymbol{\beta}) \\ & = \Psi_N(\boldsymbol{\lambda})^T \boldsymbol{\beta} + \frac{1}{2N} \sum_{\mathbf{k}} \boldsymbol{\beta}^T A_{\mathbf{k}} \boldsymbol{\beta} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) + \Delta_1(\boldsymbol{\beta}). \end{aligned}$$

The matrix $A_{\mathbf{k}}$ is given by $W_{\mathbf{k}} W_{\mathbf{k}}^T$:

$$A_{\mathbf{k}} = W_{\mathbf{k}} W_{\mathbf{k}}^T = \begin{pmatrix} 1 & |H|^{-1/2} (\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})^T \\ |H|^{-1/2} (\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})^T & |H|^{-1/4} (\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})^T \end{pmatrix}$$

and it converges in probability to

$$\frac{1}{N} \sum_{\mathbf{k}} A_{\mathbf{k}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}) \to_{\mathbb{P}} A,$$

where A is given by

$$A = \frac{-1}{4\pi^2} \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \int \mathbf{u}\mathbf{u}^T K(\mathbf{u}) d\mathbf{u} \end{pmatrix}.$$

The residual term $\Delta_1(\boldsymbol{\beta})$ is $\mathcal{O}_{\mathbb{P}}(1)$, provided that $N^{(\zeta-1)/\zeta} |H|^{1/2} \geq c_0 \log N$:

$$\Delta_1(\boldsymbol{\beta}) = \frac{-|H|^{1/2}}{2} (N|H|^{1/2})^{-1} \sum_{\mathbf{k}} \boldsymbol{\beta}^T W_{\mathbf{k}} W_{\mathbf{k}}^T \boldsymbol{\beta} (e^{Y_{\mathbf{k}}^{**} - c_{\mathbf{k}}} - 1) K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}).$$

Since $|r_N\boldsymbol{\beta}^T W_{\mathbf{k}} - \bar{m}_{\mathbf{k}}| \geq |c_{\mathbf{k}} - \bar{m}_{\mathbf{k}}|$, $c_{\mathbf{k}}$ can be written as $c_{\mathbf{k}} = \boldsymbol{\beta}^T W_{\mathbf{k}} + \alpha^T W_{\mathbf{k}}$, where $\|\alpha\| \leq c_1 r_N$, for some $c_1 > 0$. Then, $\Delta_1(\boldsymbol{\beta})$ can be decomposed in two addends:

$$\Delta_1^1(\boldsymbol{\beta}) = \frac{-|H|^{1/2}}{2}(N|H|^{1/2})^{-1}\sum_{\mathbf{k}} q_2(\boldsymbol{\beta}^T W_{\mathbf{k}} + \alpha^T W_{\mathbf{k}}, Y_{\mathbf{k}}^{**})\boldsymbol{\beta}^T W_{\mathbf{k}} W_{\mathbf{k}}^T \boldsymbol{\beta} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}),$$

$$\Delta_1^2(\boldsymbol{\beta}) = \frac{|H|^{1/2}}{2}(N|H|^{1/2})^{-1}\sum_{\mathbf{k}} \boldsymbol{\beta}^T W_{\mathbf{k}} W_{\mathbf{k}}^T \boldsymbol{\beta} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}).$$

Uniform bound $\mathcal{O}_{\mathbb{P}}(1)$ for both addends is obtained from condition on $\Phi_{n,j}$, for the particular case of $j = 1$ and $j = 3$. Then, the expression for $l(\boldsymbol{\beta})$ is proved. □

Lemma 15. *If f_θ is twice differentiable, we have the following representation for the difference between the non parametric estimation $\hat{m}_{LK}^*$ and the log-spectral density under the null hypothesis m_θ, in a frequency $\boldsymbol{\lambda}$ and under conditions in Theorem 3:*

$$\hat{m}_{LK}^*(\boldsymbol{\lambda}) - m_\theta(\boldsymbol{\lambda}) = r_N^2\Gamma(\boldsymbol{\lambda})^{-1}\sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} K\left(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})\right)(1 + o_{\mathbb{P}}(1)) + H_N(\boldsymbol{\lambda}) \tag{4.81}$$

and H_N is given by (4.73).

Proof. Using the expression for $l(\boldsymbol{\beta})$ obtained in Lemma 14 and applying the convexity lemma of Pollard (1991) we obtain the maximizer $\hat{\boldsymbol{\beta}}$ of the expression for $l(\boldsymbol{\beta})$ is given by

$$\hat{\boldsymbol{\beta}} = B^{-1}\Psi_N(\boldsymbol{\lambda}) + o_{\mathbb{P}}(1).$$

The inverse of matrix A is given by:

$$B^{-1} = -\pi^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & b_2 & -b_{12} \\ 0 & -b_{12} & b_1 \end{pmatrix},$$

$$b_j = \frac{\int u_j^2 K(\mathbf{u})d\mathbf{u}}{a_1}, j = 1, 2, \quad b_{12} = \frac{-\int u_1 u_2 K(\mathbf{u})d\mathbf{u}}{a_1},$$

$$a_1 = \int u_1^2 K(\mathbf{u})d\mathbf{u} \int u_2^2 K(\mathbf{u})d\mathbf{u} - \left(\int u_1 u_2 K(\mathbf{u})d\mathbf{u}\right)^2,$$

where u_1 and u_2 denote the first and the second components of vector $\mathbf{u} \in \mathbb{R}^2$. The first component of $\boldsymbol{\beta}$ is

$$\hat{\beta}_{(1)} = (N|H|^{1/2})^{-1/2}(\hat{m}_{LK}(\boldsymbol{\lambda}) - m_\theta(\boldsymbol{\lambda})).$$

We obtain, from the expression for $\Psi_n(\boldsymbol{\lambda})$ in (4.80):

$$\Psi_N(\boldsymbol{\lambda}) = |H|^{1/2}(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(e^{Y^{**}_{\mathbf{k}}-\bar{m}_{\mathbf{k}}}-1)W_{\mathbf{k}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}) =$$
$$|H|^{1/2}(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(e^{Y^{**}_{\mathbf{k}}-\bar{m}_{\mathbf{k}}}-1)W_{\mathbf{k}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}) =$$
$$|H|^{1/2}(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(e^{Y^{**}_{\mathbf{k}}-\boldsymbol{\beta}^T W_{\mathbf{k}}}-1)W_{\mathbf{k}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}) =$$
$$|H|^{1/2}(n|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(e^{Y^{**}_{\mathbf{k}}-\boldsymbol{\beta}^T W_{\mathbf{k}}}+e^{Y^{**}_{\mathbf{k}}-m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}-e^{Y^{**}_{\mathbf{k}}-m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}-1)W_{\mathbf{k}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}) =$$
$$|H|^{1/2}(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(e^{Y^{**}_{\mathbf{k}}-m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}-1)W_{\mathbf{k}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}) +$$
$$|H|^{1/2}(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(e^{Y^{**}_{\mathbf{k}}-\boldsymbol{\beta}^T W_{\mathbf{k}}}-e^{Y^{**}_{\mathbf{k}}-m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})})W_{\mathbf{k}}K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}).$$

The result is proved just considering the first component of $\Psi_n(\boldsymbol{\lambda})$.

$$\Psi^{(1)}_N(\boldsymbol{\lambda}) = |H|^{1/2}(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(e^{Y^{**}_{\mathbf{k}}-m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}-1)K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}) +$$
$$|H|^{1/2}(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(e^{Y^{**}_{\mathbf{k}}-\boldsymbol{\beta}^T W_{\mathbf{k}}}-e^{Y^{**}_{\mathbf{k}}-m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})})K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}) =$$
$$(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}\varepsilon_{\mathbf{k}}K(H^{-1/2}(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})) +$$
$$(N|H|^{1/2})^{-1/2}\sum_{\mathbf{k}}(q_1(\boldsymbol{\beta}^T W_{\mathbf{k}}, Y^{**}_{\mathbf{k}})-\varepsilon_k)K(H^{-1/2}(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})).$$

□

Lemma 16. *Under assumption (1)-(3), we have*

$$\sup_{\boldsymbol{\lambda}\in[0,\pi]^2}|\hat{m}_{LK}(\boldsymbol{\lambda})-\hat{m}^*_{LK}(\boldsymbol{\lambda})| = O_{\mathbb{P}}(N^{-1/2}\log N)$$

Proof. The proof of this Lemma is obatined using similar arguments as that for the proof of Lemma 15. Recall the expression for the local loglikelihood given by:

$$\sum_{\mathbf{k}}\left[Y_{\mathbf{k}}-a-b^T(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})-e^{Y_{\mathbf{k}}-a-b^T(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})}\right]K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}). \tag{4.82}$$

This expression can be written in terms of the vector $\boldsymbol{\beta}_2$ as

$$L(\boldsymbol{\beta}_2) = \sum_{\mathbf{k}}\left[Y_{\mathbf{k}}-\bar{m}_{\mathbf{k}}-(N|H|^{1/2})^{-1/2}\boldsymbol{\beta}_2^T W_{\mathbf{k}}-e^{Y_{\mathbf{k}}-\bar{m}_{\mathbf{k}}-(N|H|^{1/2})^{-1/2}\boldsymbol{\beta}_2^T W_{\mathbf{k}}}\right]K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}})$$

and the difference:

$$L(\boldsymbol{\beta}_2)-L(0) = \sum_{\mathbf{k}}\left[-(N|H|^{1/2})^{-1/2}\boldsymbol{\beta}_2^T W_{\mathbf{k}}-e^{Y_{\mathbf{k}}-\bar{m}_{\mathbf{k}}-(N|H|^{1/2})^{-1/2}\boldsymbol{\beta}_2^T W_{\mathbf{k}}}\right]K_H(\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathbf{k}}).$$

If we set

$$U_n(\boldsymbol{\beta}_2) = \sum_{\mathbf{k}} r_N(\boldsymbol{\lambda}_{\mathbf{k}}) \left[e^{-\bar{m}_{\mathbf{k}} - (n|H|^{1/2})^{-1/2} \boldsymbol{\beta}_2^T W_{\mathbf{k}}} - e^{-\bar{m}_{\mathbf{k}}} \right] K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}})$$

Then,

$$L(\boldsymbol{\beta}_2) - L(0) = l(\boldsymbol{\beta}_2) - U_N(\boldsymbol{\beta}_2)$$

An uniform bound for $U_N(\boldsymbol{\beta}_2)$ is easily found just by Taylor's expansion and using the bound $\max_{\mathbf{k}} |r_N(\boldsymbol{\lambda}_{\mathbf{k}})| = \mathcal{O}_{\mathbb{P}}(N^{-1/2} \log N)$:

$$|H|^{1/2} \sup_{\boldsymbol{\lambda} \in [0,\pi]^2} U_N(\boldsymbol{\beta}_2) = O_{\mathbb{P}}(|H|^{-1/4} \log N)$$

With the same arguments as in Lemma 15, we show that the following bounds also hold uniformly in $\boldsymbol{\lambda}$:

$$|H|^{1/2}(l(\boldsymbol{\beta}_2) - U_N(\boldsymbol{\beta}_2)) = |H|^{1/2}(L(\boldsymbol{\beta}_2) - L(\mathbf{0})) = \Psi_N(\boldsymbol{\lambda})^T \boldsymbol{\beta}_2 + \frac{1}{2} \boldsymbol{\beta}_2^T A \boldsymbol{\beta}_2 + \Delta_2(\boldsymbol{\beta}_2)$$

$$\Delta_1(\boldsymbol{\beta}_2) = O_{\mathbb{P}}(1), \qquad \Delta_2(\boldsymbol{\beta}_2) = O_{\mathbb{P}}(1)$$

$$\nabla \Delta_1(\boldsymbol{\beta}_2) = O_{\mathbb{P}}((N|H|^{1/2})^{-1/2} \log |H|^{1/2} \alpha_N + |H|)$$

$$\Delta_2(\boldsymbol{\beta}_2) = \nabla \Delta_1(\boldsymbol{\beta}_2) + O_{\mathbb{P}}(|H|^{1/4} \log N)$$

where $\alpha_N \to \infty$. Using the same arguments as that for the proof of Theorem 2 in Carroll *et al.* (1997) and the proof of the quadratic approximation lemma in Fan and Gijbels (1995) we obtain:

$$(N|H|^{1/2})^{1/2} \{\hat{m}_{LK}(\boldsymbol{\lambda}) - m(\boldsymbol{\lambda})\} = (\pi^2, 0, 0) \Psi_N(\boldsymbol{\lambda}) + O_{\mathbb{P}}\left(|H|^{1/4} \log N \right), \tag{4.83}$$

and

$$(N|H|^{1/2})^{1/2} \{\hat{m}^{**}_{LK}(\boldsymbol{\lambda}) - m(\boldsymbol{\lambda})\} = (\pi^2, 0, 0) \Psi_N(\boldsymbol{\lambda}) + O_{\mathbb{P}}\left(\frac{\log |H|^{1/2}}{\sqrt{N|H|^{1/2}}} \alpha_N + |H| \right). \tag{4.84}$$

□

Lemma 17. *Assume that $\varepsilon_1, \ldots, \varepsilon_N$ are independent identically distributed random variables, with $E(\varepsilon_1) = 0$ and $E(|\varepsilon_1|^s) < \infty$, for some $s > 2$. Assume that $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are fixed design points in $[0,1]^2 \subset \mathbb{R}^2$ such that $\mathbf{x}_i \in A_i \subset \mathbb{R}^2$, $\cup_{i=1}^N A_i = [0,1]^2$, $A_i \cap A_j = \oslash$, where A_i is Jordan-measurable, with $\max_i \mu(A_i) = \mathcal{O}(N^{-1})$, where μ is the Jordan measure and*

$$\max_i d(A_i) = \mathcal{O}(N^{-1}),$$

where $d(B) = \sup_{\mathbf{x},\mathbf{y} \in B} \|\mathbf{x} - \mathbf{y}\|$, $\|\cdot\|$ is the L^2-norm. Assume that W is a weight function satisfying a Lipchitz condition and:

$$\max_i |W_i(\mathbf{x}))| \geq c_0 N^{-1},$$

uniformly in $\mathbf{x} \in [0,1]^2$*, for a constant* c_0*. Finally assume that there is a sequence* $\alpha_N \to 0$ *and constants* $\eta \in (0, s-2)$*,* $c > 1/2$ *such that, for all* $\mathbf{x} \in [0,1]^2$*:*

$$N^{2/(s-\eta)} \max_i |W_i(\mathbf{x})| logN \leq \alpha_N c, \quad \textit{and} \quad \left(\sum_i W_i(\mathbf{x}) \log N \right)^2 \leq \alpha_N c. \tag{4.85}$$

Then:

$$\sup_{\mathbf{x}\in[0,1]^2} \left| \sum_i W_i(\mathbf{x}) \varepsilon_i \right| = \mathcal{O}(\alpha_N).$$

Proof. This lemma is a straightforward extension of Theorem 11.2 in Müller (1988). The proof is similar, since the stochastic part is not affected by the dimension. □

Lemma 18. *Assume conditions in Lemma 17 hold and suppose that the weight functions are kernel weights:*

$$W_i(\mathbf{x}) = |H|^{-1/2} K(H^{-1/2}(\mathbf{x} - \mathbf{x}_i)).$$

Then,

$$\sup_{\mathbf{x}\in[0,1]^2} \frac{1}{N} \left| \sum_i K_H(\mathbf{x} - \mathbf{x}_i) \varepsilon_i \right| = o((N|H|^{1/2})^{-1/2} (-\log |H|^{1/2}) \beta_N),$$

where the sequence $\beta_N \to \infty$ *and provided that there exists* $s > 2$*,* $\eta \in (0, s-2)$ *such that* $N^{2/(s-\eta)} |H|^{-1/2} \log N \to C$*, for some constant* C*.*

Proof. The proof is inmediate from Lemma 17. The condition on s and η is obtained from the restriction (4.85) on the kernel weights. □

Lemma 19. *The terms* B_1*,* B_2 *and* B_3 *in Lemma 13 are bounded by:*

$$B_1 = \mathcal{O}_{\mathbb{P}} \left(\frac{\log N}{\sqrt{N}} |H|^{-1/2} \log |H|^{1/2} \alpha_N \right), + \mathcal{O}_{\mathbb{P}}(\log^2 N)$$

$$B_2 = \mathcal{O}_{\mathbb{P}}(\log^2 N),$$

$$B_3 = \mathcal{O}_{\mathbb{P}}(|H|^{-1/4} \log N (-\log |H|^{1/2}) \alpha_N),$$

where $\alpha_N \to \infty$*.*

Proof. Recall the expression for B_2 is given by:

$$B_2 = \sum_{\mathbf{k}} e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} \left(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) \right)^2.$$

By Taylor's expansion of $e^{\mathbf{x}}$ and Lemma 16:

$$
\begin{aligned}
&B_2 \\
&= \sum_{\mathbf{k}} [1 + Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) + \mathcal{O}_{\mathbb{P}}((Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^2)] \cdot (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^2 \approx \\
&\sum_{\mathbf{k}} (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^2 + \sum_{\mathbf{k}} (Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))^2 \leq \\
&\leq N \sup_{\mathbf{k}} |\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})|^2 = \mathcal{O}_{\mathbb{P}}(\log^2 N).
\end{aligned}
$$

Just taking into account that:

$$
e^{Y_{\mathbf{k}}} = I(\boldsymbol{\lambda}_{\mathbf{k}}) = f_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) V_{\mathbf{k}} + r_N(\boldsymbol{\lambda}_{\mathbf{k}}) = V_{\mathbf{k}} e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} + r_N(\boldsymbol{\lambda}_{\mathbf{k}})
$$

the term B_1 is decomposed in two addends:

$$
\begin{aligned}
B_1 &= \sum_{\mathbf{k}} (1 - e^{Y_{\mathbf{k}} - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}) \, (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) \\
&= \sum_{\mathbf{k}} (1 - V_{\mathbf{k}} e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})})(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) + \\
&\sum_{\mathbf{k}} r_N(\boldsymbol{\lambda}_{\mathbf{k}}) e^{-\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) = \\
&B_{1,1} + B_{1,2}.
\end{aligned}
$$

The second addend, $B_{1,2}$ can be bounded by:

$$
\begin{aligned}
&B_{1,2} \leq \sum_{\mathbf{k}} \sup_{\mathbf{k}} |r_N(\boldsymbol{\lambda}_{\mathbf{k}})| \cdot |\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})| \leq \\
&\sum_{\mathbf{k}} \mathcal{O}_{\mathbb{P}}(N^{-1/2} \log N) \sup_{\mathbf{k}} |\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})| = \mathcal{O}_{\mathbb{P}}\left(\log^2 N\right).
\end{aligned}
$$

For the first addedn, $B_{1,1}$, applying Taylor's expansion on $e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}$ around the origin, we have:

$$
\begin{aligned}
B_{1,1} &= \sum_{\mathbf{k}} \left(1 - V_{\mathbf{k}} e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}\right) (\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) \\
&= \sum_{\mathbf{k}} (1 - V_{\mathbf{k}} - V_{\mathbf{k}}(m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) e^{c_{\mathbf{k}}})(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})),
\end{aligned}
$$

where $c_{\mathbf{k}}$ satisfies that $|m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})| \geq |c_{\mathbf{k}}|$. Then,

$$
\begin{aligned}
B_{1,1} &= \sum_{\mathbf{k}} (V_{\mathbf{k}} - 1)(\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) \\
&- \sum_{\mathbf{k}} V_{\mathbf{k}} (m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))(\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) e^{c_{\mathbf{k}}}.
\end{aligned}
$$

The variables $V_{\mathbf{k}}$ are zero-mean, so the first term in $B_{1,2}$ is $o_{\mathbb{P}}(1)$. For the first addend, applying Lemma 16:

$$
\begin{aligned}
B_{1,1} &= \sum_{\mathbf{k}} V_{\mathbf{k}}(m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}\boldsymbol{\lambda}_{\mathbf{k}})(\hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}))e^{c_{\mathbf{k}}} \\
&\leq \mathcal{O}_{\mathbb{P}}(N^{-1/2}\log N)\sum_{\mathbf{k}}(m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - \hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})) \\
&= \mathcal{O}_{\mathbb{P}}(N^{-1/2}\log N)\sum_{\mathbf{k}}\Big((\pi^2,0,0)\Psi_N(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&+\mathcal{O}_{\mathbb{P}}((N|H|^{1/2})^{-1/2})\log|H|^{1/2}\alpha_N + |H|\Big)(N|H|^{1/2})^{-1/2} \\
&= \mathcal{O}_{\mathbb{P}}\left(\frac{\log N}{\sqrt{N}}\right)\sum_{\mathbf{k}}\left(\frac{|H|^{1/4}}{\sqrt{N}}\sum_{\mathbf{j}}(e^{Y^{**}_{\mathbf{j}}-\bar{m}_{\mathbf{j}}}-1)K_H(\boldsymbol{\lambda}_{\mathbf{k}}-\boldsymbol{\lambda}_{\mathbf{j}})\right. \\
&\left.+\mathcal{O}_{\mathbb{P}}((N|H|^{1/2})^{-1/2}\log|H|^{1/2}\alpha_N + |H|)\right),
\end{aligned}
$$

applying the expressions (4.83) and (4.84). By a Taylor's expansion on $e^{Y^{**}_{\mathbf{j}}-\bar{m}_{\mathbf{j}}}$ around the origin, Lemma 18 can be applied on the sum in the first addend. Now, applying Lemmas (6) and (16),

$$
B_{1,2} = \mathcal{O}_{\mathbb{P}}\left(\frac{\log N}{\sqrt{N}}|H|^{-1/2}\log|H|^{1/2}\alpha_N\right),
$$

For the last term, B_3, also using Lemmas 6, 16, and Lemma 18 in a Taylor's expansion for the expression of $\Psi_N(\boldsymbol{\lambda}_{\mathbf{k}})$ we derive:

$$
\begin{aligned}
&B_3 = \sum_{\mathbf{k}}\frac{R_N(\boldsymbol{\lambda}_{\mathbf{k}})}{f_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}\left\{e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})-\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}-1\right\} \leq \\
&N\max_{\mathbf{k}}\frac{R_N(\boldsymbol{\lambda}_{\mathbf{k}})}{f_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}\sup_{\mathbf{k}}\left|e^{m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})-\hat{m}^*_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})}-1\right| \leq \\
&N\max_{\mathbf{k}}\frac{R_N(\boldsymbol{\lambda}_{\mathbf{k}})}{f_\theta(\boldsymbol{\lambda}_{\mathbf{k}})}\left(N|H|^{1/2}\right)^{-1/2}\cdot \\
&\quad\sup_{\mathbf{k}}\left|(\pi^2,0,0)\Psi_N(\boldsymbol{\lambda}_{\mathbf{k}}) + \mathcal{O}_{\mathbb{P}}\left((N|H|^{1/2})^{-1/2}\log|H|^{1/2}\alpha_N + |H|\right)\right| = \\
&\mathcal{O}_{\mathbb{P}}\left(\log N|H|^{3/4}\right) + \mathcal{O}_{\mathbb{P}}\left((N|H|^{1/2})^{-1/2}\log N\log|H|^{1/2}\right) \\
&\qquad+\mathcal{O}_{\mathbb{P}}\left(|H|^{-1/4}\log N(-\log|H|^{1/2})\alpha_N\right) \\
&= \mathcal{O}_{\mathbb{P}}(|H|^{-1/4}\log N(-\log|H|^{1/2})\alpha_N).
\end{aligned}
$$

□

Chapter 5

Comparison of spatial dependence structures

Contents

In this part of our work, we will establish statistical methodology in order to analyze changes in the dependence structure for different spatial processes or for a process observed on a regular grid at different time moments. We propose a test statistic for testing the hypothesis $H_0 : f_1 = \ldots = f_L$, where each f_l denotes the spectral density of each observed process, for $l = 1, \ldots, L$.

As we have already seen, much effort has been devoted to the problem of estimating/modelling the dependence structure of spatial data, from both parametric and nonparametric approaches. This problem can be focused from the spatial domain, taking the variogram or the covariogram as the target function. From these two points of view, can be also focused the construction of goodness-of-fit tests problem, as we have pointed at the beginning of Chapter 4.

It is clear from the previous chapter that an alternative to these techniques is to consider the signal process and describe the dependence structure using the spectral density (instead of the covariance function). Considering this spectral scheme, we have already provided two test statistics, using distances on the spectral and on the log-spectral domain. These goodness-of-fit testing techniques take advantage of the representation of the spatial periodogram (as the response variable in a multiplicative regression model) and the spatial log-periodogram, which can be written as the exogenous variable in a regression model, where the regression function is the log-spectral density. Once again, we will use the representation of the log-periodogram for testing the equality of a set of log-spectral densities. For that purpose, we will adapt to our context some methodology from the regression setting.

In regression context, King *et al.* (1991) study the problem of comparing two regression curves under independence and Gaussian errors. The general case of comparing $L \geq 2$ regression curves is studied in Dette and Neumeyer (2001). Vilar-Fernández and González-Manteiga (2004) provide a goodness-of-fit technique for testing the equality of regression curves, under fixed design and dependent errors. The goal of this chapter is, based on the ideas in Vilar-Fernández and González-Manteiga (2004), to provide a test for testing the hypothesis that the spectral densities of L observations of a spatial random process are equal, without specifying a parametric model. The same applies for L observations from different processes. In spatial statistics, the design points for different realizations of a process are, in many cases, the same. For instance, when these realizations represent the evolution of the process along time. In this setting, it is not unreasonable to assume that observations are taken on the same set of locations, along time.

On this scope, Zhu *et al.* (2002) establish a statistical methodology to analyze changes in the spatial cumulative distribution function (SCDF), over time. Under shrinking asymptotics (a mix-

ture between increasing and infilling domain asymptotics), following Lahiri (1999), the authors prove asymptotic normal distribution of two test statistics, for comparing two time moments. The first statistic is based on the difference between the empirical estimators of the SCDF, at both time moments. The second statistic is a weighted integrated squared difference between the empirical counterparts of the SCDF. Both testing techniques are devoted to the detection of differences over time, but they are not specifically focused on the detection of changes in the dependence structure, as it is our purpose.

As a particular case, this technique allows to detect changes on the dependence structure of a process observed at different time moments. This capacity makes the technique relevant when studying spatio-temporal processes. Invariance of the spatial dependence along time makes feasible the use of stationary spatio-temporal dependence models (e.g. Fernández-Casal *et al.* (2003)).

The application of our technique is related to biomonitoring studies. Biomonitoring studies have been hold over the last years in order to determine levels of heavy metal concentration all over Europe. We will consider the mosses dataset introduced in Section 1.4.2, focusing on Selenium and Mercury concentrations. The contents of this chapter can be seen in Crujeiras *et al.* (2006a) and Crujeiras *et al.* (2006c).

5.1 Some background.

We will start introducing the notation for this part of the work, which is slightly different from that used in previous chapters. Just note that the number of the grid points (number of observations) and the number of Fourier frequencies have a different notation. Let Z_l be a zero mean second-order stationary spatial process, observed on a regular grid D_l, for $l = 1, \ldots, L$. That is, $\{Z_l(\mathbf{s}), \mathbf{s} \in D_l = \mathbf{a}_l + D\}$, with $D = \{1, \ldots, d_1\} \times \{1, \ldots, d_2\}$. The case $\mathbf{a}_1 = \ldots = \mathbf{a}_L$ implies that the processes are observed on the same grid of locations. Denote by $N_d = d_1 d_2$ the number of points in any of the grids D_l, with $l = 1, \ldots, L$. The covariance function of the processes are defined by:

$$C_l(\mathbf{u}) = E(Z_l(\mathbf{s}), Z_l(\mathbf{s} + \mathbf{u})), \quad \mathbf{s}, \mathbf{u} \in \mathbb{Z}^2. \tag{5.1}$$

Assuming that $\sum_{\mathbf{u}} |C_l(\mathbf{u})| < \infty$, by Khinchin's theorem (see Section 1.3.5), the covariance function of a stationary random process can be written, for $l = 1, \ldots, L$ as:

$$C_l(\mathbf{u}) = \int_{\Pi^2} e^{-i\mathbf{u}^T\boldsymbol{\lambda}} f_l(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \tag{5.2}$$

where f_l, the spectral density, is bounded and continuous for all $l = 1, \ldots, L$. The classical nonparametric estimator of the spatial spectral density, the periodogram, is given by:

$$I_l(\boldsymbol{\lambda}_\mathbf{k}) = \frac{1}{(2\pi)^2 N_d} \left| \sum_{\mathbf{s} \in D_l} Z_l(\mathbf{s}) e^{-i\mathbf{s}^T \boldsymbol{\lambda}_\mathbf{k}} \right|^2, \tag{5.3}$$

where $\mathbf{s}^T \boldsymbol{\lambda}_\mathbf{k}$ denotes the scalar product in $\mathbb{R}^2$. The periodogram is usually computed at the set of bidimensional Fourier frequencies, $\boldsymbol{\lambda}_\mathbf{k}^T = (\lambda_{k_1}, \lambda_{k_2})$:

$$\lambda_{k_i} = \frac{2\pi k_i}{d_i}, \quad k_i = 0, \pm 1, \ldots, \pm n_i = \lfloor \frac{d_i - 1}{2} \rfloor, \quad i = 1, 2 \tag{5.4}$$

and denote by $N = (2n_1 + 1)(2n_2 + 1)$ the number of Fourier frequencies. The periodogram (5.3) can be also written in terms of the sample covariances as:

$$I_l(\boldsymbol{\lambda}_\mathbf{k}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{u} \in \mathcal{U}} \hat{C}_l(\mathbf{u}) e^{-i\mathbf{u}^T \boldsymbol{\lambda}_\mathbf{k}}, \quad l = 1, \ldots, L \tag{5.5}$$

where $\mathcal{U} = \{\mathbf{u} = (u_1, u_2); u_i = 1 - d_i, \ldots, d_i - 1, i = 1, 2\}$ and the sample covariances, for Z_l with $l = 1, \ldots, L$, are given by:

$$\hat{C}_l(\mathbf{v}) = \frac{1}{N_d} \sum_{\mathbf{s} \in D_l(\mathbf{v})} Z_l(\mathbf{s}) Z_l(\mathbf{s} + \mathbf{v}), \quad D_l(\mathbf{v}) = \{\mathbf{s} \in D_l; \mathbf{s} + \mathbf{v} \in D_l\}. \tag{5.6}$$

We will suppose that the spatial process Z_l can be represented as:

$$Z_l(\mathbf{s}) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_{jk}^l \varepsilon_l(s_1 - j, s_2 - k), \tag{5.7}$$

where the error variables ε_l are independent and identically distributed as $N(0, \sigma_{\varepsilon_l}^2)$, for $l = 1, \ldots, L$. Then, the corresponding spectral density f_l can be written as:

$$f_l(\boldsymbol{\lambda}) = |A_l(\boldsymbol{\lambda})|^2 f_{\varepsilon_l}(\boldsymbol{\lambda}), \quad \boldsymbol{\lambda} \in \Pi^2 \tag{5.8}$$

where $f_{\varepsilon_l}(\boldsymbol{\lambda}) = \dfrac{\sigma_{\varepsilon_l}^2}{(2\pi)^2}$ and

$$A_l(\boldsymbol{\lambda}) = \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \psi_{jk}^l e^{-i(j,k)\boldsymbol{\lambda}}, \quad (j,k)\boldsymbol{\lambda} = j\lambda_1 + k\lambda_2.$$

In this case, the periodogram for each process Z_l, with $l = 1, \ldots, L$, admits the following representation:

$$I_l(\boldsymbol{\lambda}_\mathbf{k}) = f_l(\boldsymbol{\lambda}_\mathbf{k}) V_\mathbf{k}^l + R_N^l(\boldsymbol{\lambda}_\mathbf{k}), \tag{5.9}$$

where the variables $V_{\mathbf{k}}^l$ are i.i.d. standard exponential distributed, and $V_{\mathbf{k}}^l$ and $V_{\mathbf{k}}^{l'}$, with $l \neq l'$ are also independent. The residual term $R_N(\boldsymbol{\lambda}_{\mathbf{k}})$ is uniformly bounded (see Crujeiras *et al.* (2006b)). Applying logarithms in (5.9) we have:

$$Y_{\mathbf{k}}^l = m_l(\boldsymbol{\lambda}_{\mathbf{k}}) + z_{\mathbf{k}}^l + r_{\mathbf{k}}^l, \quad l = 1, \ldots, L \tag{5.10}$$

where $m_l = \log f_l$ is the log-spectral density, the variables $z_{\mathbf{k}}^l = \log V_{\mathbf{k}}^l$ are i.i.d. with density function $h(x) = e^{x-e^x}$, and the residual term $r_{\mathbf{k}}^l$ is given by:

$$r_{\mathbf{k}}^l = \log\left(1 + \frac{R_N^l(\boldsymbol{\lambda}_{\mathbf{k}})}{f_l(\boldsymbol{\lambda}_{\mathbf{k}})V_{\mathbf{k}}^l}\right).$$

Our main purpose is to test whether the spatial spectral densities are the same, or equivalently, in terms of the spatial log-spectral densities:

$$\begin{array}{ll} H_0: & m_1 = \ldots = m_L, \\ H_a: & m_l \neq m_j, \text{ for some } l \neq j. \end{array} \tag{5.11}$$

In this context, the comparison can be made by considering nonparametric estimators of the spatial log-spectral densities. Different nonparametric estimators of the spatial log-spectral density could be obtained considering a smoothed combination of log-periodogram values, that is:

$$\hat{m}_l(\boldsymbol{\lambda}_{\mathbf{k}}) = \sum_{\mathbf{i}} W_{\mathbf{i}}^l(\boldsymbol{\lambda}_{\mathbf{k}}) Y_{\mathbf{i}}^l. \tag{5.12}$$

The weights $W_{\mathbf{i}}^l$ can be defined as Gasser-Muller weights, for instance:

$$W_{\mathbf{i}}^l(\boldsymbol{\lambda}) = \frac{(2\pi)^2}{|H|^{1/2}} \int_{A_{\mathbf{i}}} K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\mu}))d\boldsymbol{\mu}, \tag{5.13}$$

where K is a bidimensional kernel function, H is a bidimensional bandwidth matrix and the integration region is such that:

$$A_{\mathbf{i}} = [a_{i_1-1}, a_{i_1}] \times [a_{i_2-1}, a_{i_2}], \quad \boldsymbol{\lambda}_{\mathbf{i}} \in A_{\mathbf{i}}, \quad \cup_i A_i = A, \quad A_i \cap A_j = \oslash, i \neq j.$$

The sets A_i in the partition of A must be Jordan measurable and $\max_i \mu(A_i) = \mathcal{O}(N^{-1})$ (see Müller (1988)). Another options are Priestley-Chao weights:

$$W_{\mathbf{i}}^l(\boldsymbol{\lambda}) = \frac{(2\pi)^2}{N} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{i}}) = \frac{(2\pi)^2}{N|H|^{1/2}} K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{i}})), \tag{5.14}$$

and Nadaraya-Watson weights:

$$W_{\mathbf{i}}^l(\boldsymbol{\lambda}) = \frac{(2\pi)^2 K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{i}})}{\sum_{\mathbf{i}} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{i}})} = \frac{(2\pi)^2 K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{i}}))}{\sum_{\mathbf{i}} K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{i}}))}. \tag{5.15}$$

Another alternative consists of considering a local-linear estimator for the spatial log-spectral density. A pilot local-linear estimator for $m_l(\boldsymbol{\lambda})$, with $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$, is obtained by multivariate local linear least squares regression by minimizing:

$$\sum_{\mathbf{k}} \left(Y_{\mathbf{k}}^l - (\beta_0, \beta_{10}, \beta_{01}) \begin{pmatrix} 1 \\ \lambda_1 - \lambda_{k1} \\ \lambda_2 - \lambda_{k2} \end{pmatrix} \right)^2 \cdot K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_{\mathbf{k}}), \tag{5.16}$$

and the nonparametric estimator is given by $\hat{m}^l(\boldsymbol{\lambda}) = \hat{\beta}_0$, where $(\hat{\beta}_0, \hat{\beta}_{10}, \hat{\beta}_{01})$ is the argument that minimizes expression (5.16).

The previous expressions for a nonparametric estimator of the log-spectral density come from the nonparametric regression context. Back to model (5.10), and taking into account the distribution of the error variables $z_{\mathbf{k}}^l$, the log-likelihood associated with this model, ignoring the term $r_{\mathbf{k}}^l$, is given by:

$$\sum_{\mathbf{k}} \left(Y_{\mathbf{k}}^l - m_l(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y_{\mathbf{k}}^l - m_l(\boldsymbol{\lambda}_{\mathbf{k}})} \right), \quad l = 1, \ldots, L. \tag{5.17}$$

From a nonparametric approach, consider the estimator obtained for the spatial log-spectral density m_l by a multidimensional local linear kernel estimator. For $\mathbf{x} \in \mathbb{R}^2$, we approximate $m_l(\boldsymbol{\lambda}_{\mathbf{k}})$ by the plane $a_l + \mathbf{b}_l^T(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x})$. Therefore, a local log-likelihood function based on (5.17) is given by:

$$\sum_{\mathbf{k}} \left(Y_{\mathbf{k}}^l - a_l - \mathbf{b}_l^T(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x}) - e^{Y_{\mathbf{k}}^l - a_l - \mathbf{b}_l^T(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x})} \right) K_H(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x}), \quad l = 1, \ldots, L \tag{5.18}$$

where K_H is the rescaled kernel, and take the maximum local log-likelihood estimator $\hat{m}_l(\mathbf{x})$ of $m_l(\mathbf{x})$ as $\hat{a}_l$ in the maximizer $(\hat{a}_l, \hat{\mathbf{b}}_l)$ of (5.18). These nonparametric estimators are used to illustrate the simulation study and the real data application.

In some cases, we may know some features of the dependence structures. For instance, by applying a goodness-of-fit test as those proposed in Crujeiras *et al.* (2006b), before testing the equality of the spatial spectral densities, we could assess whether the spectral densities belong to the same parametric family. In that case, the testing problem would be stated as:

$$\begin{array}{ll} H_0: & \theta_1 = \ldots = \theta_L, \\ H_a: & \theta_l \neq \theta_j, \quad \text{for some } l \neq j, \end{array} \tag{5.19}$$

with $m_{\theta_l} = \log f_{\theta_l}$, $l = 1, \ldots, L$ and $f_{\theta_l} \in \mathcal{F}_\Theta$, where $\mathcal{F}_\Theta$ denotes a parametric family of spectral densities.

5.2 An L_2 test for comparing spatial spectral densities.

As we have already commented in the introduction, Zhu *et al.* (2002) develop hypothesis testing methods to detect differences in a spatial random process, at two different time points. The testing techniques are based on the SCDF. This random function provides a spatial statistical summary of the random field and it is defined as:

$$F_{\infty,Z}(z;D) = \frac{1}{|D|}\int_D \mathbf{1}(Z(\mathbf{s}) \leq z)d\mathbf{s}, \tag{5.20}$$

where $\{Z(\mathbf{s}), \mathbf{s} \in D\}$, with $\mathbf{s}$ a continuous spatial index, $D \subset \mathbb{R}^d$, $|D|$ denotes the volume of D and $\mathbf{1}$ is the indicator function. For a finite sample of the process $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_N)\}$, the empirical counterpart of the SCDF in (5.20), namely the empirical spatial cumulative distribution function (ESCDF) is given by:

$$F_{N,Z}(z;D) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(Z(\mathbf{s}_i) \leq z). \tag{5.21}$$

In Lahiri (1999), asymptotic theoretical results on the SCDF are given, considering a shrinking asymptotic framework (Cressie (1993), pp. 100-101). With the motivation of detecting changes or trends in ecological resources over time, and for the particular case of two time points, the authors derive the large-sample distribution of a normalized test statistic based on the difference of the ESCDFs at two time moments. A second procedure quantifies the change using a weighted integrated squared difference between the SCDFs.

In the spatial context, we do not know more references on this topic of spatial processes comparison. In our case, we are interested in detecting changes on the dependence structure and, for that purpose, we will consider a spectral approach.

Consider the following test statistic, based on a L^2-distance:

$$Q = \sum_{l=2}^{L}\left(\sum_{j=1}^{l-1}\left(\int_{\Pi^2}\left(\widehat{m_l}(\boldsymbol{\lambda}) - \widehat{m_j}(\boldsymbol{\lambda})\right)^2 \omega(\boldsymbol{\lambda})d\boldsymbol{\lambda}\right)\right), \tag{5.22}$$

where ω is a weighting function in Π^2. This weighting function ω is chosen in such a way the edge-effect on the estimation is avoided. In our context, we consider a weighting function that filters frequencies around the origin and those ones with components near to $\frac{2\pi n_i}{d_i}$. Besides, in these cases, the log-periodogram values may present a higher variability. This edge-effect error is also corrected by the local linear estimator.

For simplicity, consider the testing problem $H_0 : m_1 = m_2$ vs. $H_a : m_1 \neq m_2$. In the general case of L processes, we proceed similarly. Assume that both Z_1 and Z_2 have been observed on grids with the same design. This implies that the corresponding Fourier frequencies are the same in both cases. Using Riemann approximation, Q can be approximated by $\hat{Q}$, which is given by:

$$\hat{Q} = \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} \left(\widehat{m_1}(\boldsymbol{\lambda}_\mathbf{k}) - \widehat{m_2}(\boldsymbol{\lambda}_\mathbf{k})\right)^2 \omega(\boldsymbol{\lambda}_\mathbf{k}). \tag{5.23}$$

In order to perform the test in practice, we will compute the distribution of the test statistic under the null hypothesis H_0. The asymptotic behaviour of $\hat{Q}$, under H_0, can be established but usually, the convergence of this type of test statistic to its limit distribution is slow (see, for instance, some works in the regression setting as Härdle and Mammen (1993)). Therefore, this distribution must be approximated by simulation methods. For that purpose, it is necessary to obtain an estimation of the spatial log-spectral density under $H_0 : m_1 = m_2 = m$.

In the case we consider, the design points (that is, the grid of Fourier frequencies) are the same, and we could build a pilot estimator of the log-spectral density under H_0, namely $\hat{m}$, as an average of the log-periodograms:

$$\tilde{Y}_\mathbf{k} = \frac{\left(Y_\mathbf{k}^1 + Y_\mathbf{k}^2\right)}{2}. \tag{5.24}$$

If our null hypothesis can be relaxed by the fact that $f_{\theta_1}, f_{\theta_2} \in \mathcal{F}_\Theta$, as in problem (5.19), then, an estimation of the log-spectral density under $H_0 : m_{\theta_1} = m_{\theta_2} = m_\theta$ can be given by a parametric estimator $m_{\hat{\theta}}$. The parameter vector θ can be estimated, under H_0, by a Whittle log-likelihood approach:

$$\hat{\theta} = \arg\max_\theta \sum_{\mathbf{k}} \left(\tilde{Y}_\mathbf{k} - m_\theta(\boldsymbol{\lambda}_\mathbf{k}) - e^{\tilde{Y}_\mathbf{k} - m_\theta(\boldsymbol{\lambda}_\mathbf{k})}\right). \tag{5.25}$$

As we have pointed, Whittle estimators are not consistent for dimensions higher or equal to two, and in the case of dimension two, these estimators show a non negligible bias. Different alternatives to achieve consistency in this estimation procedure can be found in Guyon (1982), Dahlhaus and Künsch (1987) or Crujeiras *et al.* (2006b).

In order to compute the test statistic (5.23), we must fix a bandwidth matrix H. The selection of the bandwidth matrix parameter is a crucial step in nonparametric estimation and testing. Nevertheless, the choice of optimal bandwidth matrices in multidimensional testing problems remains unsolved and usually, the standard approach consists of examining the behaviour of the test over a range of bandwidths.

Instead of trying a range of bandwidths, an automatic bandwidth selection criteria could be also used. For instance, we could take an optimal bandwidth matrix for the estimation problem, under H_0.

Since log-periodogram values are asymptotically independent, for a large enough sample, we may expect good approximations of $\hat{H}$ by using a cross-validation criteria. For the testing problem (5.19), the bandwidth matrix $\hat{H}$ may be selected such that:

$$\hat{H} = \arg\min_{H} \sum_{\mathbf{k}} \left(\hat{m}^{-\mathbf{k}}(H, \boldsymbol{\lambda}_{\mathbf{k}}) - m_{\hat{\theta}}(H, \boldsymbol{\lambda}_{\mathbf{k}}) \right)^2, \tag{5.26}$$

where $\hat{m}^{-\mathbf{k}}(H, \cdot)$ is the nonparametric estimator of the spatial log-spectral density obtained when ignoring the frequency $\boldsymbol{\lambda}_{\mathbf{k}}$ for obtaining the nonparametric estimator of m at this frequency.

In the nonparametric testing problem (5.11), the bandwidth matrix could be obtained from:

$$\hat{H} = \arg\min_{H} \sum_{\mathbf{k}} \left(\hat{m}^{-\mathbf{k}}(H, \boldsymbol{\lambda}_{\mathbf{k}}) - \tilde{Y}_{\mathbf{k}} \right)^2. \tag{5.27}$$

5.2.1 Asymptotic theory.

Consider $\{Z_l(\mathbf{s}), \mathbf{s} \in D_l\}$, with $l = 1, \ldots, L$, L realizations of a spatial stochastic process (for instance, realizations taken on L time moments) or L realizations of different spatial processes. Our main purpose is to solve the testing problem (5.11).

In this context, the comparisson can be made by considering nonparametric estimators of the spatial log-spectral densities. Consider the test statistic (5.22), based on an L^2-distance.

A1 The spatial processes can be represented as:

$$Z^l(\mathbf{s}) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_{ij}^l \varepsilon_l(s_1 - i, s_2 - j), \quad l = 1, \ldots, L,$$

where ε_l are i.i.d. $N(0, \sigma_{\varepsilon_l}^2)$ (and independent among themselves) random variables and $\sum |\psi_{ij}^l|^2 < \infty$.

A2 The spectral densities are non-vanishing:

$$\inf_{\boldsymbol{\lambda} \in \Pi^2} f_l(\boldsymbol{\lambda}) > 0, \quad l = 1, \ldots, L.$$

A3 We consider Gasser-Muller type weights, given by (5.13), or Priestley-Chao weights, given by (5.14). Besides, $W_{\mathbf{i}}^1 = \ldots = W_{\mathbf{i}}^L$.

A4 The bidimensional kernel function K is continuously differentiable, with compact support and $\int K^2(\mathbf{u})d\mathbf{u} < \infty$.

A5 The bidimensional bandwidth matrix, H satisfies $N|H|^{1/2} \to \infty$, as $N \to \infty$, with $n_1, n_2 \to \infty$ and $n_1/n_2 \to c$, for some constant c.

Consider first the testing problem $H_0 : m_1 = m_2$ vs. $H_a : m_1 \neq m_2$ and assume that both Z^1 and Z^2 have been observed on grids with the same design. Consider the Riemann approximation to the (5.22), namely $\hat{Q}$, given by (5.23).

Theorem 5. *Assume conditions (A1)-(A5) hold. Then, under the null hypothesis that $H_0 : m_1 = m_2$, we have that, as $N \to \infty$:*

$$\sqrt{N|H|^{1/2}}\left(\hat{Q} - \frac{(2\pi)^4}{12N|H|^{1/2}}C_K I_\omega\right) \to N(0, \sigma_Q^2), \tag{5.28}$$

in distribution, with

$$C_K = \int K^2(\mathbf{u})d\mathbf{u}, \quad I_\omega = \int \omega(\mathbf{v})d\mathbf{v} \quad \text{and the asymptotic variance is}$$

$$\sigma_Q^2 = \frac{(2\pi)^8}{72}\int (K * K)^2(\mathbf{u})d\mathbf{u}\int \omega^2(\mathbf{v})d\mathbf{v},$$

where $*$ denotes the convolution operator.

Also in this context of two dependence structures comparison, consider that the null hypothesis is false and assume:

$$m_1(\boldsymbol{\lambda}) - m_2(\boldsymbol{\lambda}) = C_N p(\boldsymbol{\lambda}), \tag{5.29}$$

where $p(\boldsymbol{\lambda})$ is a non-zero function. We will see that the test statistic $\hat{Q}$ allows for detecting local alternatives at a distance of order $N^{-1/2}|H|^{-1/8}$.

Theorem 6. *Assume conditions (A1)-(A5) hold. Then, if (5.29) holds, we have that, as $N \to \infty$:*

$$\sqrt{N|H|^{1/2}}\left(\hat{Q} - \frac{(2\pi)^4}{12N|H|^{1/2}}C_K I_\omega\right) \to N\left(\int p^2(\mathbf{v})\omega(\mathbf{v})d\mathbf{v}, \sigma_Q^2\right),$$

in distribution, with C_K, I_ω and σ_Q^2 as in Theorem 1, and with $C_N^2 = (N^2|H|^{1/2})^{-1/2}$.

It is easy to generalize Theorems 5 and 6 for random fields on $\mathbb{R}^d$. Considering a d-variate kernel function K satisfying condition A4 and a d-dimensional bandwidth matrix H, satisfying condition $A5$. The corresponding asymptotic mean and variance in (5.28) are given by:

$$\frac{4}{3}\pi^{2d}\frac{1}{N|H|^{1/2}}C_K I_\omega, \quad \text{and}$$

$$\sigma^2_{Q,d} = \frac{2^{2d+1}}{9}\pi^{4+2d}\int (K*K)^2(\mathbf{u})d\mathbf{u}\int \omega^2(\mathbf{v})d\mathbf{v},$$

where the weighting function ω is now defined on $\Pi^d = [-\pi,\pi]^d$. Thus, in the particular case of $d = 1$, we provide a testing technique for comparing temporal spectral densities. In this case, we have a scalar bandwidth parameter h, which plays the role of $H^{1/2}$ in the general dimension setting.

If the spatial process Z_l, for $l = 1,\ldots,L$, are observed on regular grids with different sizes, then the corresponding frequency spectrum is not the same. The asymptotic behaviour of $\hat{Q}$ could be determined following similar arguments to those in (Vilar-Fernández and González-Manteiga (2004), Theorem 3), under some conditions on the asymptotic rates of the samples.

5.2.2 Bootstrap procedures for calibrating p-values.

Consider the testing problem (5.19). This *a priori* information will simplify the algorithm for callibrating the p-value of the test in practice. An estimation of the test statistic, under H_0, can be given by a Monte Carlo approach. In order to calibrate the p-value of the test statistic $\hat{Q}$, the following algorithm can be employed in practice.

- **Algorithm 1.**

Step 1. Compute the observed test statistic $\hat{Q}^{obs}$.

Step 2. Draw two random samples of size $d_1 \times d_2$, with the log-spectral density under H_0, that is, $m_{\hat{\theta}}$.

Step 3. Compute the test statistic for these generated random samples $\hat{Q}^{(b)}$.

Step 4. Repeat *Step 1* and *Step 2* B times and obtain the tests statistic $\hat{Q}^{(1)},\ldots,\hat{Q}^{(B)}$.

Step 5. Compute the p-value of the test statistic as the percentage of bootstrap replicates $\{\hat{Q}^{(1)},\ldots,\hat{Q}^{(B)}\}$ that exceed the observed value $Q^{(obs)}$.

In this algorithm, a parametric estimation of the spatial log-spectral density is needed in *Step 2.* This parametric estimator is usually obtained by a Whittle-log likelihood approach, based on an average of the periodograms in the different observations. Whittle estimators are obtained by maximizing (5.17).

Also in *Step 2*, in order to generate a random sample from a spatial process with a certain spatial spectral density, one could use a specific algorithm, for instance, when we consider linear-by-linear process (see Alonso *et al.* (1996)). When an specific algorithm is not available, then we must use a standard technique for the simulation of spatial processes. In this case, Cholesky factorization based method (see Cressie (1993), for example) could be used. Another alternative is spectral simulation procedures, as the Modified Fourier Integral Method (Crujeiras and Fernández-Casal (2006)), which has been introduced in Chapter 3.

In the nonparametric testing problem, when we ignore whether the spectral densities belong to the same parametric familiy, a totally nonparametric algorithm for approximating the p-value of the test must be considered.

- **Algorithm 2.**

Step 1. Compute the observed test statistic $\hat{Q}^{obs}$.

Step 2. Draw two random samples of size N_d, with the log-spectral density under $H_0 : m_1 = m_2$ as follows:

2.a Obtain a non parametric estimation $\hat{m}$ of the log-spectral density. For example compute the log-periodograms, taking the average at each frequency (as in (5.24)) and smooth this average to get $\tilde{Y}^s_{\mathbf{k}}$ (see Robinson (2006)).

2.b Apply the Inverse Fourier Transform on $\tilde{I}^s(\boldsymbol{\lambda}_{\mathbf{k}}) = e^{\tilde{Y}^s_{\mathbf{k}}}$ and get an estimation of the covariance function $\hat{C}(\mathbf{u})$, with $\mathbf{u} \in \mathcal{U}$.

2.c Obtain two realizations of the process, on a grid of size N_d, from the estimated covariances, $\hat{C}$.

Step 3. Compute the test statistic for these generated random samples $\hat{Q}^{(b)}$.

Step 4. Repeat *Step 1* and *Step 2* B times and obtain the tests statistic $\hat{Q}^{(1)}, \ldots, \hat{Q}^{(B)}$.

Step 5. Compute the p-value of the test statistic as the percentage of bootstrap replicates $\{\hat{Q}^{(1)}, \ldots, \hat{Q}^{(B)}\}$ that exceed the observed value $\hat{Q}^{(obs)}$.

In *Step 2.a* we must take into account that, in order to generate a sample on a grid $\{1, \ldots, d_1\} \times \{1, \ldots, d_2\}$ the covariances $\hat{C}(\mathbf{u})$ must cover a wider grid of size $\{1, \ldots, k_1\} \times \{1, \ldots, k_2\}$, with $k_i = 2d_i - 1$, with $i = 1, 2$ (see Priestley (1981)). Therefore, the frequency spectrum must be also recalculated. In *Step 2.c*, Cholesky factorization method could be applied.

These algorithms can be easily generalized to the general case of checking for differences within a collection of $L > 2$ processes, or $L > 2$ observations of the same process.

5.3 Simulation results.

We illustrate the performance of the test statistic considering bidimensional autoregressive process ($BAR(1)$):

$$Z_l(i,j) = \beta_1^l Z(i-1,j) + \beta_2^l Z(i,j-1) - \beta_1^l \beta_2^l Z(i-1,j-1) + \varepsilon_l(i,j), \quad l = 1,2, \tag{5.30}$$

where $\varepsilon_l(i,j)$ are independent identically distributed Gaussian random variables, with zero mean and variance $\sigma^2_{\varepsilon_l}$. The spectral densities corresponding to Z_1 and Z_2 are given by:

$$f_l(\boldsymbol{\lambda}) = \frac{\sigma^2_\varepsilon}{(2\pi)^2} \cdot \frac{1}{1 + (\beta_1^l)^2 - 2\beta_1^l \cos(\lambda_1)} \cdot \frac{1}{1 + (\beta_2^l)^2 - 2\beta_2^l \cos(\lambda_2)}, \quad l = 1,2. \tag{5.31}$$

In order to study the performance of the test, in terms of size and power, we consider different values for β_1^l and β_2^l, from 0.0 (corresponding to the case of independence) to 0.9. One thousand replicates of the process are generated on a 20×20 regular grid. Random sample generations of this process are obtained as in Alonso *et al.* (1996).

We set the null hypothesis that Z_1 and Z_2 are $BAR(1)$ processes with the same dependence structure, that is, testing problem (5.19). Therefore, *Algorithm 1* is implemented in this case. A multiplicative Epanechnikov bidimensional kernel is considered. The weighting function ω filters the frequencies near the origin and those with the largest components, in order to avoid the edge effect. The bandwidth parameter is chosen using the cross-validation criteria (5.26). We consider diagonal bandwidth matrices, whose elements are proportional to the spacing between frequencies, that is:

$$H = r \cdot diag\left(\frac{2\pi}{n_1}, \frac{2\pi}{n_2}\right). \tag{5.32}$$

The nonparametric estimator for the spatial log-spectral density is obtained from the local-linear method, specified in equation (5.16).

The size of the test is shown in Table 5.1, at three different significance levels α: 0.01, 0.05 and 0.10. $\boldsymbol{\beta}^j = (\beta_1^j, \beta_2^j)$, for $j = 1,2$, denote the parameters in Z_1 and Z_2, respectively. The percentage of rejections of the test statistic is computed from 1000 simulations. Some results on the power of the test are shown in Table 5.2. The test shows a good behaviour in all cases.

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.0, 0.0)$	0.010	0.050	0.093
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.3, 0.3)$	0.007	0.033	0.076
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.010	0.041	0.081
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.048	0.107	0.193

Table 5.1: Size of the test, with **Algorithm 1**. 20×20 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (5.30), for the sample from Z_j, $j = 1, 2$. Significance level α.

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.05, 0.05)$	0.015	0.054	0.112
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.1, 0.1)$	0.017	0.065	0.133
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.2, 0.2)$	0.067	0.196	0.293
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.3, 0.3)$	0.290	0.550	0.670
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.990	0.990	1.000
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.9, 0.9)$	1.000	1.000	1.000
$\boldsymbol{\beta}^1 = (0.3, 0.3), \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.220	0.500	0.590
$\boldsymbol{\beta}^1 = (0.3, 0.3), \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.910	0.980	0.990
$\boldsymbol{\beta}^1 = (0.6, 0.6), \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.200	0.380	0.520

Table 5.2: Power of the test, with **Algorithm 1**. 20×20 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (5.30), for the sample from Z_j, $j = 1, 2$. Significance level α.

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.0, 0.0)$	0.014	0.049	0.095
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.3, 0.3)$	0.011	0.046	0.102
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.015	0.056	0.099
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.057	0.131	0.210

Table 5.3: Size of the test, with **Algorithm 2**. 20×20 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (5.30), for the sample from Z_j, $j = 1, 2$. Significance level α.

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.05, 0.05)$	0.015	0.054	0.109
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.1, 0.1)$	0.015	0.069	0.130
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.2, 0.2)$	0.074	0.207	0.317
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.3, 0.3)$	0.326	0.579	0.700
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.994	0.999	1.000
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.9, 0.9)$	1.000	1.000	1.000
$\boldsymbol{\beta}^1 = (0.3, 0.3), \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.252	0.489	0.638
$\boldsymbol{\beta}^1 = (0.3, 0.3), \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.927	0.983	0.995
$\boldsymbol{\beta}^1 = (0.6, 0.6), \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.265	0.469	0.590

Table 5.4: Power of the test, with **Algorithm 2**. 20×20 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (5.30), for the sample from Z_j, $j = 1, 2$. Significance level α.

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.05, 0.05)$	0.034	0.112	0.192
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.1, 0.1)$	0.202	0.390	0.518
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.2, 0.2)$	0.910	0.974	0.988

Table 5.5: Power of the test, with **Algorithm 2**. 40×40 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (5.30), for the sample from Z_j, $j = 1, 2$. Significance level α.

When no a priori knwoledge on the form of the spectral densities is available, then *Algorithm 2* must be implemented. Under the same conditions on the size of the grid, kernel function, bandwidth selection and number of Bootstrap replicates, we run new simulations. In this case, Z_1 and Z_2 are simulated from model (5.30) but we do not use the fact that both spectral densities belong to the same family.

Results of the test, using the completely nonparametric algorithm, are given in Tables 5.3 and 5.4. Comparing these results with the ones shown in Tables 5.1 and 5.2, we can see that there is not a great loss in terms of size and power, when the fact that the two log-spectral densities belong to the same parametric family.

In Table 5.5, we show the power results for the nonparametric algorithm, in a 40×40 regular grid. We can see that its performance is better than for the 20×20 grid case.

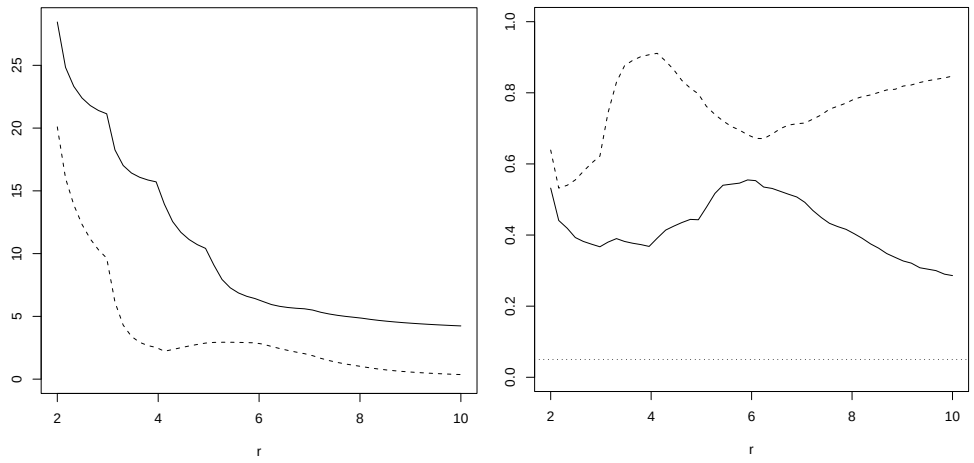

Figure 5.1: Left panel: test statistics. Right panel: p-values. Solid line: based on the local linear estimator (5.16). Dashed line: based on local loglikelihood estimator (5.18). Dotted line: significance level 0.05. r denotes the scaling parameter in (5.32).

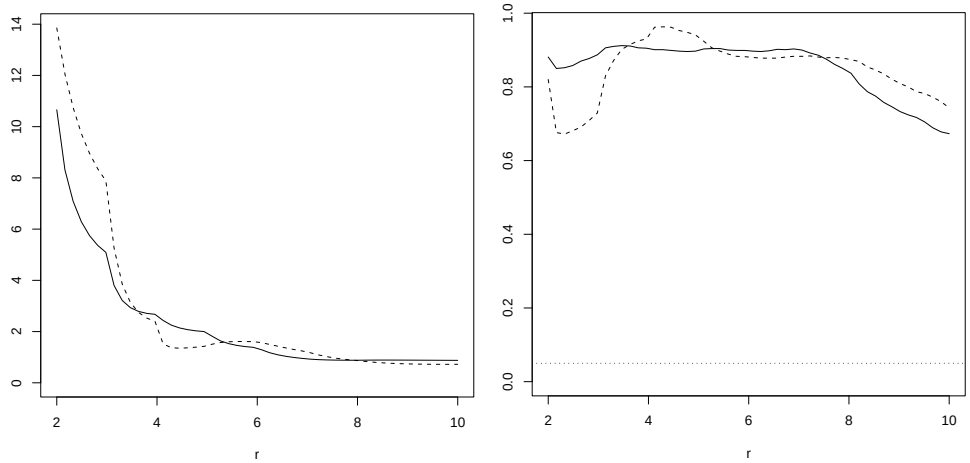

Figure 5.2: Left panel: test statistics for $\log(Hg)$ concentrations. Right panel: p-values. Solid line: based on the local linear estimator (5.16). Dashed line: based on local loglikelihood estimator (5.18). Dotted line: significance level 0.05. r denotes the scaling parameter in (5.32).

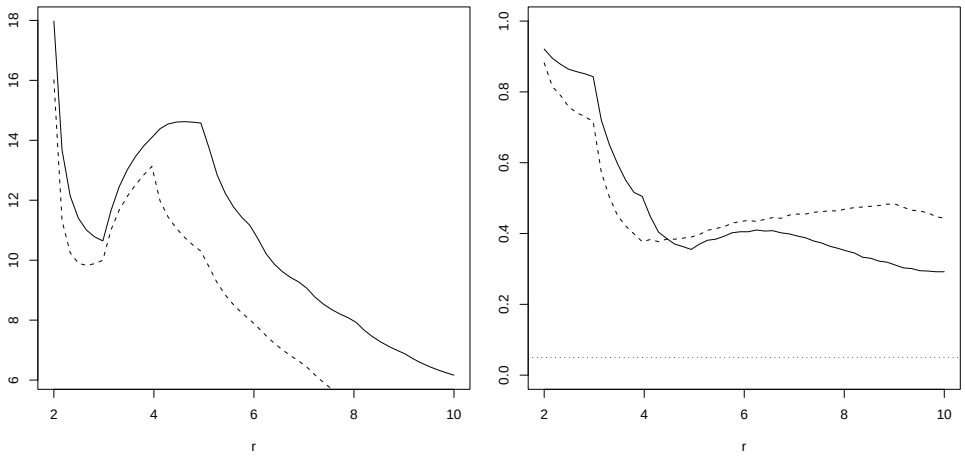

Figure 5.3: Left panel: test statistics for $\log(Se)$ concentrations. Right panel: p-values. Solid line: based on the local linear estimator (5.16). Dashed line: based on local loglikelihood estimator (5.18). Dotted line: significance level 0.05. r denotes the scaling parameter in (5.32).

5.4 Real data application.

We apply the testing technique to Mercury and Selenium concentration datasets. Mercury and Selenium concentrations were measured (over a the same regular grid) on 2000, 2002 and 2004 (March and September, in this last year). Our main goal is to check whether the dependence structure in the process observed in March 2004 and September 2004 is the same. The importance of controlling Hg and Se concentrations has beeen pointed out in Section 1.4.2.

In this case, we consider two nonparametric estimators for the spatial log-spectral density. First, we consider a local linear estimator, given by (5.16). Secondly, the Whittle estimator from (5.18) is used. The kernel function is a multiplicative Epanechnikov kernel and the weighting function is chosen to avoid the edge-effect. The algorithm for approximating the p-value of the test statistic is the nonparametric Algorithm 2. In Figure 5.1 we show the values of the tests (right panel) and the corresponding p-values (left panel) along a range of bandwidths. In Figure 5.2, we show the results of the tests and the corresponding p-values when applying a logarithmic transform to the data. There is no evidence that the dependence structure in Hg concentrations has changed from March to September.

Histogram for $\log(Se)$ concentrations has already been given in Section 1.4.2. In Figure 5.3 we show the tests statistics and the corresponding p-values from Algorithm 2, considering local linear (5.16) and local Whittle (5.18) estimation. Once again, there is no significative change in the dependence structure of selenium concentrations.

5.5 Appendix Chapter 5.

Let's introduce the following notation. Consider the following regression model:

$$Y_{\mathbf{k}}^{l*} = m_l(\boldsymbol{\lambda}_{\mathbf{k}}) + z_{\mathbf{k}}^{l*}, \quad l = 1, 2. \tag{5.33}$$

where $Y_{\mathbf{k}}^{l*} = Y_{\mathbf{k}}^{l} - C_0 - r_{\mathbf{k}}^{l}$, and $z_{\mathbf{k}}^{l*} = z_{\mathbf{k}}^{l} - C_0$, where $C_0 = E(z_{\mathbf{k}}^{l})$ is the Euler constant. Denote by $\widehat{m_l}$ the nonparametric estimator of m_l in (5.33)

$$\widehat{m_l}^*(\boldsymbol{\lambda}_{\mathbf{k}}) = \sum_{\mathbf{j}} W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) Y_{\mathbf{j}}^{l*}.$$

and denote by $B_{\mathbf{k}}^{l}$:

$$B_{\mathbf{k}}^{l} = \sum_{\mathbf{j}} W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) r_{\mathbf{j}}^{l}, \quad l = 1, 2.$$

5.5.1 Proof of Theorem 5

Lemma 20. *The test statistic $\hat{Q}$ can be decomposed in three addends:*

$$\hat{Q} = \hat{Q}_1 + \hat{Q}_2 + \hat{Q}_3,$$

where

$$\hat{Q}_1 = \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} (\widehat{m_1}^*(\boldsymbol{\lambda}_{\mathbf{k}}) - \widehat{m_2}^*(\boldsymbol{\lambda}_{\mathbf{k}}))^2 \omega(\boldsymbol{\lambda}_{\mathbf{k}}),$$

$$\hat{Q}_2 = \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} (B_{\mathbf{k}}^1 - B_{\mathbf{k}}^2)^2 \omega(\boldsymbol{\lambda}_{\mathbf{k}}),$$

$$\hat{Q}_3 = 2\frac{(2\pi)^2}{N} \sum_{\mathbf{k}} (\widehat{m_1}^*(\boldsymbol{\lambda}_{\mathbf{k}}) - \widehat{m_2}^*(\boldsymbol{\lambda}_{\mathbf{k}}))(B_{\mathbf{k}}^1 - B_{\mathbf{k}}^2) \omega(\boldsymbol{\lambda}_{\mathbf{k}}).$$

Proof. It is straightforward from the definitions of the non parametric estimator in regression model (5.33) and the quantities $B_{\mathbf{k}}^1$, $B_{\mathbf{k}}^2$.

$$\begin{aligned}
\hat{Q} &= \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} (\widehat{m_1}(\boldsymbol{\lambda}_{\mathbf{k}}) - \widehat{m_2}(\boldsymbol{\lambda}_{\mathbf{k}}))^2 \omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&= \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} (\widehat{m_1}^*(\boldsymbol{\lambda}_{\mathbf{k}}) - \widehat{m_2}^*(\boldsymbol{\lambda}_{\mathbf{k}}) + B_{\mathbf{k}}^1 - B_{\mathbf{k}}^2)^2 \omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&= \hat{Q}_1 + \hat{Q}_2 + \hat{Q}_3.
\end{aligned}$$

□

Lemma 21. *Under conditions (A1)-(A5) and under H_0, we have that:*

$$\hat{Q}_2 = \mathcal{O}_{\mathbb{P}}\left(\frac{\log^2 N}{N^2|H|^{1/2}}\right),$$

$$\hat{Q}_3 = \mathcal{O}_{\mathbb{P}}\left(\frac{\log^2 N}{N}\right).$$

Proof. Recall that, by definitions of $B^1_{\mathbf{k}}$ and $B^2_{\mathbf{k}}$, the statistic $\hat{Q}_2$ can be written as:

$$\hat{Q}_2 = \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}(B^1_{\mathbf{k}} - B^2_{\mathbf{k}})^2\omega(\boldsymbol{\lambda}_{\mathbf{k}}),$$

where

$$B^1_{\mathbf{k}} = \sum_{\mathbf{j}} W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})r^1_{\mathbf{j}}, \quad B^2_{\mathbf{k}} = \sum_{\mathbf{j}} W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})r^2_{\mathbf{j}}.$$

Then, $\hat{Q}_2$ can be decomposed as:

$$\begin{aligned}\hat{Q}_2 &= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})(r^1_{\mathbf{j}} - r^2_{\mathbf{j}})^2\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\ &+ \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{i}}\sum_{\mathbf{j}\neq\mathbf{i}} W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}})(r^1_{\mathbf{j}} - r^2_{\mathbf{j}})(r^1_{\mathbf{i}} - r^2_{\mathbf{i}})\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\ &= \hat{Q}_{2,1} + \hat{Q}_{2,2}.\end{aligned}$$

Recall the expression for the residual part $r^l_{\mathbf{j}}$, for $l = 1, 2$:

$$r^l_{\mathbf{j}} = \log\left(1 + \frac{R^l_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f_l(\boldsymbol{\lambda}_{\mathbf{j}})V_{\mathbf{j}}}\right), \quad l = 1, 2$$

and apply a Taylor expansion around 0:

$$r^l_{\mathbf{j}} = -\frac{R^l_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f_l(\boldsymbol{\lambda}_{\mathbf{j}})V_{\mathbf{j}}} - \frac{1}{2(1+x_j)^2}\left(\frac{R^l_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f_l(\boldsymbol{\lambda}_{\mathbf{j}})V_{\mathbf{j}}}\right)^2,$$

where $x_j \in \left(0, \frac{R^l_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f_l(\boldsymbol{\lambda}_{\mathbf{j}})V_{\mathbf{j}}}\right)$. Since,

$$\max_{\mathbf{j}} |R^l_N(\boldsymbol{\lambda}_{\mathbf{j}})| = \mathcal{O}_{\mathbb{P}}(N^{-1/2}\log N),$$

for $l = 1, 2$, just following Kooperberg *et al.* (1995), the Lagrange remainder in the Taylor expansion can be uniformly bounded by:

$$\max_{\mathbf{j}}\left|\frac{1}{2(1+x_j)^2}\left(\frac{R^l_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f_l(\boldsymbol{\lambda}_{\mathbf{j}})V_{\mathbf{j}}}\right)^2\right| = \mathcal{O}_{\mathbb{P}}\left(\frac{\log^2 N}{N}\right).$$

Denote the Lagrange remainder from the Taylor expansion by $LR^l_{\mathbf{j}}$, for $l = 1, 2$. Then:

$$\hat{Q}_{2,1} = \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} \sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) \left(\frac{R^2_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f_2(\boldsymbol{\lambda}_{\mathbf{j}})V_{\mathbf{j}}} - \frac{R^1_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f_1(\boldsymbol{\lambda}_{\mathbf{j}})V_{\mathbf{j}}} + LR^2_{\mathbf{j}} - LR^1_{\mathbf{j}} \right)^2 \omega(\boldsymbol{\lambda}_{\mathbf{k}}).$$

Since we assume that $H_0 : f_1 = f_2$ holds, we will omit the subindex in the spectral densities. Besides, since the $V_{\mathbf{j}}$ variables are independent (both between and within populations 1 and 2), we will denote this variables by V. Computing the square, $\hat{Q}_{2,1}$ can be decomposed again in three addends as:

$$\hat{Q}_{2,1} = \hat{Q}^1_{2,1} + \hat{Q}^2_{2,1} + \hat{Q}^3_{2,1},$$

where

$$\hat{Q}^1_{2,1} = \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} \sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) \left(\frac{R^2_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})V} - \frac{R^1_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})V} \right)^2 \omega(\boldsymbol{\lambda}_{\mathbf{k}}),$$

$$\hat{Q}^2_{2,1} = \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} \sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) (LR^2_{\mathbf{j}} - LR^1_{\mathbf{j}})^2 \omega(\boldsymbol{\lambda}_{\mathbf{k}}),$$

and

$$\hat{Q}^3_{2,1} = 2\frac{(2\pi)^2}{N} \sum_{\mathbf{k}} \sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) \left(\frac{R^2_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})V} - \frac{R^1_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})V} \right) (LR^2_{\mathbf{j}} - LR^1_{\mathbf{j}}) \omega(\boldsymbol{\lambda}_{\mathbf{k}}).$$

Let's find a bound for $\hat{Q}^2_{2,1}$, the addend involving the Lagrange remainders.

$$\begin{aligned}
\hat{Q}^2_{2,1} &= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})(LR^1_{\mathbf{j}} - LR^2_{\mathbf{j}})^2\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&\leq \max_{\mathbf{j}} |LR^1_{\mathbf{j}} - LR^2_{\mathbf{j}}|^2 \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&= \mathcal{O}_{\mathbb{P}}\left(\frac{\log^4 N}{N^2}\right)\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}}\frac{(2\pi)^4}{N^2|H|}K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}}))\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&= \mathcal{O}_{\mathbb{P}}\left(\frac{\log^4 N}{N^2}\right)\frac{(2\pi)^6}{N}\left(\sum_{\mathbf{k}}\sum_{\mathbf{j}\neq\mathbf{k}}\frac{1}{N^2|H|}K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}}))\omega(\boldsymbol{\lambda}_{\mathbf{k}})\right. \\
&\qquad\qquad \left. + \sum_{\mathbf{k}}\frac{1}{N|H|}K^2(H^{-1/2}\mathbf{0})\omega(\boldsymbol{\lambda}_{\mathbf{k}})\right) \\
&= \mathcal{O}_{\mathbb{P}}\left(\frac{\log^4 N}{N^4|H|^{1/2}}\right) \\
&\qquad + \mathcal{O}_{\mathbb{P}}\left(\frac{\log^4 N}{N^2}\right)\frac{(2\pi)^6}{N^3|H|}\sum_{\mathbf{k}}\sum_{\mathbf{j}\neq\mathbf{k}}K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}}))\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&\leq \mathcal{O}_{\mathbb{P}}\left(\frac{\log^4 N}{N^4|H|^{1/2}}\right) \\
&\qquad + \mathcal{O}_{\mathbb{P}}\left(\frac{\log^4 N}{N^2}\right)\frac{(2\pi)^6}{N^2|H|}\max_{\mathbf{k}}\sum_{\mathbf{j}\neq\mathbf{k}}K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}})) \\
&\approx \mathcal{O}_{\mathbb{P}}\left(\frac{\log^4 N}{N^4|H|^{1/2}}\right) + \mathcal{O}_{\mathbb{P}}\left(\frac{\log^4 N}{N^2}\right)\frac{1}{N|H|^{1/2}}C_K \\
&= \mathcal{O}_{\mathbb{P}}\left(\frac{\log^4 N}{N^3|H|^{1/2}}\right),
\end{aligned}$$

where the last inequality follows from $\max_{\mathbf{k}}\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \leq c$, for some constant c. Following similar arguments as above, we will found bounds for $\hat{Q}_{2,1}$ and $\hat{Q}_{2,3}$. Let's start with $\hat{Q}^1_{2,1}$:

$$\begin{aligned}
\hat{Q}^1_{2,1} &= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})\left(\frac{R^2_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})V} - \frac{R^1_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})V}\right)^2\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&= \hat{Q}^{1,1}_{2,1} + \hat{Q}^{1,2}_{2,1} + \hat{Q}^{1,3}_{2,1}
\end{aligned}$$

where

$$\hat{Q}^{1,1}_{2,1} = \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})\left(\frac{R^1_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})V}\right)^2\omega(\boldsymbol{\lambda}_{\mathbf{k}}),$$

$$\hat{Q}^{1,3}_{2,1} = 2\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}} W^2_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})\frac{R^1_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})V}\frac{R^2_N(\boldsymbol{\lambda}_{\mathbf{j}})}{f(\boldsymbol{\lambda}_{\mathbf{j}})V}\omega(\boldsymbol{\lambda}_{\mathbf{k}})$$

and $\hat{Q}_{2,1}^{1,2}$ is the similar to $\hat{Q}_{2,1}^{1,1}$, but replacing each $R_N^1(\boldsymbol{\lambda}_\mathbf{j})$ for $R_N^2(\boldsymbol{\lambda}_\mathbf{j})$. We will find a bound for $\hat{Q}_{2,1}^{1,1}$. Similar computations lead to the same bound for the other addends.

$$\begin{aligned}
\hat{Q}_{2,1}^{1,1} &\leq \frac{(2\pi)^2}{N} N \max_{\mathbf{j}} \left(\frac{R_N^1(\boldsymbol{\lambda}_\mathbf{j})}{f(\boldsymbol{\lambda}_\mathbf{j})V} \right)^2 \sum_{\mathbf{j}} \sum_{\mathbf{k}} W_\mathbf{j}^2(\boldsymbol{\lambda}_\mathbf{k}) \omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \mathcal{O}_\mathbb{P}\left(\frac{\log^2 N}{N} \right) \sum_{\mathbf{j}} \sum_{\mathbf{k}} \frac{(2\pi)^4}{N^2|H|} K^2(H^{-1/2}(\boldsymbol{\lambda}_\mathbf{k} - \boldsymbol{\lambda}_\mathbf{j})) \omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&\approx \mathcal{O}_\mathbb{P}\left(\frac{\log^2 N}{N} \right) \frac{1}{N|H|^{1/2}} \int \omega(\mathbf{v}) d\mathbf{v} \int K^2(\mathbf{u}) d\mathbf{u} \\
&= \mathcal{O}_\mathbb{P}\left(\frac{\log^2 N}{N^2|H|^{1/2}} \right).
\end{aligned}$$

Let's find a bound for the third addend, $\hat{Q}_{2,1}^3$:

$$\begin{aligned}
\hat{Q}_{2,1}^3 &= 2\frac{(2\pi)^2}{N} \sum_{\mathbf{k}} \sum_{\mathbf{j}} W_\mathbf{j}^2(\boldsymbol{\lambda}_\mathbf{k}) \left(\frac{R_N^1(\boldsymbol{\lambda}_\mathbf{j})}{f(\boldsymbol{\lambda}_\mathbf{j})V} - \frac{R_N^2(\boldsymbol{\lambda}_\mathbf{j})}{f(\boldsymbol{\lambda}_\mathbf{j})V} \right) (LR_\mathbf{j}^2 - LR_\mathbf{j}^1) \omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&\leq 2\frac{(2\pi)^2}{N} \max_{\mathbf{j}} \left| \frac{R_N^1(\boldsymbol{\lambda}_\mathbf{j})}{f(\boldsymbol{\lambda}_\mathbf{j})V} - \frac{R_N^2(\boldsymbol{\lambda}_\mathbf{j})}{f(\boldsymbol{\lambda}_\mathbf{j})V} \right| |LR_\mathbf{j}^2 - LR_\mathbf{j}^1| \\
&\qquad \times \sum_{\mathbf{j}} \sum_{\mathbf{k}} \left(\frac{1}{N^2|H|} K^2(H^{-1/2}(\boldsymbol{\lambda}_\mathbf{k} - \boldsymbol{\lambda}_\mathbf{j})) \right) \omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \mathcal{O}_\mathbb{P}\left(\frac{\log N}{N^{1/2}} \right) \mathcal{O}_\mathbb{P}\left(\frac{\log^2 N}{N} \right) \frac{(2\pi)^6}{N} \\
&\qquad \times \sum_{\mathbf{j}} \sum_{\mathbf{k}} \frac{1}{N^2|H|} K^2(H^{-1/2}(\boldsymbol{\lambda}_\mathbf{k} - \boldsymbol{\lambda}_\mathbf{j})) \omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \mathcal{O}_\mathbb{P}\left(\frac{\log^3 N}{N^{5/2}|H|^{1/2}} \right).
\end{aligned}$$

Therefore,

$$\hat{Q}_2 \leq \mathcal{O}_\mathbb{P}\left(\frac{\log^2 N}{N^2|H|^{1/2}} \right).$$

Now, we will find a bound for $\hat{Q}_3$. This statistic can be written as:

$$\begin{aligned}
\hat{Q}_3 &= 2\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\widehat{m_1}(\boldsymbol{\lambda}_\mathbf{k}) - \widehat{m_2}(\boldsymbol{\lambda}_\mathbf{k})\right)(B^1_\mathbf{k} - B^2_\mathbf{k})\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= 2\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{j}} W_\mathbf{j}(\boldsymbol{\lambda}_\mathbf{k})Y^1_\mathbf{j} - \sum_{\mathbf{i}} W_\mathbf{i}(\boldsymbol{\lambda}_\mathbf{k})Y^2_\mathbf{i}\right)\cdot \\
&\qquad\qquad \left(\sum_{\mathbf{j}} W_\mathbf{j}(\boldsymbol{\lambda}_\mathbf{k})r^1_\mathbf{j} - \sum_{\mathbf{i}} W_\mathbf{i}(\boldsymbol{\lambda}_\mathbf{k})r^2_\mathbf{j}\right)\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&\leq 2\max_{\mathbf{k}}|B^1_\mathbf{k} - B^2_\mathbf{k}|\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}} W_\mathbf{j}(\boldsymbol{\lambda}_\mathbf{k})(Y^1_\mathbf{j} - Y^2_\mathbf{j})\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&\leq \mathcal{O}_\mathbb{P}\left(\frac{\log N}{N^{1/2}}\right)\max_{\mathbf{j}}|Y^1_\mathbf{j} - Y^2_\mathbf{j}|\frac{(2\pi)^6}{N} \\
&\qquad\qquad \times\sum_{\mathbf{k}}\sum_{\mathbf{j}}\frac{1}{N|H|^{1/2}}K(H^{-1/2}(\boldsymbol{\lambda}_\mathbf{k} - \boldsymbol{\lambda}_\mathbf{j}))\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \mathcal{O}_\mathbb{P}\left(\frac{\log^2 N}{N}\right)
\end{aligned}$$

since $H_0 : f_1 = f_2$, it implies that $Y^1_\mathbf{j} - Y^2_\mathbf{j} = r^1_\mathbf{j} - r^2_\mathbf{j}$, for every Fourier frequency. □

Lemma 22. *Under conditions (A1)-(A5) and under H_0, we have that*

$$\sqrt{N^2|H|^{1/2}}\left(\hat{Q}_1 - \frac{(2\pi)^4}{12N|H|^{1/2}}C_K I_\omega\right) \to N(0, \sigma^2_Q),$$

in distribution, where

$$C_K = \int K^2(\mathbf{u})d\mathbf{u}, \quad I_\omega = \int \omega(\mathbf{v})d\mathbf{v} \quad \textit{and the asymptotic variance is}$$

$$\sigma^2_Q = \frac{(2\pi)^8}{72}\int (K * K)^2(\mathbf{u})d\mathbf{u}\int \omega^2(\mathbf{v})d\mathbf{v},$$

where $*$ denotes the convolution operator.

Proof. Define the following random variables:

$$\Lambda_\mathbf{k} = z^{1*}_\mathbf{k} - z^{2*}_\mathbf{k}.$$

Therefore:

$$E(\Lambda_\mathbf{k}) = 0, \quad E(\Lambda^2_\mathbf{k}) = \frac{\pi^2}{3} \quad \text{and} \quad Cov(\Lambda_\mathbf{k}, \Lambda_\mathbf{j}) = 0 \text{ for } \mathbf{j} \neq \mathbf{k}.$$

since the variables $z^{1*}_\mathbf{k}$ and $z^{2*}_\mathbf{k}$ are independent with variance $\pi^2/6$. Under H_0, the difference between log-periodogram values in both population at a fixed Fourier frequency $\boldsymbol{\lambda}_\mathbf{k}$ is given by:

$Y_{\mathbf{k}}^{1*} - Y_{\mathbf{k}}^{2*} = z_{\mathbf{k}}^{1*} - z_{\mathbf{k}}^{2*} = \Lambda_{\mathbf{k}}$. Then, the statistic $\hat{Q}_1$ can be decomposed in two addends, in the following way:

$$\begin{aligned}\hat{Q}_1 &= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\widehat{m_1}^*(\boldsymbol{\lambda}_{\mathbf{k}}) - \widehat{m_2}^*(\boldsymbol{\lambda}_{\mathbf{k}})\right)^2\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\ &= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{j}} W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})Y_{\mathbf{k}}^{1*} - \sum_{\mathbf{i}} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}})Y_{\mathbf{i}}^{2*}\right)^2\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\ &= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{j}} W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})\Lambda_{\mathbf{j}}\right)^2\omega(\boldsymbol{\lambda}_{\mathbf{k}}) = \hat{Q}_{1,1} + \hat{Q}_{1,2},\end{aligned}$$

where $\hat{Q}_{1,1}$ and $\hat{Q}_{1,2}$ are given by:

$$\hat{Q}_{1,1} = \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}} W_{\mathbf{j}}^2(\boldsymbol{\lambda}_{\mathbf{k}})\Lambda_{\mathbf{j}}^2\omega(\boldsymbol{\lambda}_{\mathbf{k}}),$$

$$\hat{Q}_{1,2} = \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}}\sum_{\mathbf{i}\neq\mathbf{j}} W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}})\Lambda_{\mathbf{j}}\Lambda_{\mathbf{i}}\omega(\boldsymbol{\lambda}_{\mathbf{k}}).$$

Define

$$b_{\mathbf{i},\mathbf{j}} = \frac{(2\pi)^2}{N}\sum_{\mathbf{k}} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}})W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})\omega(\boldsymbol{\lambda}_{\mathbf{k}}).$$

Then:

$$\hat{Q}_{1,1} = \sum_{\mathbf{j}} b_{\mathbf{j},\mathbf{j}}\Lambda_{\mathbf{j}}^2, \quad \text{and} \quad \hat{Q}_{1,2} = \sum_{\mathbf{i}\neq\mathbf{j}} b_{\mathbf{i},\mathbf{j}}\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}.$$

First, we will study the behaviour of $\hat{Q}_{1,1}$. For simplicity, consider Priestley-Chao weights, this statistic is given by:

$$\begin{aligned}\hat{Q}_{1,1} &= \sum_{\mathbf{j}} b_{\mathbf{j},\mathbf{j}}\Lambda_{\mathbf{j}}^2 = \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}} W_{\mathbf{j}}^2(\boldsymbol{\lambda}_{\mathbf{k}})\omega(\boldsymbol{\lambda}_{\mathbf{k}})\Lambda_{\mathbf{j}}^2 \\ &= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\omega(\boldsymbol{\lambda}_{\mathbf{k}})\sum_{\mathbf{j}}\frac{(2\pi)^2}{N^2|H|}K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}}))\Lambda_{\mathbf{j}}^2.\end{aligned}$$

Taking expectations and using Riemann approximation, it is easy to see that:

$$\begin{aligned}
&E(\hat{Q}_{1,1}) \\
&= \frac{(2\pi)^2}{N^2|H|^{1/2}} \sum_{\mathbf{j}} E(\Lambda_{\mathbf{j}}^2) \left(\frac{(2\pi)^4}{N|H|^{1/2}} \sum_{\mathbf{k}} K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}}))\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \right) \\
&= \frac{\pi^2}{3} \frac{(2\pi)^2}{N^2|H|^{1/2}} \left(\sum_{\mathbf{k}} \omega(\boldsymbol{\lambda}_{\mathbf{k}}) \frac{(2\pi)^4}{N|H|^{1/2}} \sum_{\mathbf{j}} K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}})) \right) \\
&= \frac{1}{12} \frac{1}{N|H|^{1/2}} \left(\frac{(2\pi)^4}{N} \sum_{\mathbf{k}} \omega(\boldsymbol{\lambda}_{\mathbf{k}}) \frac{(2\pi)^4}{N|H|^{1/2}} \sum_{\mathbf{j}} K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}})) \right) \\
&= \frac{1}{12} \frac{1}{N|H|^{1/2}} \left(\frac{(2\pi)^4}{N} \sum_{\mathbf{k}} \omega(\boldsymbol{\lambda}_{\mathbf{k}}) \frac{(2\pi)^4}{N|H|^{1/2}} \sum_{\mathbf{j}} K_{\mathbf{k}}^2(H^{-1/2}\boldsymbol{\lambda}_{\mathbf{j}}) \right) \\
&\approx \frac{(2\pi)^4}{12} \frac{1}{N|H|^{1/2}} \int \omega(\mathbf{v}) d\mathbf{v} \int K^2(\mathbf{u}) d\mathbf{u}.
\end{aligned}$$

Let's check the order of the variance of $\hat{Q}_{1,1}$. Denote by $c_2 = Var(\Lambda_{\mathbf{j}}^2)$ This variance can be computed taking into account that:

$$Var(\hat{Q}_{1,1}) = \frac{(2\pi)^{12}}{N^6|H|^2} c_2 \sum_{\mathbf{j}} \alpha_{\mathbf{j}}^2,$$

where the coefficients $\alpha_{\mathbf{j}}$ are given by:

$$\alpha_{\mathbf{j}} = \sum_{\mathbf{k}} K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}}))\omega(\boldsymbol{\lambda}_{\mathbf{k}}).$$

Then,

$$\begin{aligned}
&Var(\hat{Q}_{1,1}) = \\
&c_2 \frac{(2\pi)^{12}}{N^6|H|^2} \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \sum_{\mathbf{j}} \omega(\boldsymbol{\lambda}_{\mathbf{k}})\omega(\boldsymbol{\lambda}_{\mathbf{k}'}) \\
&\qquad \times K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}}))K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}'} - \boldsymbol{\lambda}_{\mathbf{j}}))
\end{aligned}$$

which can be approximated, using a changes of variable and Riemann summation, by:

$$Var(\hat{Q}_{1,1}) \approx C_K^2 c_2 \frac{(2\pi)^8}{N^4|H|} \sum_{\mathbf{k}} \omega(\boldsymbol{\lambda}_{\mathbf{k}}) \sum_{\mathbf{k}'} \omega(\boldsymbol{\lambda}_{\mathbf{k}'}) = \mathcal{O}\left(\frac{1}{N^2|H|} \right)$$

Therefore, applying Markov's inequality, it follows that:

$$\hat{Q}_{1,1} = \frac{(2\pi)^4}{12N|H|^{1/2}} C_K I_\omega + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{N|H|^{1/2}} \right).$$

Now, we will check the behaviour of $\hat{Q}_{1,2}$. We will see that this term does not make contributions in bias. Recall the expression of $\hat{Q}_{1,2}$:

$$\hat{Q}_{1,2} = \sum_{\mathbf{i}\neq\mathbf{j}} b_{\mathbf{i},\mathbf{j}}\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}.$$

Then, computing its expectation:

$$E(\hat{Q}_{1,2}) = E\left(\sum_{\mathbf{i}\neq\mathbf{j}} b_{\mathbf{i},\mathbf{j}}\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}\right) = E\left(\sum_{\mathbf{i}\neq\mathbf{j}}\sum_{\mathbf{k}} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}})W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})\omega(\boldsymbol{\lambda}_{\mathbf{k}})\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}\right)$$
$$= \sum_{\mathbf{i}\neq\mathbf{j}}\sum_{\mathbf{k}} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}})W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})\omega(\boldsymbol{\lambda}_{\mathbf{k}})E(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}) = 0,$$

since $\Lambda_{\mathbf{i}}$ and $\Lambda_{\mathbf{j}}$ are uncorrelated, for $\mathbf{i} \neq \mathbf{j}$.

Let's compute the variance of $\hat{Q}_{1,2}$. Since the variables $\Lambda_{\mathbf{j}}$ are i.i.d., we have:

$$Var(\hat{Q}_{1,2}) = \sum_{\mathbf{i}\neq\mathbf{j}}\sum_{\mathbf{u}\neq\mathbf{v}} b_{\mathbf{ij}}b_{\mathbf{uv}}Cov(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}},\Lambda_{\mathbf{u}}\Lambda_{\mathbf{v}})$$
$$= \sum_{\mathbf{i}\neq\mathbf{j}}\sum_{\mathbf{u}\neq\mathbf{v}} b_{\mathbf{ij}}b_{\mathbf{uv}}\left(E(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}\Lambda_{\mathbf{u}}\Lambda_{\mathbf{v}}) - E(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}})E(\Lambda_{\mathbf{u}}\Lambda_{\mathbf{v}})\right)$$
$$= \sum_{\mathbf{i}\neq\mathbf{j}}\sum_{\mathbf{u}\neq\mathbf{v}} b_{\mathbf{ij}}b_{\mathbf{uv}}E(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}\Lambda_{\mathbf{u}}\Lambda_{\mathbf{v}}),$$

since $\mathbf{i} \neq \mathbf{j}$ and therefore: $E(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}) = E(\Lambda_{\mathbf{i}})E(\Lambda_{\mathbf{j}}) = 0$. The same applies for $\mathbf{u} \neq \mathbf{j}$. For $E(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}\Lambda_{\mathbf{u}}\Lambda_{\mathbf{v}})$ to be different from zero, one of these two conditions hold: $\mathbf{i} = \mathbf{u}$ and $\mathbf{j} = \mathbf{v}$ or $\mathbf{i} = \mathbf{v}$ and $\mathbf{j} = \mathbf{u}$. Then:

$$Var(\hat{Q}_{1,2}) = \sum_{\mathbf{i}}\sum_{\mathbf{j}\neq\mathbf{i}} b_{\mathbf{ij}}\sum_{\mathbf{u}}\sum_{\mathbf{u}\neq\mathbf{v}} b_{\mathbf{uv}}E(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}\Lambda_{\mathbf{u}}\Lambda_{\mathbf{v}})$$
$$= \sum_{\mathbf{i}}\sum_{\mathbf{j}\neq\mathbf{i}} b_{\mathbf{ij}}\sum_{\mathbf{u}=\mathbf{i}}\sum_{\mathbf{v}=\mathbf{j}} b_{\mathbf{uv}}E(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}\Lambda_{\mathbf{u}}\Lambda_{\mathbf{v}})$$
$$+ \sum_{\mathbf{i}}\sum_{\mathbf{j}\neq\mathbf{i}} b_{\mathbf{ij}}\sum_{\mathbf{u}=\mathbf{j}}\sum_{\mathbf{v}=\mathbf{i}} b_{\mathbf{uv}}E(\Lambda_{\mathbf{i}}\Lambda_{\mathbf{j}}\Lambda_{\mathbf{u}}\Lambda_{\mathbf{v}})$$
$$= 2\sum_{\mathbf{i}}\sum_{\mathbf{j}\neq\mathbf{i}} b_{\mathbf{ij}}^2E(\Lambda_{\mathbf{i}}^2)E(\Lambda_{\mathbf{j}}^2) = \frac{2\pi^4}{9}\sum_{\mathbf{i}}\sum_{\mathbf{j}\neq\mathbf{i}} b_{\mathbf{ij}}^2.$$

Consider the following approximation for the product of two $b_{\mathbf{ij}}$ coefficients:

$$b_{\mathbf{ij}}b_{\mathbf{uv}} \approx \frac{(2\pi)^4}{N^4|H|}K * K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{i}} - \boldsymbol{\lambda}_{\mathbf{j}}))\omega(\boldsymbol{\lambda}_{\mathbf{i}})K * K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{u}} - \boldsymbol{\lambda}_{\mathbf{v}}))\omega(\boldsymbol{\lambda}_{\mathbf{u}}).$$

Then,

$$Var(\hat{Q}_{1,2}) \approx \frac{2\pi^4}{9} \sum_{\mathbf{i}} \sum_{\mathbf{j}\neq\mathbf{i}} \frac{(2\pi)^8}{N^4|H|} (K * K)^2 (H^{-1/2}(\boldsymbol{\lambda}_\mathbf{i} - \boldsymbol{\lambda}_\mathbf{j}))\omega^2(\boldsymbol{\lambda}_\mathbf{i})$$
$$\approx \frac{(2\pi)^8}{72} \frac{1}{N|H|^{1/2}} \int (K * K)^2(\mathbf{u}) d\mathbf{u} \int \omega(\mathbf{v}) d\mathbf{v}.$$

Therefore, the asymptotic variance of $\hat{Q}$ is given by:

$$\sigma^2_{\hat{Q}} = \lim_{N\to\infty} N^2|H|^{1/2}\sigma^2_{\hat{Q}_{1,2}} = \frac{(2\pi)^8}{72} \int (K * K(\mathbf{u}))^2 \, d\mathbf{u} \int \omega^2(\mathbf{v}) d\mathbf{v}.$$

In order to prove the asymptotic normal distribution of $\hat{Q}_{1,2}$, we will apply Theorem 5.2 in de Jong (1987). For that purpose, we must write $\hat{Q}_{1,2}$ as a quadratic form, namely $\hat{Q}_{1,2} = \sum_{i,j} a_{i,j} X_i X_j$, where i and j are one-dimensional indexes and X_i are i.i.d. random variables with zero mean and unit variance.

First, define a new subindex for the Fourier frequencies $\boldsymbol{\lambda}_\mathbf{k}$, with $\mathbf{k} = (k_1, k_2)$ and $k_l = 0, \pm 1, \ldots, \pm n_l$, for $l = 1, 2$. Consider $\boldsymbol{\lambda}_\mathbf{k} = \boldsymbol{\lambda}_{\mathbf{k}'}$ where $\mathbf{k}' = (k'_1, k'_2)$, with $k'_l = 1, \ldots, m_l = 2n_l + 1$, in such a way that $k'_l = k_l + n_1 + 1$ for $l = 1, 2$. Recall that $N = m_1 \times m_2$. Denote by $\mathcal{M}_{N\times N}$ the space of square matrix with size N. The new coefficients, with one dimensional indexes, are given by the following matrix:

$$A = (a_{ij}), \quad A \in \mathcal{M}_{N\times N},$$

and each entry of this matrix is defined by $a_{ij} = \frac{\pi}{\sqrt{3}} b_\mathbf{ij}$ and $a_{ii} = 0$, where the bidimensional indexes are given by:

$$\mathbf{i} = (i_1, i_2), \quad \text{if } (i_1 - 1)m_2 \leq i \leq i_1 m_2 \quad \text{and } i = (i_1 - 1)m_2 + i_2, \tag{5.34}$$

$$\mathbf{j} = (j_1, j_2), \quad \text{if } (j_1 - 1)m_2 \leq j \leq j_1 m_2 \quad \text{and } j = (j_1 - 1)m_2 + j_2. \tag{5.35}$$

Now, define the variables:

$$X_i = \frac{\sqrt{3}}{\pi} \Lambda_\mathbf{i}, \quad \text{where } i = (i_1 - 1)m_2 + i_2.$$

With this definitions, $\hat{Q}_{1,2}$ can be written as a quadratic form with one-dimensional indexes:

$$\hat{Q}_{1,2} = \sum_{i,j} a_{i,j} X_i X_j.$$

Asymptotic normality is proved if the following conditions are satisfied:

(i) There exists a sequence of real numbers $k(N)$ such that:

$$k(N)^4 \sigma^2_{\hat{Q}} \max_i \sum_j a^2_{ij} \to 0, \quad N \to \infty.$$

(ii) The random variables X_i satisfy:

$$\max_i E\left(X_i^2 1_{\{|X_i|>k(N)\}}\right) \to 0, \quad N \to \infty.$$

(iii) The eigenvalues of the matrix $A = (a_{ij})$ are negligible:

$$\sigma_{\hat{Q}}^2 \max_i \mu_i^2 \to 0, \quad N \to \infty.$$

Let's prove that condition (i) is satisfied. The coefficients in the quadratic form are given by:

$$a_{ij} = \frac{\pi}{\sqrt{3}} \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}}) W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) \omega(\boldsymbol{\lambda}_{\mathbf{k}}),$$

where $\mathbf{i}$ and $\mathbf{j}$ are determined by (5.34) and (5.35), respectively. Then,

$$\begin{aligned}
a_{ij}^2 &= \frac{\pi^2(2\pi)^4}{3N^2} \left(\sum_{\mathbf{k}} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}}) W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) \omega(\boldsymbol{\lambda}_{\mathbf{k}}) \right)^2 \\
&= \frac{\pi^2(2\pi)^4}{3N^2} \sum_{\mathbf{k}} W_{\mathbf{i}}^2(\boldsymbol{\lambda}_{\mathbf{k}}) W_{\mathbf{j}}^2(\boldsymbol{\lambda}_{\mathbf{k}}) \omega^2(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&+ \frac{\pi^2(2\pi)^4}{3N^2} \sum_{\mathbf{k}} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}}) W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) \omega(\boldsymbol{\lambda}_{\mathbf{k}}) \sum_{\mathbf{k}\neq\mathbf{k}'} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}'}) W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}'}) \omega(\boldsymbol{\lambda}_{\mathbf{k}'}) \\
&= a_{ij}^{2A} + a_{ij}^{2B}.
\end{aligned}$$

Consider now:

$$\max_i \sum_j a_{ij}^{2A} = \max_i \sum_j \frac{\pi^2(2\pi)^4}{3N^2} \sum_{\mathbf{k}} W_{\mathbf{i}}^2(\boldsymbol{\lambda}_{\mathbf{k}}) W_{\mathbf{j}}^2(\boldsymbol{\lambda}_{\mathbf{k}}) \omega^2(\boldsymbol{\lambda}_{\mathbf{k}}).$$

This term can be approximated (up to a factor $(2\pi)^4$) by:

$$\begin{aligned}
\max_i \sum_j a_{ij}^{2A} &= \frac{\pi^2(2\pi)^4}{3N^2} \max_{\mathbf{i}} \sum_{\mathbf{k}} \omega^2(\boldsymbol{\lambda}_{\mathbf{k}}) W_{\mathbf{i}}^2(\boldsymbol{\lambda}_{\mathbf{k}}) \sum_{\mathbf{j}} W_{\mathbf{j}}^2(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&= \frac{\pi^2(2\pi)^4}{3N^2} \max_{\mathbf{i}} \sum_{\mathbf{k}} \omega^2(\boldsymbol{\lambda}_{\mathbf{k}}) \frac{1}{N^2|H|} K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{i}})) \\
&\qquad \times \sum_{\mathbf{j}} \frac{1}{N^2|H|} K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}})) \\
&\approx \frac{\pi^2(2\pi)^2}{3N^3} \max_{\mathbf{i}} \sum_{\mathbf{k}} \omega^2(\boldsymbol{\lambda}_{\mathbf{k}}) \frac{1}{N^2|H|} K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{i}})) C_K \\
&= C_K \frac{\pi^2}{3N^4|H|^{1/2}} \max_{\mathbf{i}} \frac{(2\pi)^2}{N|H|^{1/2}} \sum_{\mathbf{k}} \omega^2(\boldsymbol{\lambda}_{\mathbf{k}}) K^2(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{i}})) \\
&\approx C_K \frac{\pi^2}{3N^4|H|^{1/2}} \max_{\mathbf{i}} (\omega^2 * K^2)(\boldsymbol{\lambda}_{\mathbf{i}}) = \mathcal{O}\left(N^{-4}|H|^{-1/2}\right).
\end{aligned}$$

Following similar arguments, the same rate is obtained for $\max_i \sum_j a_{ij}^{1B}$ Therefore,

$$k^4(N)\sigma_{\hat{Q}}^2 \max_i \sum_j a_{ij} = \mathcal{O}\left(\frac{k^4(N)}{N^6|H|}\right),$$

which tends to zero if the sequence $k(N) \to \infty$ satisfies that $\dfrac{k^4(N)}{N^6|H|} \to 0$.

Condition (ii) holds since the variables X_i are i.i.d. with second order moment $E(X_i^2) = 1$. It remains to show that condition (iii) also holds. Since A is a symmetric $N \times N$ matrix, there exists an ortogonal matrix U such that $U^{-1}AU$ is diagonal. This result implies that A is diagonalizable with real eigenvalues, $\{\mu_i\}$, with $i = 1, \ldots, N$. The $\|\cdot\|_\infty$ norm of the matrix A is given by:

$$\|A\|_\infty = \max_i \sum_j |a_{ij}|,$$

The spectral ratio of the matrix is defined as the maximum of the eigenvalues of A and it can be bounded by any norm in the matrix space $\mathcal{M}_{N\times N}$; therefore, for the particular case of the supremun norm $\|\cdot\|_\infty$:

$$\max_i |\mu_i| \leq \max_i \sum_j |a_{ij}|.$$

Besides,

$$\max_i |\mu_i|^2 \leq \left(\max_i |\mu_i|\right)^2 \leq \left(\max_i \sum_j |a_{ij}|\right)^2,$$

so we will give a bound for the last term in the inequality.

$$\begin{aligned}
\max_i \sum_j |a_{ij}| &= \max_{\mathbf{i}} \sum_{\mathbf{j}} \frac{\pi(2\pi)^2}{\sqrt{3}N} \left|\sum_{\mathbf{k}} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}}) W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}}) \omega(\boldsymbol{\lambda}_{\mathbf{k}})\right| \\
&\leq \frac{\pi(2\pi)^2}{\sqrt{3}N} \frac{(2\pi)^4}{N^2|H|} \max_{\mathbf{i}} \sum_{\mathbf{j}} \sum_{\mathbf{k}} \left|K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{i}}))\right| \\
&\qquad \times \left|K(H^{-1/2}(\boldsymbol{\lambda}_{\mathbf{k}} - \boldsymbol{\lambda}_{\mathbf{j}}))\right| |\omega(\boldsymbol{\lambda}_{\mathbf{k}})| \\
&\approx \frac{2^4\pi^5}{\sqrt{3}N(2\pi)^2} \max_{\mathbf{i}} (\omega * K)(\boldsymbol{\lambda}_{\mathbf{i}}) \int K(\mathbf{u}) d\mathbf{u} = \mathcal{O}(N^{-1}).
\end{aligned}$$

Therefore,

$$\sigma_{\hat{Q}}^2 \max_i |\mu_i|^2 = \mathcal{O}\left(\frac{1}{N^4|H|^{1/2}}\right) \to 0.$$

Then, the asymptotic convergence to a normal distribution is proved. □

Proof of Theorem 5. The theorem is proved combining the results from Lemmas 20-22. □

5.5.2 Proof of Theorem 6

Proof of Theorem 6. Consider the decomposition of the test statistic given in Lemma 20: $\hat{Q} = \hat{Q}_{1a} + \hat{Q}_{2a} + \hat{Q}_{3a}$. The sketch of the proof is as follows: first, we will find bounds for $\hat{Q}_{2a}$ and $\hat{Q}_{3a}$. Second, $\hat{Q}_{1a}$ is decomposed in three addends $\hat{Q}_{1a1}$, $\hat{Q}_{1a2}$ and $\hat{Q}_1$. The asymptotic normality of $\hat{Q}_1$ is proved in Theorem 5. We will find a bound for $\hat{Q}_{1a2}$ and we will derive the asymptotic value of $\hat{Q}_{1a1}$.

From Lemma 21, it is obvious that $\hat{Q}_{2a} = \mathcal{O}_{\mathbb{P}}\left(\frac{\log^2 N}{N^2|H|^{1/2}}\right)$, since the equality of the spectral densities in both models is not necessary. Let's find a bound for $\hat{Q}_{3a}$.

$$\begin{aligned}
\hat{Q}_{3a} &= 2\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}(\widehat{m_1}(\boldsymbol{\lambda}_{\mathbf{k}}) - \widehat{m_2}(\boldsymbol{\lambda}_{\mathbf{k}}))(B^1_{\mathbf{k}} - B^2_{\mathbf{k}})\omega(\boldsymbol{\lambda}_{\mathbf{k}}) \\
&\leq 2\max_{\mathbf{k}}|B^1_{\mathbf{k}} - B^2_{\mathbf{k}}|\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}}W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})(C_N p(\boldsymbol{\lambda}_{\mathbf{j}}) + r^1_{\mathbf{j}} - r^2_{\mathbf{j}}) \\
&= 2\max_{\mathbf{k}}|B^1_{\mathbf{k}} - B^2_{\mathbf{k}}|C_N\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}}W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})p(\boldsymbol{\lambda}_{\mathbf{j}}) \\
&+2\max_{\mathbf{k}}|B^1_{\mathbf{k}} - B^2_{\mathbf{k}}|\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\sum_{\mathbf{j}}W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})(r^1_{\mathbf{j}} - r^2_{\mathbf{j}}) \\
&= \hat{Q}^1_{3a} + \hat{Q}^2_{3a}.
\end{aligned}$$

From Lemma 21, it is obvious that:

$$\hat{Q}^2_{3a} \leq \mathcal{O}_{\mathbb{P}}\left(\frac{\log^2 N}{N}\right),$$

and for $\hat{Q}^1_{3a}$, proceeding in a similar way, we obtain:

$$\hat{Q}^1_{3a} \leq \mathcal{O}_{\mathbb{P}}\left(\frac{\log^2 N}{N^{3/2}|H|^{1/8}}\right).$$

Therefore,

$$\hat{Q}_{3a} = \mathcal{O}_{\mathbb{P}}\left(\frac{\log^2 N}{N}\right).$$

The first addend in the decomposition of $\hat{Q}$ can be written as:

$$\begin{aligned}
\hat{Q}_{1a} &= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\widehat{m^1}^*(\boldsymbol{\lambda}_\mathbf{k}) - \widehat{m^2}^*(\boldsymbol{\lambda}_\mathbf{k})\right)^2\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{i}} W_\mathbf{i}(\boldsymbol{\lambda}_\mathbf{k})Y_\mathbf{i}^{1*} - \sum_{\mathbf{j}} W_\mathbf{j}(\boldsymbol{\lambda}_\mathbf{k})Y_\mathbf{j}^{2*}\right)^2\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{i}} W_\mathbf{i}(\boldsymbol{\lambda}_\mathbf{k})(m^1(\boldsymbol{\lambda}_\mathbf{k}) - m^2(\boldsymbol{\lambda}_\mathbf{k}) + z_\mathbf{k}^{1*} - z_\mathbf{k}^{2*})\right)^2\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{i}} W_\mathbf{i}(\boldsymbol{\lambda}_\mathbf{k})C_N p(\boldsymbol{\lambda}_\mathbf{k}) + \sum_{\mathbf{j}} W_\mathbf{j}(\boldsymbol{\lambda}_\mathbf{k})\Lambda_\mathbf{j}\right)^2\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{i}} W_\mathbf{i}(\boldsymbol{\lambda}_\mathbf{k})C_N p(\boldsymbol{\lambda}_\mathbf{k})\right)^2\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&+ \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{j}} W_\mathbf{j}(\boldsymbol{\lambda}_\mathbf{k})\Lambda_\mathbf{j}\right)^2\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&+ 2\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{i}} W_\mathbf{i}(\boldsymbol{\lambda}_\mathbf{k})C_N p(\boldsymbol{\lambda}_\mathbf{k})\right)\left(\sum_{\mathbf{j}} W_\mathbf{j}(\boldsymbol{\lambda}_\mathbf{k})\Lambda_\mathbf{j}\right)\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \hat{Q}_{1a1} + \hat{Q}_1 + \hat{Q}_{1a2}.
\end{aligned}$$

The asymptotic distribution of $\hat{Q}_1$ has been proved in Theorem 6. We will study the asymptotic behaviour of $\hat{Q}_{1a1}$.

$$\begin{aligned}
\hat{Q}_{1a1} &= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{i}} W_\mathbf{i}(\boldsymbol{\lambda}_\mathbf{k})C_N p(\boldsymbol{\lambda}_\mathbf{k})\right)^2\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{i}} \frac{(2\pi)^2}{N|H|^{1/2}}K(H^{-1/2}(\boldsymbol{\lambda}_\mathbf{k} - \boldsymbol{\lambda}_\mathbf{i}))C_N p(\boldsymbol{\lambda}_\mathbf{i})\right)^2\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&= \frac{(2\pi)^2}{N}C_N\frac{(2\pi)^2}{N^2|H|}\sum_{\mathbf{k}}\left(K(H^{-1/2}(\boldsymbol{\lambda}_\mathbf{k} - \boldsymbol{\lambda}_\mathbf{i}))p(\boldsymbol{\lambda}_\mathbf{i})\right)^2\omega(\boldsymbol{\lambda}_\mathbf{k}) \\
&\approx \int p^2(\mathbf{v})\omega(\mathbf{v})d\mathbf{v},
\end{aligned}$$

where the last approximation holds if $C_N^2 = N^{-1}|H|^{-1/4}$.

We must find a bound for $\hat{Q}_{1a2},$. The expected value of this term is zero, since $E(\Lambda) = 0$. And

for the variance, we have that:

$$
\begin{aligned}
&Var(\hat{Q}_{1a2}) = \\
&= Var\left(2\frac{(2\pi)^2}{N}\sum_{\mathbf{k}}\left(\sum_{\mathbf{i}} W_{\mathbf{i}}(\boldsymbol{\lambda}_{\mathbf{k}})C_N p(\boldsymbol{\lambda}_{\mathbf{i}})\right)\left(\sum_{\mathbf{j}} W_{\mathbf{j}}(\boldsymbol{\lambda}_{\mathbf{k}})\Lambda_{\mathbf{j}}\right)\omega(\boldsymbol{\lambda}_{\mathbf{k}})\right) \\
&= 4\frac{(2\pi)^4}{N^2}C_N^2\sum_{\mathbf{k}}\sum_{\mathbf{i}}\sum_{\mathbf{j}} W_{\mathbf{i}}^2(\boldsymbol{\lambda}_{\mathbf{k}})W_{\mathbf{j}}^2(\boldsymbol{\lambda}_{\mathbf{k}})p^2(\boldsymbol{\lambda}_{\mathbf{i}})\omega^2(\boldsymbol{\lambda}_{\mathbf{k}})Var(\Lambda_{\mathbf{j}}) \\
&= \mathcal{O}_{\mathbb{P}}\left(\frac{C_N^2}{N^2|H|^{1/4}}\right) = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{N^3|H|^{5/4}}\right).
\end{aligned}
$$

Then, applying the Markov inequality,

$$
\hat{Q}_{1a2} = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{N^{3/2}|H|^{5/8}}\right).
$$

□

Further research

We will describe briefly some of the future research lines we have started to explored. Keeping our interest focused on the dependence structure of spatial processes and particularly, on the construction of goodness-of-fit test techniques (Chapter 4). An alternative could be to consider goodness-of-fit tests based on Empirical Processes. From another point of view, our theoretical developments could be translated into the wavelet spectrum, and the goodness-of-fit problem could be tackled from this context.

One of the assumptions on the spatial design, related to classical Fourier analysis, along this dissertation is that sampling locations are regularly spaced as it happens, for instance, with satellite data. But in many applications in geostatistics, this restriction may be too strong. Considering the Wavelet Transform overcomes this drawback.

Chapter 3 in this manuscript was devoted to simulation methods for spatially dependent data, all of them under stationarity and gaussianity assumptions. We will try to extend the spectral simulation methods to a non-gaussian non-stationarity situation, by considering resampling techniques on the spectral domain.

In Chapter 5, we pointed out the interest of the test for comparing spatial spectral densities in the spatio-temporal context. If the spatial dependence remains invariant along time, then the process is temporal stationary and simple spatio-temporal models for describing the dependence structure could be used. A class of simple models for spatio-temporal dependence are separable covariances. From the spectral domain setting, we can also propose a test for separability.

Finally, even though our main interest has been focused on the dependence structure, the small-scale variability, we provide some ideas for constructing a goodness-of-fit test for the large-scale variability component.

Goodness-of-fit test based for the spatial spectral density on Empirical Processes.

As we have seen in the overview on goodness-of-fit test techniques for regression models (Chapter 4, Section 4.1), in order to build a goodness-of-fit test for the spatial log-spectral density, we have chosen a generalized likelihood ratio test statistics, but other choices would be also possible.

An appealing alternative is the construction of goodness-of-fit tests based on Empirical Processes. In the spatial statistics literature, Lahiri (1999) consider a empirical process in order to determine the asymptotic distribution of the empirical spatial cumulative distribution function (ESCDF) predictor in (5.21).

From (2.15) we could consider a nonparametric estimator of the integrated regression function $m = \log f$:

$$I(\boldsymbol{\lambda}) = \frac{1}{N} \sum_{\mathbf{k}} \mathbf{1}_{\{\boldsymbol{\lambda}_{\mathbf{k}} < \boldsymbol{\lambda}\}} Y_{\mathbf{k}}, \tag{5.36}$$

where $<$ denotes the lexicographic order in $\mathbb{Z}^2$ and $\sum_{\mathbf{k}}$ denotes the sum over all the Fourier frequencies. An empirical process on this context could be defined as:

$$\eta_N(\boldsymbol{\lambda}) = \frac{1}{N} \sum_{\mathbf{k}} \mathbf{1}_{\{\boldsymbol{\lambda}_{\mathbf{k}} < \boldsymbol{\lambda}\}} (Y_{\mathbf{k}} - m(\boldsymbol{\lambda}_{\mathbf{k}})). \tag{5.37}$$

In order to study the asymptotic distribution of (5.37), we must note that we have fixed design points (the Fourier frequencies) and we somehow control the distribution of the error term. The distribution of the $z_{\mathbf{k}}$ variables is known and the behaviour of the $r_{\mathbf{k}}$ term can be uniformly bounded. A similar problem has been studied in Diebolt and Zuber (2001), for the regression context but the basis of the theory developed in their paper can be found in Stute (1997).

Since the spectral density is unknown, we must consider an estimated empirical process:

$$\hat{\eta}_N(\boldsymbol{\lambda}) = \frac{1}{N} \sum_{\mathbf{k}} \mathbf{1}_{\{\boldsymbol{\lambda}_{\mathbf{k}} < \boldsymbol{\lambda}\}} (Y_{\mathbf{k}} - m_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}})),$$

where the parameter estimator satisfy a consistency requirement. The critical issue in the theoretical developments is proving the tightness of the empirical process (5.37).

Wavelet approach.

The goodness-of-fit test statistic (4.10) based on the spatial periodogram can be adapted in terms of the scalogram (the *periodogram* for Wavelet Transform). The introduction of wavelet techniques

allows for more flexibility in the statement of the testing problems.

For a spatial random field Z defined on $\mathbb{R}^d$, its dependence structure can be characterized in terms of a continuous spectrum of the covariance operator. In particular, if Z is a second-order stationary process, this continuous spectrum coincides with the spectral density.

Besides, the covariance operator generates a bilinear form which defines the inner product of the associated reproducing kernel Hilbert space $\mathcal{H}(Z)$. If this space is isomorphic to a fractional Sobolev space, then Z admits an orthogonal decomposition:

$$Z(s) = \sum_j \sum_{\mathbf{k}} \gamma_{j\mathbf{k}}(s) Z_{j\mathbf{k}} + \sum_{\mathbf{k}} \varphi_{0\mathbf{k}}(s) Z_{0\mathbf{k}}, \tag{5.38}$$

where $\gamma_{j\mathbf{k}}$ and $\varphi_{0\mathbf{k}}$ are defined in terms of the kernel generating by self-convolution the covariance and the wavelet basis (see Angulo and Ruiz-Medina (1999)). Apart from the technical details and mathematical concepts that encloses this approach, representation (5.38) reminds us to the moving average representation considered in our developments.

Definition 3. *The extended-scalogram, at a level j and translations $\mathbf{k}$ and $\mathbf{l}$, is defined as:*

$$S(j,\mathbf{k},\mathbf{l}) = \left(\sum_s Z(s)\psi_{j\mathbf{k}}(s) \right) \overline{\left(\sum_v Z(v)\psi_{j\mathbf{l}}(v) \right)}. \tag{5.39}$$

The expectation of the scalogram is the Wavelet Transform of the covariance, at level j and translations $\mathbf{k}$ and $\mathbf{l}$. For a fixed level of resolution j, a goodness-of-fit test for the covariance of a spatial process could be formulated from the wavelet spectrum, as a Weighted Least Square criterion on the difference:

$$\frac{S(j,\mathbf{k},\mathbf{l})}{TW^{2D}(\hat{C})_{(j,\mathbf{k},\mathbf{l})}} - 1,$$

where $TW^{2D}(\hat{C})_{(j,\mathbf{k},\mathbf{l})}$ denotes the $2D$-wavelet transform of an estimator $\hat{C}$ of the covariance function.

A nonparametric resampling method.

The uniform distribution of the phases is a sufficient condition for the stationarity of the process. Besides, the Gaussian distribution of the real and imaginary parts of the process (regardless the distribution of the spatial process) is only achieved asymptotically. Then, the algorithms presented in the Chapter 3 may not perform well for a non-Gaussian context. The classical spectral simulation method can be extended to a non-Gaussian situation.

Since conditions $(i)-(v)$ in Section 3.1.1 hold for any stationary spatial process Z, and such a process admits a representation in a Fourier-Stieltjes form as (1.62), it is possible from a realization of Z obtain a discrete approximation to the associated spectral process $\mathcal{Y}$ given by:

$$\mathcal{Y}(\boldsymbol{\lambda}_\mathbf{k}) = \frac{1}{(2\pi)^2 N} \sum_{\mathbf{s}\in D} Z(\mathbf{s}) e^{-i\mathbf{s}^T \boldsymbol{\lambda}_\mathbf{k}} \tag{5.40}$$

This spectral complex process can be splitted in its real and imaginary parts, $U(\boldsymbol{\lambda}_\mathbf{k})$ and $V(\boldsymbol{\lambda}_\mathbf{k})$. For a finite set of observations, that is, from a realization of the spatial process Z, we may not be able to determine the distribution of U and V particularly if the distribution of Z is far from being Gaussian.

1. $\{Z(\mathbf{s}), \mathbf{s} \in D \subset \mathbb{R}^2\}$, observed data, with $D = \{0, \ldots, n_1 - 1\} \times \{0, \ldots, n_2 - 1\}$ and denote by $N = n_1 \times n_2$. Assume the covariance function C is known.

2. Get $\{\mathcal{Y}(\boldsymbol{\lambda}_\mathbf{k}), \boldsymbol{\lambda}_\mathbf{k} \in \Pi^2\}$ the associated spectral process given by (5.40) at the Fourier frequencies and obtain the real and imaginary parts of the process:

$$U(\boldsymbol{\lambda}_\mathbf{k}) = \frac{1}{(2\pi)^2 N} \sum_{\mathbf{s}\in D} Z(\mathbf{s}) \cos(\mathbf{s}^T \boldsymbol{\lambda}_\mathbf{k}),$$

$$V(\boldsymbol{\lambda}_\mathbf{k}) = \frac{1}{(2\pi)^2 N} \sum_{\mathbf{s}\in D} Z(\mathbf{s}) \sin(\mathbf{s}^T \boldsymbol{\lambda}_\mathbf{k}).$$

3. $U(\boldsymbol{\lambda}_\mathbf{k})$ have the same distribution for all $\boldsymbol{\lambda}_\mathbf{k}$ and we have seen that these random variables are uncorrelated. Consider

$$\tilde{U}(\boldsymbol{\lambda}_\mathbf{k}) = \sqrt{\frac{2(2\pi)^2}{\sigma^2_\mathbf{k}}} U(\boldsymbol{\lambda}_\mathbf{k}),$$

   for $\boldsymbol{\lambda}_\mathbf{k} \neq (0,0)$. The rescaled sample $\{\tilde{U}(\boldsymbol{\lambda}_\mathbf{k})\}$ can ben considered as a random sample of the distribution $F^{\tilde{U}}$. Consider also the empirical distribution $F_N^{\tilde{U}}$ and draw a sample from this distribution: $\{\tilde{U}^*(\boldsymbol{\lambda}_\mathbf{k})\}$.

4. Proceed for $V(\boldsymbol{\lambda}_\mathbf{k})$ as we have done with $U(\boldsymbol{\lambda}_\mathbf{k})$, and get, from $F_N^{\tilde{V}}$ a random sample $\{\tilde{V}^*(\boldsymbol{\lambda}_\mathbf{k})\}$.

5. Rescaled the generated samples and get $U^*(\boldsymbol{\lambda}_\mathbf{k}) = \sqrt{\frac{\sigma^2_\mathbf{k}}{2(2\pi)^2}} \tilde{U}^*(\boldsymbol{\lambda}_\mathbf{k})$ (similarly, $V^*(\boldsymbol{\lambda}_\mathbf{k})$), for $\boldsymbol{\lambda}_\mathbf{k} \neq (0,0)$.

6. For the origin, assuming that $E(Z(\mathbf{s})) = 0$, draw $U^*(0,0) \sim N\left(0, \frac{\sigma_0^2}{(2\pi)^2}\right)$. In practice, we must estimate the mean from the data by $\bar{Z}$ and simulate $U^*(0,0) \sim N\left(\bar{Z}, \frac{\sigma_0^2}{(2\pi)^2}\right)$.

7. Build a realization of the spectral process as:

$$\mathcal{Y}^*(\boldsymbol{\lambda}_\mathbf{k}) = U^*(\boldsymbol{\lambda}_\mathbf{k}) + iV^*(\boldsymbol{\lambda}_\mathbf{k}).$$

8. Consider the inverse discrete Fourier transform of the realization of the spectral process $\mathcal{Y}^*(\boldsymbol{\lambda}_\mathbf{k})$ and get $Z^*(\mathbf{s})$, for $\mathbf{s} \in \mathcal{D}$.

This resampling method is consistent in the sense that the distribution of the simulated spectral process $\mathcal{Y}^*$ converges to the distribution of $\mathcal{Y}$. It is straightforward just regarding the independence of the real and imaginary parts and applying Glivenko-Cantelli's theorem for the uniform convergence of the empirical distribution. Some considerations on the continuous case must be made.

A test for separability.

As we have already commented, applications in many scientific fields involve spatio-temporal processes. In these cases, the interested is not only focused on the spatial dependence structure, but on the spatio-temporal one. The simplest way for modelling spatio-temporal processes is by considering a separable covariance function.

A spatio-temporal process $Z(\mathbf{s}, t)$ with covariance function $C(\mathbf{u}, h) = Cov(Z(\mathbf{s}, \mathbf{t}), Z(\mathbf{s}+\mathbf{u}, t+h))$, is said to be separable if C can be factorized as: $C(\mathbf{u}, t) = C_S(\mathbf{u})C_T(t)$, where C_S and C_T are spatial and temporal covariance functions, respectively. However, separable models are not always adequate for describing the spatio-temporal dependence structure. In this context, Fuentes (2006b) proposes a formal test for separability. The test is based on the spectral representation of the process and it consists basically in studying whether the coherence function is constant across frequencies. An advantage of this method is that it does not neet data to be regularly spaced.

Our idea is also based on spectral analysis, but on a different way. Since the log-periodogram can be written as the response variable in a regression model:

$$\log I(\boldsymbol{\lambda}_\mathbf{k}, \tau) = m(\boldsymbol{\lambda}_\mathbf{k}, \tau) + z_{\mathbf{k},\tau} + r_{\mathbf{k},\tau}. \tag{5.41}$$

If we consider a spatio-temporal process (and a spatio-temporal periodogram), testing for the separability of the process will be equivalent to test for the additivity of the log-spectral density, as the regression function in (5.41).

Goodness-of-fit test for the trend function.

Large-scale variability is often incorporated into the mean structure of a model, along with any relevant covariates. Trend terms may be incorporated in a simple way without any scientific knowledge. Sometimes, this is accomplished through the use of nonparametric smoothing functions. The goal of incorporating such trend terms in a model is often to produce some form of stationarity in the residuals. For instance, the estimation of trends for spatio-temporal processes in airborne concentrations, has been the subject of several works using different methods such as linear regression models, nonparametric models or generalized additive models (e.g. Lefohn and Shadwick (1991), Holland *et al.* (2000)).

Our aim would be to test the validity of an estimated trend function through the study of an empirical process of the error term.

Consider a spatial (or spatio-temporal) process Z, observed on a regular grid with spacing between neighbour coordinates Δ, with the following decomposition:

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \varepsilon(\mathbf{s}), \tag{5.42}$$

where μ is a deterministic trend function and ε is the error term which models the dependence structure, not necessarily weakly stationary. Similarly to the Integrated Regression Function introduced in (4.4), we can define the *Cummulative Mean Function* (CMF), for a fixed $\mathbf{s}_0 \in \mathbb{R}^{2+}$:

$$F(\mathbf{s}_0) = \int_{\mathbf{0}}^{\mathbf{s}_0} \mu(\mathbf{s}) dP(\mathbf{s}),$$

where P is the distribution of the spatial locations. Under the condition of uniformly distributed locations over an observation region $D \subset \mathbb{R}^2$ (regular grid), we can write the CMF as follows:

$$F(\mathbf{s}_0) = \frac{1}{|D|} \int_D \mu(\mathbf{s}) \mathbf{1}_{\{\mathbf{s} \leq \mathbf{s}_0\}} d\mathbf{s},$$

where $|D|$ denotes the volume of D, and a natural estimator for the CMF is given by:

$$F_N(\mathbf{s}_0) = \frac{1}{N} \sum_{\mathbf{s} \in D} Z(\Delta\mathbf{s}) \mathbf{1}_{\{\Delta\mathbf{s} \leq \mathbf{s}_0\}}.$$

We define the Empirial Spatial Process (ESP) as:

$$\eta_{N,\Delta}(\mathbf{s}_0) = \frac{\Delta^2}{N} \sum_{\mathbf{s} \in D} \varepsilon(\Delta\mathbf{s}) \mathbf{1}_{\{\Delta\mathbf{s} \leq \mathbf{s}_0\}}, \tag{5.43}$$

where Δ denots the spacing between neighbour coordinates in the grid, as usual. We assume, for simplicity, equal spacement in all directions. Assume that the error process is second-order

stationary. Considering the spectral representation of the error process ε, as it has been introduced in Section 1.3.4, and assuming that the spectral density of the error process f_ε, decays at a rate proportional to $\|\boldsymbol{\lambda}\|^{-\tau}$, $\tau > 2$, at high frequencies, it is easy to see that the covariance function of the ESP (5.43) converges to:

$$C_\eta(\mathbf{s}, \mathbf{s}') = \int FT(\mathbf{1}_\mathbf{s})(-\boldsymbol{\lambda}) FT(\mathbf{1}_{\mathbf{s}'})(\boldsymbol{\lambda}) f_\varepsilon(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \tag{5.44}$$

under a shrinking asymptotic framework. FT denotes the Fourier Transform operator, so (5.44) can be written as:

$$C_\eta(\mathbf{s}, \mathbf{s}') = \int e^{-i(\mathbf{s}-\mathbf{s}')^T\boldsymbol{\lambda}} A(\boldsymbol{\lambda}) A(-\boldsymbol{\lambda}) f_\varepsilon(\boldsymbol{\lambda}) d\boldsymbol{\lambda},$$

where

$$A(\boldsymbol{\lambda}) = \frac{1}{\lambda_1\lambda_2} + \frac{\pi(\delta(\lambda_1) + \delta(\lambda_2))}{i} + \pi^2\delta(\lambda_1)\delta(\lambda_2),$$

where $i^2 = (-1)$ and δ is the Dirac-delta function.

If the error process ε is non-stationary, then the non-stationary spectral density is required to decay at a rate proportional to $\|(\boldsymbol{\lambda}, \boldsymbol{\omega})\|^{-\tau}$, with $\tau > 4$, for high frequencies. Once again, under shrinking asymptotics we have that the covariance function of the ESP (5.43) converges to:

$$C_\eta(\mathbf{s}, \mathbf{s}') = \int\int FT(\mathbf{1}_\mathbf{s})(\boldsymbol{\lambda}) FT(\mathbf{1}_{\mathbf{s}'})(\boldsymbol{\omega}) f_\varepsilon(\boldsymbol{\lambda}, \boldsymbol{\omega}) d\boldsymbol{\lambda} d\boldsymbol{\omega}. \tag{5.45}$$

Both (5.44) and (5.45) depend on the unknown spectral density of the error process. We could adjust a parametric model by applying a goodness-of-fit tests, as those proposed in Chapter 4.

Note that the ESP is non-stationary, thus in order to estimate its spectral density, we could use the non-stationary periodogram introduced in (Fuentes (2002)):

$$I(\boldsymbol{\lambda}, \boldsymbol{\omega}) = \frac{\Delta^2}{N\Delta^2} \sum_{\mathbf{s}\in D} \eta_{N,\Delta}(\Delta\mathbf{s}) e^{-i\Delta\mathbf{s}^T\boldsymbol{\lambda}} \sum_{\mathbf{x}\in D} \eta_{N,\Delta}(\Delta\mathbf{x}) e^{-i\Delta\mathbf{x}^T\boldsymbol{\omega}}. \tag{5.46}$$

An idea for testing

$$\begin{aligned} H_0: &\quad \mu \in \mathcal{M}_\theta = \{\mu_\theta, \theta \in \Theta\}, \\ H_a: &\quad \mu \notin \mathcal{M}_\theta = \{\mu_\theta, \theta \in \Theta\}. \end{aligned}$$

is based on the comparison of the periodogram of the ESP (5.46) with its spatial spectral density, obtained as the Fourier Transform of (5.44) or (5.45). This comparison could be done using a Kullback-Leibler discrepancy or a Bootstrap approach.

A more complete research would be the study of the asymptotic behaviour of the ESP. The convergence of the finite dimensional distributions to a Gaussian limit can be obtained, but once again, proving the stochastic equicontinuity (tightness) may pose some challenges.

Software.

All the techniques presented along this manuscript (estimation methods, simulation algorithms and goodness-of-fit test procedures) have been implemented in Visual Fortran. Aware of the importance of allowing the scientific community for using these methods, our next objetive is to create a library for R statistical software.

Resumo en galego

A estatística espacial é unha das metodoloxías fundamentais para unha grande variedade de disciplinas, como a ecoloxía, a hidroloxía ou as ciencias medioambientais. En todos estes campos, os experimentos que se levan a cabo recollen datos que poden estar relacionados a unhas certas coordenadas no espazo. Neste proceso, os especialistas das distintas áreas, soen ter a idea intuitiva de que os valores dunha variable en localizacións próximas tenderán a ser máis semellantes que os tomados en localizacións alonxadas xeográficamente. Polo tanto, parece obvio que estes datos non poden ser tratados como independentes.

Algunhas referencias xerais no eido da estatística espacial son os libros de Cressie (1993), Chilés e Delfiner (1999) (para o caso xeoestatístico), Stein (1999) ou o máis recente de Schabenberger e Gotway (2005).

O deseño de modelos espaciais para representar a variabilidade dun proceso é un dos obxectivos principais en estatística espacial. A variabilidade nun modelo pode ser debida a dúas fontes. Por unha banda, a variabilidade a pequena escala (dependencia) e, pola outra, a variabilidade a grande escala (tendencia). Esta última ten sido modelada a través de modelos de regresión, como os modelos lineais xeneralizados ou os modelos aditivos. O noso interese centrarase na estrutura de dependencia do proceso.

Aínda dentro da estatística espacial, poderiamos facer a seguinte distinción entre os procesos estocásticos no espazo:

i) Procesos Xeoestatísticos. O proceso Z toma valores de xeito continuo sobre unha rexión $D \subset \mathbb{R}^2$. Por exemplo, supoñamos que nos interesa medir a concentración que existe dun certo metal pesado no solo ou no ambiente. As medicións da concentración de metal poderían tomarse en calqueira punto, posto que o proceso toma valores en calquera localización (véxase Cressie (1993) ou Chilés e Delfiner (1999)).

ii) Procesos Reticulares. O proceso Z toma valores nun conxunto finito de puntos no espazo, $D = \{s_1, \ldots, s_n\}$. Un exemplo: en estudos epidemiolóxicos, os datos de índice de mortandade por unha certa enfermidade en Galicia veñen dados por comarcas (as localizacións sobre as que toma valores o proceso son puntos asociados a cada comarca) (véxase, por exemplo, Cressie (1993), Parte II).

iii) Procesos Puntuais. O proceso Z toma valores nunha rexión $D \subset \mathbb{R}^2$, pero as posicións onde se atopan realizacións deste proceso Z distribúense de xeito aleatorio sobre D, sen que o investigador teña control sobre elas. Esta é a situación que se plantexa en estudos forestais (medicións en árbores: o investigador non ten control sobre onde tomar as medicións) (por exemplo, Stoyan *et al.* (1995) e Diggle (2003)).

O modelado da dependencia espacial ten unha especial importancia no contexto xeoestatístico. A xeoestatística refírese aos procesos espaciais continuos. Como xa se comentou, a medición de tales cantidades asociadas a procesos continuous, pódese facer en calquera localización no espazo. Pero as medidas non se toman en tódolos puntos, e a predición é un dos principais obxectivos da análise xeoestatística.

Neste contexto, as técnicas de predición difiren das clásicas porque involucran ao modelo espacial: a predición xeoestatística ten que conta a estrutura de dependencia do proceso. É por iso que se teñen invertido moitos esforzos na descrición do comportamento da estrutura de dependencia, en particular, baixo suposicións de estacionariedade. Con todo, non se ten prestado atención ao problema do contraste da bondade de axuste.

O obxectivo principal deste traballo é propoñer contrastes de bondade de axuste que permitan contrastar a validez dun certo modelo para explicar a estrutura de dependencia dun proceso espacial estacionario. Centrarémos a nosa investigación no dominio espectral, polo que a estrutura de dependencia será modelada a través da densidade espectral.

A continuación expoñemos un breve resumo de cada un dos capítulos que constitúen esta tese doutoral, facendo mención dos principais avances obtidos en cada un deles.

Capítulo 1. Estatística espacial e métodos espectrais. Neste capítulo facemos unha breve revisión das diferentes situacións nas que nos atopamos cando dispoñemos de datos con dependencia espacial (e.g. Cressie (1993) ou Chilés e Delfiner (1999)). Faremos unha revisión do caso xeoestatístico, centrándonos no problema de modelado da dependencia e interpolación kriging.

Para poder explicar a estrutura de dependencia e facer posible a inferencia, unha das hipóteses básicas que manexaremos é a de estacionariedade de segunda orde (estancionariedade débil ou homoxeneidade do proceso).

Definición. *Un proceso espacial Z dise debilmente estacionario se:*

$$E(Z(\mathbf{s})) = \mu(\mathbf{s}), \quad \forall \mathbf{s} \in D, \tag{5.47}$$

$$Cov(Z(\mathbf{s}+\mathbf{u}), Z(\mathbf{s})) = C(\mathbf{u}), \quad \forall \mathbf{s}, \mathbf{s}+\mathbf{u} \in D. \tag{5.48}$$

En (5.47), $\mu(\mathbf{s})$ denota a función de tendencia, que recolle a variabilidade a grande escala do proceso. A estacionariedade débil implica que a dependencia entre dúas observacións do proceso non depende das posicións nas que son tomadas, senón do vector diferenza entre as localizacións. Un caso máis sinxelo é aquel no que a dependencia é función só da distancia entre as posicións: isotropía.

Definición. *Un proceso espacial Z dise intrinsecamente estacionario se:*

$$E(Z(\mathbf{s}+\mathbf{u}) - Z(\mathbf{s})) = 0, \quad \forall \mathbf{s}, \mathbf{s}+\mathbf{u} \in D, \tag{5.49}$$

$$Var(Z(\mathbf{s}+\mathbf{u}) - Z(\mathbf{s})) = 2\gamma(\mathbf{u}), \quad \forall \mathbf{s}, \mathbf{s}+\mathbf{u} \in D. \tag{5.50}$$

A estacionariedade intrínseca supón que, para todo $\mathbf{u}$, o incremento $(Z(\mathbf{s}+\mathbf{u}) - Z(\mathbf{s}))$ é debilmente estacionario.

Tanto o covariograma C como o variograma 2γ describen a estructura de dependencia dos datos. Estas dúas funcións do proceso teñen interese por si mesmas, posto que modelizan a dependencia; pero ademais, son fundamentais cando o noso obxectivo é a predición do proceso en puntos onde non temos observacións. O variograma ou o covariograma interveñen dentro dos métodos de interpolación óptima de datos espaciais: kriging (Stein (1999)).

Existe unha ampla literatura sobre a estimación destas dúas funcións, de xeito paramétrico e unha alternativa máis recente a través de técnicas de estimación non paramétrica. Con todo, non imos considerar o covariograma ou o variograma para o modelado da dependencia, posto que non traballaremos no dominio dos datos, senón no dominio das frecuencias. Centrarémonos en procesos espaciais $\{Z(\mathbf{s}), \mathbf{s} \in D \subset \mathbb{R}^2\}$ que son debilmente estacionarios, posto que todo proceso deste tipo admite unha representación en forma de integral de Fourier-Stieltjes:

$$Z(\mathbf{s}) = \int_{\mathbb{R}^2} e^{i\mathbf{x}^T\boldsymbol{\omega}} \mathcal{Y}(d\boldsymbol{\omega}), \tag{5.51}$$

(véxase (Yaglom (1987))) onde $\mathcal{Y}$ é unha medida aleatoria ortogonal complexa. Esta expresión coñécese como a representación espectral de Z. Ademais, existe unha identidade entre a clase de funcións de covarianzas continuas en $\mathbb{R}^d$ e a clase de funcións definidas positivas en $\mathbb{R}^d$ (Teorema de Bochner). O Teorema de Khinchin establece que unha función real continua $C(\mathbf{u})$ definida sobre $\mathbb{R}^d$ é unha función de covarianzas (un covariograma) se e só se é a Transformada de Fourier Inversa dunha medida acotada, simétrica e positiva $F(d\boldsymbol{\omega})$. No caso $d = 2$:

$$C(\mathbf{u}) = \int_{\mathbb{R}^2} e^{i\mathbf{u}^T\boldsymbol{\omega}} F(d\boldsymbol{\omega}), \quad \text{con} \int_{\mathbb{R}^2} F(d\boldsymbol{\omega}) < \infty. \tag{5.52}$$

Ademais, se $C(\mathbf{u})$ decae suficientemente rápido para asegurar que $C \in L^1(\mathbb{R}^2)$, a medida F é a integral dunha función acotada e continua $f(\boldsymbol{\omega})$, que se coñece como a densidade espectral do proceso. Polo tanto, a densidade espectral pódese escribir como a Transformada de Fourier da covarianza, formando un par de Fourier:

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} C(\mathbf{u}) e^{-i\mathbf{u}^T\boldsymbol{\omega}} d\mathbf{u}, \quad \boldsymbol{\omega} \in \mathbb{R}^2, \tag{5.53}$$

$$C(\mathbf{u}) = \int_{\mathbb{R}^2} f(\boldsymbol{\omega}) e^{i\mathbf{u}^T\boldsymbol{\omega}} d\boldsymbol{\omega}. \tag{5.54}$$

No caso de que traballemos con procesos discretos, as integrais das fórmulas anteriores pasarían a expresarse en termos de sumatorios.

Na última sección deste capítulo incluímos unha sección na que se recollen os aspectos máis importantes da representación espectral de campos aleatorios (e.g. Grenander (1981), Yaglom and Yaglom (1987)), como base para os desenrolos teóricos posteriores.

Capítulo 2. Técnicas espectrais para o modelado da dependencia espacial. Introducimos neste capítulo o concepto de periodogram espacial, como estimador non paramétrico da densidade espectral. Tamén facemos unha breve revisión das técnicas de estimación para a densidade espectral, dende unha perspectiva paramétrica e non paramétrica. O capítulo compleméntase con algunhas consideracións sobre estimadores da densidade espectral derivados do periodograma. No apéndice deste capítulo inclúense as probas dos resultados obtidos. Por simplicidade, centrarémonos no caso de procesos discretos, aínda que os desenrolos correspondentes ao caso continuo pódense atopar no Capítulo 2.

O periodograma (tamén denominado densidade espectral mostral) é o estimador non paramétrico clásico da densidade espectral. Para un proceso espacial Z observado nunha grella regular $D = \{\mathbf{s} = (s_1, s_2) : s_1 = 0, \ldots, n_1 - 1, s_2 = 0, \ldots, n_2 - 1\}$ con $N = n_1 n_2$ puntos, o periodograma

espacial na frecuencia $\boldsymbol{\lambda}$ vén dado por:

$$I(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2 N} \left| \sum_{\mathbf{s}\in D} Z(\mathbf{s}) e^{-i\mathbf{s}^T\boldsymbol{\lambda}} \right|^2, \quad \boldsymbol{\lambda} \in \Pi^2 = [-\pi,\pi] \times [-\pi,\pi]. \tag{5.55}$$

O periodogram soe avaliarse no conxunto das frecuencias de Fourier bidimensionais (como xa explicamos na Sección 1.3.1) $\boldsymbol{\lambda}_{\mathbf{k}}^T = (\lambda_{k_1}, \lambda_{k_2})$:

$$\begin{aligned} \lambda_{k_1} &= \tfrac{2\pi k_1}{n_1}, \quad k_1 = 0, \pm 1, \ldots, \pm m_1, \quad \text{onde } m_1 = [(n_1-1)/2], \\ \lambda_{k_2} &= \tfrac{2\pi k_2}{n_2}, \quad k_2 = 0, \pm 1, \ldots, \pm m_2, \quad \text{onde } m_2 = [(n_2-1)/2]. \end{aligned}$$

O periodograma defínese en termos dos datos observados en (5.55), pero dado que a densidade espectral é a Transformada de Fourier da función de covarianzas, non resultaría extraño escribir o periodograma en termos dun estimador da covarianza. Se definimos as covarianzas mostrais como:

$$\hat{C}(\mathbf{u}) = \frac{1}{N} \sum_{\mathbf{s}\in D(\mathbf{u})} Z(\mathbf{u}) Z(\mathbf{s}+\mathbf{u}), \tag{5.56}$$

onde $D(\mathbf{u}) = \{\mathbf{s} \in D; \mathbf{s}+\mathbf{u} \in D\}$, o periodograma pódese escribir como:

$$I(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{u}\in\mathcal{U}} C(\mathbf{u}) e^{-i\mathbf{u}^T\boldsymbol{\lambda}}, \tag{5.57}$$

onde $\mathbf{u} \in \mathcal{U} = \{(u_1,u_2); 1-n_1 \le u_1 \le n_1-1, 1-n_2 \le u_2 \le n_2-1\}$.

O periodograma é un estimador asintóticamente insesgado da densidade espectral, pero non é consistente. A representación (5.57) utilizarase para construír estimadores consistentes da densidade espectral, a través da suavización das covarianzas.

Neste capítulo estudiamos o sesgo e a dependencia dun periodograma, no caso dun proceso discreto, cando se fai *tapering* nas observacións. Tamén se propón unha clase de estimadores consistentes do periodograma para procesos continuos. O problema destes estimadores, baseados nun suavizado das covarianzas, é que presentan efecto-fronteira. As implicacións da selección do núcleo de suavizado, o parámetro ventana e o espaciado entre os datos neste efecto-fronteira, estúdianse na Sección 2.4.

A pesar da súa falta de consistencia como estimador da densidade espectral, o periodograma (5.55) presenta unha formulación atractiva para a clase dos procesos lineais:

$$Z(\mathbf{s}) = \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} a_{jl} \varepsilon(s_1-j, s_2-l), \quad \sum_{l=-\infty}^{\infty} a_{jl}^2 < \infty \tag{5.58}$$

onde ε son variables aleatorias independentes e identicamente distribuídas $N(0, \sigma_\varepsilon^2)$. Esta representación non resulta excesivamente restritiva, dado que todo proceso Gaussiano estacionario pode ser representado deste xeito. Ademais, pode tamén intrepretarse como unha aproximación discreta aos procesos lineais continuous. O periodograma dunha realización deste proceso pódese escribir en termos do periodograma do proceso de innovación ε. En consecuencia:

$$I(\boldsymbol{\lambda}_\mathbf{k}) = f(\boldsymbol{\lambda}_\mathbf{k})V_\mathbf{k} + R_n(\boldsymbol{\lambda}_\mathbf{k}) \tag{5.59}$$

onde as $V_\mathbf{k}$'s son independentes e identicamente distribuídas $Exp(1)$ (véxase Brockwell e Davis (1991). Se aplicamos logaritmos en (5.59) teremos

$$Y_\mathbf{k} = m(\boldsymbol{\lambda}_\mathbf{k}) + z_\mathbf{k} + r_\mathbf{k} \tag{5.60}$$

onde $m = \log f$ e o termo residual

$$r_\mathbf{k} = \log\left[1 + \frac{R_n(\boldsymbol{\lambda}_\mathbf{k})}{f(\boldsymbol{\lambda}_\mathbf{k})V_\mathbf{k}}\right]. \tag{5.61}$$

As variables $z_\mathbf{k}$ son independentes e identicamente distribuídas $Gum(0, 1)$. Nos desenrolos posteriores sacaremos vantaxe das representacións (5.59) e (5.60) do periodograma espacial e do seu logaritmo, tanto para a proposta de métodos de simulación como para a construción de contrastes de bondade de axuste para a densidade espectral no espazo.

Capítulo 3. Simulación de estruturas de dependencia espacial. Neste capítulo, revisamos o método clásico de simulación espectral para procesos espaciais, o denominado Método Integral de Fourier (véxase Pardo-Igúzquiza e Chica-Olmo (1993) e Chilés e Delfiner (1999)) e propoñemos unha modificación do mesmo que presenta un mellor comportamento.

A nosa proposta está inspirada na representación do periodograma como resposta nun modelo de regresión multiplicativo (5.59), onde a densidade espectral aparece perturbada por unha variable exponencial de media unidade. Os algoritmos do Método Integral de Fourier e do Método Integral de Fourier Modificado pódense atopar na Sección 3.2. As modificacións propostas sobre o Método Integral de Fourier tamén abranguen o caso de xeneración de realizacións de procesos continuos.

A través dun estudo de simulación, amósase o bo funcionamente do Método Integral de Fourier Modificado, tanto no caso de procesos discretos como continuos. Tamén se pon te manifesto que a extensión directa de métodos de simulación no contexto das series temporais pode dar lugar a resultados pouco satisfactorios.

Capítulo 4. Contrastes de bondade de axuste para a densidade espectral espacial. Neste capítulo propóñense dúas técnicas de bondade de axuste para a densidade espectral. Neste eido, podemos plantexar un contraste de hipóteses que teña como hipótese nula que a densidade espectral dun proceso da forma (5.58) pertence a unha certa familia paramétrica:

$$\begin{array}{rrcl} H_0: & f & \in & \mathcal{F}_\theta = \{f_\theta; \quad \theta \in \Theta\}, \\ H_a: & f & \notin & \mathcal{F}_\theta = \{f_\theta; \quad \theta \in \Theta\}. \end{array} \tag{5.62}$$

Ou ben, considerando o logaritmo da densidade espectral, o problema pódese escribir como

$$\begin{array}{rrcl} \tilde{H}_0: & m & \in & \mathcal{M}_\theta = \{m_\theta; \quad \theta \in \Theta\}, \\ \tilde{H}_a: & m & \notin & \mathcal{M}_\theta = \{m_\theta; \quad \theta \in \Theta\}. \end{array} \tag{5.63}$$

Baixo algunhas condicións de regularidade nos coeficientes de (5.58), proponse o seguinte estatístico de contraste (de xeito similar a Paparoditis (2000), para series de tempo):

$$T_P = N|H|^{1/4} \int_{\Pi^2} \left(\frac{1}{N|H|^{1/2}} \sum_{\mathbf{k}} K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k})) \left(\frac{I(\boldsymbol{\lambda})}{f_{\hat{\theta}}(\boldsymbol{\lambda})} - 1 \right) \right)^2 d\boldsymbol{\lambda}, \tag{5.64}$$

onde $\sum_{\mathbf{k}}$ denota a suma sobre tódalas frecuencias de Fourier frequencies e $\hat{\theta}$ é o estimador de Whittle (véxase Sección 2.5). N denota o número de datos, K é un kernel bivariante e H é a matriz ventana. Os resultados obtidos sobre este estatístico son os seguintes:

Teorema 1. *Baixo as suposicións (1)-(4) da Sección 4.3.1 e baixo a hipótese nula* $H_0 : f_\theta \in \mathcal{F}_\theta$

$$T_P - \mu_H \to N(0, \tau^2) \quad \textit{en distribución},$$

onde μ_H *e* τ^2 *veñen dadas por:*

$$\mu_H = |H|^{-1/4} \int K^2(\mathbf{u}) d\mathbf{u}, \tag{5.65}$$

$$\tau^2 = \frac{1}{2\pi^2} \int_{2\Pi^2} \left(\int_{\Pi^2} K(\mathbf{u}) K(\mathbf{u} + \mathbf{v}) d\mathbf{u} \right)^2 d\mathbf{v}, \quad 2\Pi^2 = [-2\pi, 2\pi] \times [-2\pi, 2\pi]. \tag{5.66}$$

Teorema 2. *Consideremos o problema* $H_0 : f \in \mathcal{F}_\theta$ *vs.* $H_a : f \in \mathcal{F} - \mathcal{F}_\theta$. *Baixo as hipóteses (1)-(3) e (5) da Sección 4.3.1, cando* $n_1, n_2 \to \infty$*:*

$$N^{-1}|H|^{-1/4} T_P \to \int_{\Pi^2} \left(\frac{f(\boldsymbol{\lambda})}{f_{\theta^*}(\boldsymbol{\lambda})} - 1 \right)^2 d\boldsymbol{\lambda}$$

en probabilidade.

A segunda proposta para un contraste (similar a Fan and Zhang (2004) para o caso unidimensional) baséase na expresión (2.15) para o logaritmo do periodograma. O estatístico de contraste proposto é un estatístico de razón de verosimilitudes xeneralizado:

$$T_{LK} = \sum_{\mathbf{k}} \left[e^{Y_{\mathbf{k}} - m_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}})} + m_{\hat{\theta}}(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y_{\mathbf{k}} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}})} - \hat{m}_{LK}(\boldsymbol{\lambda}_{\mathbf{k}}) \right], \tag{5.67}$$

onde $\hat{m}_{LK}$ é un estimador non-paramétrico da log-densidade espectral (Sección 2.5.1).

Teorema 3. *Baixo as hipóteses (1)-(4) da Sección 4.3, se* $N^{(\zeta-1)/\zeta}|H|^{1/2} \geq c \log^{\delta} N$, *para unha constante* c *e unha* $\delta > (\zeta-1)/(\zeta-2))$, $\zeta > 2$ *supoñendo que se estamos baixo* H_0,

$$\sigma^{-1}(T_{LK} - \mu_H + b_H) \to N(0,1),$$

onde μ_H, b_H *e* σ^2 *son:*

$$\mu_H = \frac{4\pi^2}{|H|^{1/2}} \left(K(\mathbf{0}) - \frac{1}{2} \int K^2(\mathbf{s}) d\mathbf{s} \right), \tag{5.68}$$

$$b_H = \sum_{\mathbf{k}} \frac{-|H|^2}{8 f_{\theta}(\boldsymbol{\lambda}_{\mathbf{k}})} \int \int \mathbf{s}^T H_{m_\theta}(\boldsymbol{\lambda}_{\mathbf{k}}) \mathbf{s} \cdot (\mathbf{s}+\mathbf{u})^T H_{m_\theta}(\boldsymbol{\lambda}_{\mathbf{k}})(\mathbf{s}+\mathbf{u}) K(\mathbf{s}) K(\mathbf{s}+\mathbf{u}) d\mathbf{s} d\mathbf{u}, \tag{5.69}$$

$$\sigma^2 = \frac{4\pi^2}{|H|^{1/2}} \int \left(2K(\mathbf{s}) - K * K(\mathbf{s})\right)^2 d\mathbf{s}, \tag{5.70}$$

e $H_{m_\theta}(\boldsymbol{\lambda}_{\mathbf{k}})$ *é a matriz Hessiana de* m_θ.

Dado que a velocidade de converxencia das distribucións de T_P e T_{LK} ao seu límite Gaussiano é lenta, propoñemos unha forma alternativa de estimar as distribucións dos estatísticos de contraste, baixo a hipótese nula, a través dunha aproximación Monte Carlo.

O comportamento dos dous tests ilústrase mediante un estudo de simulación e a aplicación a datos reais. No apéndice deste capítulo inclúense as probas dos resultados obtidos.

Capítulo 5. Comparación de estruturas de dependencia. Este último capítulo dedícase a un test para comparar dúas ou máis densidades espectrais. De xeito equivalente, estase a propoñer un test para ver se a estrutura de dependencia dun conxunto de observacións varía ao longo do tempo. Este test está baseado nunha distancia L_2, como en Vilar-Fernández e González-Manteiga (2004), para a comparación de curvas de regresión.

O noso obxectivo é contrastar se as densidades espectrais de varias observacións dun mesmo proceso (ou de procesos distintos) son a mesma. De xeito equivalente, en termos da log-densidade espectral, o contraste pode formularse como:

$$\begin{aligned} H_0 &: \quad m_1 = \ldots = m_L, \\ H_a &: \quad m_l \neq m_j, \text{ para algún } l \neq j. \end{aligned} \tag{5.71}$$

Neste contexto, a comparación pódese facer a través de estimadores non-paramétricos da log-densidade espectral, da forma:

$$\hat{m}_l(\boldsymbol{\lambda}_\mathbf{k}) = \sum_{\mathbf{i}} W_\mathbf{i}^l(\boldsymbol{\lambda}_\mathbf{k}) Y_\mathbf{i}^l, \tag{5.72}$$

onde os pesos $W_\mathbf{i}^l$ pódense definir como pesos de Gasser-Muller, de Priestley-Chao ou locais-lineais. No noso contexto, os puntos do deseño son fixos (son as frecuencias de Fourier), polo que estes pesos son asintóticamente equivalentes. O estatístico de contraste proposto será:

$$\hat{Q} = \frac{(2\pi)^2}{N} \sum_{l=2}^{L} \sum_{j=1}^{l-1} \sum_{\mathbf{k}} \left(\widehat{m^l}(\boldsymbol{\lambda}_\mathbf{k}) - \widehat{m^j}(\boldsymbol{\lambda}_\mathbf{k}) \right)^2 \omega(\boldsymbol{\lambda}_\mathbf{k}). \tag{5.73}$$

No caso máis sinxelo de comparar dúas densidades espectrais, obtemos os seguintes resultados:

Teorema 4. *Supoñamos que se satisfacen as condicións (A1)-(A5) da Sección 5.2.1. Entón, baixo a hipótese nula $H_0 : m_1 = m_2$ temos que:*

$$\sqrt{N|H|^{1/2}} \left(\hat{Q} - \frac{1}{12N|H|^{1/2}} C_K I_\omega \right) \to N(0, \sigma_Q^2),$$

en distribución, con

$$C_K = \int K(\mathbf{u}) d\mathbf{u}, \quad I_\omega = \int \omega(\mathbf{v}) d\mathbf{v} \quad \text{e a varianza asintótica é}$$

$$\sigma_Q^2 = \frac{1}{18} \int (K * K)^2(\mathbf{u}) d\mathbf{u} \int \omega(\mathbf{v}) d\mathbf{v},$$

onde $*$ denota o operador de convolución.

Tamén neste contexto de comparar dúas estrutras de dependencia, consideremos que a hipótese nula é false e supoñamos que:

$$m_1(\boldsymbol{\lambda}) - m_2(\boldsymbol{\lambda}) = C_N p(\boldsymbol{\lambda}), \tag{5.74}$$

onde $p(\boldsymbol{\lambda})$ é unha función determinística non nula.

Teorema 5. *Supoñamos que se satisfacen as condicións (A1)-(A5) da Sección 5.2.1. Entón, se (5.74) é certa, temos que:*

$$\sqrt{N|H|^{1/2}} \left(\hat{Q} - \frac{1}{12N|H|^{1/2}} C_K I_\omega \right) \to N \left(\int p^2(\mathbf{v}) \omega(\mathbf{v}) d\mathbf{v}, \sigma_Q^2 \right),$$

en distribución, onde C_K, I_ω e σ_Q^2 son como no teorema anterior, e con $C_N^2 = (N^2|H|^{1/2})^{-1/2}$.

Para a aplicación deste contraste na práctica, propóense dous algoritmos. O primeiro deles utilízase cando temos algunha infomación sobre a forma paramétrica das densidades espectrais, e

o segundo cando non dispoñemos de información ningunha a priori. Tamén incluímos algúns resultados de simulación e a aplicación a datos reais. Os desenrolos teóricos achéganse nun apéndice.

Agradecementos.
Este traballo foi financiado polo Ministerio de Educación e Ciencia e FEDER, a través do proxectos BFM2002-03213, MTM2005-0020 e a bolsa predoutoral de Formación de Persoal Investigador BES2003-0581.

Bibliography

Abramowitz, M. and Stegun, I. (1965). *Handbook of MAthematical Functions*. New York: Dover.

Alonso, F. J.; Angulo, J. M.; and Bueso, M. C. (1996). Estimation and smoothing from incomplete data for a class of lattice processes. *Comm. Statist. Theory Methods*, volume 25, no. 3:pp. 667–681.

Angulo, J. M. and Ruiz-Medina, M. D. (1999). Multi-resolution approximation to the stochastic inverse problem. *Adv. in Appl. Probab.*, volume 31, no. 4:pp. 1039–1057.

Armstrong, M. and Delfiner, P. (1980). Towards a more robust variogram: A case study on coal. *Centre de Géostatisque, Fontainebleau*, volume 671.

Banerjee, S.; Carlin, B. P.; and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall.

Barry, R. P. and Ver Hoef, J. M. (1996). Blackbox kriging: spatial prediction without specifying variogram models. *J. Agric. Biol. Environ. Stat.*, volume 1, no. 3:pp. 297–322.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B*, volume 36:pp. 192–236.

Borgman, L.; Taheri, M.; and Hagan, R. (1984). Three-dimensional frequency domain simulations of geological variables. In G. Verly; A. Journel; and A. Marechal (Eds.), *Geostatistics for natural resources characterization*. The Netherlands.

Box, G. and Muller, M. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, volume 1:pp. 610–611.

Briggs, W. and Henson, V. (1995). *The DFT. An Owner's Manual for the Discrete Fourier Transform*. Siam.

Brilinger, D. R. (1974). Fourier analysis of stationary processes. In *Proceedings of the IEEE*, 12, pp. 1628–1643.

Brillinger, D. R. (1970). The frequency analysis of relations between stationary spatial series. In R. Pyke (Ed.), *Proceedings, 12th Biennial Seminar on Time Series, Stochastic Processes, Convexity, Combinatories. Canadian Mahtematical Congress.* Montreal.

Brillinger, D. R. (1981). *Time series*. Oakland, Calif.: Holden-Day Inc., second edition.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*. Springer Series in Statistics. New York: Springer-Verlag, second edition.

Carroll, R. J.; Fan, J.; Gijbels, I.; and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.*, volume 92, no. 438:pp. 477–489.

Chilès, J.-P. and Delfiner, P. (1999). *Geostatistics*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. New York: John Wiley & Sons Inc. Modeling spatial uncertainty, A Wiley-Interscience Publication.

Cooley, J. and Tukey, J. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computations*, volume 19:pp. 297–301.

Cressie, N. and Hawkins, D. M. (1980). Robust estimation of the variogram. I. *J. Internat. Assoc. Math. Geol.*, volume 12, no. 2:pp. 115–125.

Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley & Sons Inc.

Crujeiras, R. M. and Fernández-Casal, R. (2006). On the spectral simulation of spatial dependence structures. *Technical Report. Department of Statistics and O.R. University of Santiago de Compostela. http://eio.usc.es/pub/reports.html*.

Crujeiras, R. M.; Fernández-Casal, R.; and González-Manteiga, W. (2006a). Comparing spatial dependence structures. *Technical Report. Department of Statistics and O.R. University of Santiago de Compostela. http://eio.usc.es/pub/reports.html*.

Crujeiras, R. M.; Fernández-Casal, R.; and González-Manteiga, W. (2006b). Goodness-of-fit tests for the spatial spectral density. *Technical Report. Department of Statistics and O.R. University of Santiago de Compostela. http://eio.usc.es/pub/reports.html*.

Crujeiras, R. M.; Fernández-Casal, R.; and González-Manteiga, W. (2006c). An l_2 test for comparing spatial spectral densities. *Technical Report. Department of Statistics and O.R. University of Santiago de Compostela. http://eio.usc.es/pub/reports.html*.

Dahlhaus, R. and Janas, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.*, volume 24, no. 5:pp. 1934–1963.

Dahlhaus, R. and Künsch, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, volume 74, no. 4:pp. 877–882.

Dahlhaus, R. and Wefelmeyer, W. (1996). Asimptotically optimal estimation in misspecified times series models. *Annals of Statistics*.

Dautray, R. and Lions, J. (1985). *Mathematical Analysis and Numerical Methods for Science and Technology*. Springer-Verlag.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields*, volume 75, no. 2:pp. 261–277.

Delgado, M.; Hidalgo, J.; and Velasco, C. (2005). Distribucin free goodness-of-fit tests for linear processes. *Ann. Statist.*, volume 33, no. 6.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variances. *Annals of Statistic*.

Dette, H. and Neumeyer, N. (2001). Nonparametric analysis of covariance. *Annals of Statistic*, volume 29, no. 5:pp. 1361–1400.

Diblasi, A. and Bowman, A. W. (2001). On the use of the variogram in checking for independence in spatial data. *Biometrics*, volume 57, no. 1:pp. 211–218.

Diebolt, J. and Zuber, J. (2001). On testing the goodness-of-fit of nonlinear heteroscedastic regression models. *Comm. Statist. Simulation Comput.*, volume 30, no. 1:pp. 195–216.

Dietrich, C. R. and Newsam, G. N. (1997). Fast and exact simulation of stationary gaussian processes through the circulant embedding of the covariance matrix. *Siam Journal on Scientific Computing*, volume 18, no. 4:pp. 1088–1107.

Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns*. Oxford University Press.

Dijkstra, E. W. (1976). *A Discipline of Programming*. Prentice Hall.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B*, volume 57, no. 2:pp. 371–394.

Fan, J.; Zhang, C.; and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.*, volume 29, no. 1:pp. 153–193.

Fan, J. and Zhang, W. (2004). Generalised likelihood ratio tests for spectral density. *Biometrika*, volume 91, no. 1:pp. 195–209.

Fernández, J.; Rey, A.; and Carballeira, A. (2000). An extended study of heavy metal deposition in galicia based on moss analysis. *Science of Total Environment*, volume 254:pp. 31–44.

Fernández-Casal, R. (2003). *Geoestadstica Espacio-Temporal. Modelos flexibles de variogramas anisotrpicos no separables*. PhD Dissertation. Departament of Statistics and Operation Research. University of Santiago de Compostela.

Fernández-Casal, R.; González-Manteiga, W.; and Febrero-Bande, M. (2003). Flexible spatio-temporal stationary variogram models. *Stat. Comput.*, volume 13, no. 2:pp. 127–136.

Folland, G. (1993). *Fourier Analisis and its applications*. Wadsworth.

Franke, J. and Härdle, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.*, volume 20, no. 1:pp. 121–145.

Fuentes, M. (2001). Fixed-domain asymptotics for variograms using subsampling. *Math. Geol.*, volume 33, no. 6:pp. 679–691.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, volume 89, no. 1:pp. 197–210.

Fuentes, M. (2006a). Approximate likelihood for large irregularly spaced spatial data. *J. Amer. Statist. Assoc.*, volume 101, no. 476.

Fuentes, M. (2006b). Testing for separability of spatial-temporal covariance functions. *J. Statisti. Plan. Inference*, volume 136, no. 2:pp. 447–466.

García-Soidán, P.; Febrero-Bande, M.; and González-Manteiga, W. (2003). Local linear regression estimation of the semivariogram. *Statistics and Probability Letters*, volume 64:pp. 169–179.

García-Soidán, P.; Febrero-Bande, M.; and González-Manteiga, W. (2004). Nonparametric kernel estimation of an isotropic variogram. *Journal of Stat. Planning Inference*, volume 121:pp. 65–92.

Genton, M. G. (1998). Highly robust variogram estimation. *Math. Geol.*, volume 30, no. 2:pp. 213–221.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.

González Manteiga, W. and Cao, R. (1993). Testing the hypothesis of a general linear model using nonparametric regression estimation. *Test*, volume 2, no. 1-2:pp. 161–188.

Grenander, U. (1981). *Abstract Inference*. Wiley.

Guan, Y.; Sherman, M.; and Calvin, J. A. (2006). Assessing isotropy for spatial point processes. *Biometrics*, volume 62, no. 1:pp. 119–125, 316.

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, volume 69, no. 1:pp. 95–105.

Hannan, E. J. (1970). *Multiple time series*. John Wiley and Sons, Inc., New York-London-Sydney.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, volume 21, no. 4:pp. 1926–1947. ISSN 0090-5364.

Hart, J. D. (1997). *Nonparametric smoothing and lack-of-fit tests*. Springer Series in Statistics. New York: Springer-Verlag.

Holland, D.; Oliveira, V.; Cox, L.; and Smith, R. (2000). Estimation of regional trends in sulfur dioxide over the eastern united states. *Environmetrics*, volume 11, no. 4:pp. 373–393.

King, E.; Hart, J.; and Wherly, T. (1991). Testing the equality of two regression curves using linear smoothers. *Statistics & Probability Letters*, volume 12:pp. 239–247.

Kooperberg, C.; Stone, C. J.; and Truong, Y. K. (1995). Rate of convergence for logspline spectral density estimation. *J. Time Ser. Anal.*, volume 16, no. 4:pp. 389–401.

Kreiss, J.-P. and Paparoditis, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *Ann. Statist.*, volume 31, no. 6:pp. 1923–1955.

Lahiri, S. (1999). Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method. *Probab. Theory Related Fields*, volume 114:pp. 55–84.

Lawson, A. B. (2006). *Statistical methods in spatial epidemiology*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd., second edition.

Lefohn, A. and Shadwick, D. (1991). Ozone, sulfur dioxide, and nitrogen dioxide trends at rural sites located in the united states. *Atmospheric Environment*, volume 25, no. 2:pp. 491–501.

Lele, S. (1995). Inner product matrices, Kriging, and nonparametric estimation of variogram. *Math. Geol.*, volume 27, no. 5:pp. 673–692.

Li, H.; Caldery, C.; and Cressie, N. (2005). Beyond moran's i: Testing for spatial dependence based on the sar model. *Department of Statistics. Preprint No. 763. The Ohio State University.*

Maglione, D. and Diblasi, A. (2004). Exploring a valid model for the variogram of an isotropic spatial process. *Stoch. Envir. Res. and Risk Ass.*, volume 18:pp. 366–376.

Martin, R. (2000). Exact gaussian maximum likelihood and simulation for regularly-spaced observations with gaussian correlations. *Biometrika*, volume 87, no. 3:pp. 727–730.

Martin, R. J. (1979). A subclass of lattice processes applied to a problem in planar sampling. *Biometrika*, volume 66, no. 2:pp. 209–217.

Mateu, J. and Ribeiro, P. (1999). Geostatistical data versus point process data: analysis of second order characteristics. *Quantitative Geology and Geostatistics*, volume 10:pp. 213–224.

Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Appl. Probability*, volume 5:pp. 439–468.

Moura, J. and Balram, N. (1992). Recursive structure of noncausal gauss-markov random fields. *IEEE Transactions on Information Theore*, volume 38, no. 2:pp. 334–354.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. New York: John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.

Müller, H.-G. (1988). *Nonparametric regression analysis of longitudinal data*, volume 46 of *Lecture Notes in Statistics*. Berlin: Springer-Verlag.

Neuman, S. and Jacobson, E. (1984). Analysis of nonintrisical spatial variability by residual kriging with application to regional groundwater levels. *Mathematical Geology*, volume 16:pp. 499–521.

Paparoditis, E. (2000). Spectral density based goodness-of-fit tests for time series models. *Scand. J. Statist.*, volume 27, no. 1:pp. 143–176.

Paparoditis, E. and Politis, D. N. (1999). The local bootstrap for periodogram statistics. *J. Time Ser. Anal.*, volume 20, no. 2:pp. 193–222.

Pardo-Igúzquiza, E. and Chica-Olmo, M. (1993). The fourier integral method: An efficient spectral method for simulation of random fields. *Mathematical Geology*, volume 25, no. 2:pp. 177–217.

Pentinnen, A.; Stoyan, D.; and Hentonnen, H. (1992). Marked point processes in forest statistics. *Forest Science*.

Porcu, E.; Crujeiras, R. M.; J., M.; and González-Manteiga, W. (2005). Spatial and spatio-temporal dependence of the periodogram for regularly spaced data. *Technical Report. Department of Mathematics Universitat Jaume I*.

Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; and Vetterling, W. T. (1989). *Numerical Recipes, Fortran version*. Cambridge University Press.

Priestley, M. B. (1981). *Spectral analysis and time series. Vol. 1.* London: Academic Press Inc.

Robinson, P. (1976). *System identification advances and case studies*, chapter Some problems in the identification and estimation of continuous time systems from discrete time series, pp. 407–439. Academic Press, INC.

Robinson, P. (1983). *Review of various approaches to power spectrum estimation*, volume 3, pp. 343–368. Elsevier Science.

Robinson, P. (2006). Nonparametric spectrum estimation for spatial data. *To appear in Journal of Statistical Planning and Inference.*

Robinson, P. and Vidal Sanz, J. (2006). Modified whittle estimation of multilateral models on a lattice. *J. Multivariate Anal.*, volume 97:pp. 1090–1120.

Ross, S. M. (1997). *Simulation*. London: Academic Press Inc., second edition.

Rue, H. (2001). Fast sampling of gaussian markov random fields. *J. R. Statist. Soc. B*, volume 64, no. 2:pp. 325–338.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, volume 22, no. 3:pp. 1346–1370.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical methods for spatial data analysis*. Texts in Statistical Science Series. Chapman & Hall.

Shao, J. (2003). *Mathematical statistics*. Springer Texts in Statistics. New York: Springer-Verlag, second edition.

Shapiro, A. and Botha, J. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis*, volume 11:pp. 87–96.

Shinozuka, M. (1971). Simulation of multivariate and multidimensional random processes. *The Journal of the Acoustical Society of America*, volume 49, no. 1:pp. 357–368.

Stein, M. L. (1995). Fixed domain asymptotics for spatial periodograms. *J. Am. Statist. Assoc.*, volume 90:pp. 1277–88.

Stein, M. L. (1999). *Interpolation of spatial data*. Springer Series in Statistics. New York: Springer-Verlag.

Stoyan, D.; Kendall, W.; and Mecke, J. (1995). *Stochastic Geometry and its Applications.* Wiley Series in Probability and Statistics: Probability and Statistics. Chichester: John Wiley & Sons Inc., second edition.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.*, volume 25, no. 2:pp. 613–641.

Stute, W.; González Manteiga, W.; and Presedo Quindimil, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.*, volume 93, no. 441:pp. 141–149.

Tawaga, H. (1997). World-wide distribution of evergreen lucidophyll oak-laurel forest. *Tropics.*

Vilar-Fernández, J. and González-Manteiga, W. (2004). Nonparametric comparison of curves with dependent errors. *Statistics*, volume 38, no. 2:pp. 81–99.

Walden, O. and Stoyan, D. (1996). On variograms in point process statistics. *Biometrical Journal*, volume 38, no. 8:pp. 895–905.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, volume 41:pp. 434–449.

Yaglom, A. M. (1987). *Correlation theory of stationary and related random functions. Vol. II.* Springer Series in Statistics. New York: Springer-Verlag.

Yaglom, A. M. and Yaglom, I. M. (1987). *Challenging mathematical problems with elementary solutions. Vol. I.* New York: Dover Publications Inc.

Yao, T. (1998). Conditional spectral simulation with phase identification. *Mathematical Geology*, volume 30, no. 3:pp. 285–308.

Yao, T. (2004). Reproduction of the mean, variance and variogram model in spectral simulation. *Mathematical Geology*, volume 36, no. 4:pp. 487–506.

Young, L. and Young, J. (1998). *Statistical Ecology.* Kluwer Academic Press.

Zhang, C. and Dette, H. (2004). A power comparison between nonparametric regression tests. *Statist. Probab. Lett.*, volume 66, no. 3:pp. 289–301.

Zheng, J. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics*, volume 75, no. 2:pp. 263–289.

Zhu, J.; Lahiri, S.; and Cressie, N. (2002). Asymptotic inference for spatial cdfs over time. *Statistica Sinica*, volume 12:pp. 843–861.