Electronic Journal of Statistics
Vol. 0 (0000)
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

Asymptotic properties of parameter estimates for random fields with tapered data

Huda Mohammed Alomari

Department of Mathematics and Statistics, La Trobe University, Melbourne, Australia and Department of Mathematics, Faculty of Sciences and Arts, Al-Makhwah, University of Al-Baha, Al-Baha, The Kingdom of Saudi Arabia
e-mail: halomari@students.latrobe.edu.au

M. Pilar Frías

Department of Statistics and and Operation Research, University of Jaén
e-mail: mpfrias@ujaen.es

Nikolai N. Leonenko

Cardiff School of Mathematics, University of Cardiff, United Kingdom
e-mail: LeonenkoN@cardiff.ac.uk

María D. Ruiz-Medina

Department of Statistics and Operation Research, University of Granada, Spain
e-mail: mruiz@ugr.es

Lyudmyla Sakhno

Department of Probability Theory, Statistics and Actuarial Mathematics, Taras Shevchenko National University of Kyiv, Ukraine
e-mail: lms@univ.kiev.ua

Antoni Torres

Department of Statistics and Operation Research, University of Granada, Spain
e-mail: atisignes@gmail.com

Abstract: In this paper we present novel results on the asymptotic behavior of the so-called Ibragimov minimum contrast estimates. The case of tapered data for various models of Gaussian random fields is investigated.

MSC 2010 subject classifications: Primary 62F12, 62M30; secondary 60G60.
Keywords and phrases: Gaussian random field, Ibragimov minimum contrast, tapered data, consistency and asymptotic normality.

Contents

1. Introduction

Parameter estimation of stationary stochastic processes and random fields constitutes a central topic in statistical inference from temporally or spatially correlated data. The maximum likelihood estimation method (MLE) is one of the most popular estimation tools. The problem of parameter estimation of fractionally integrated processes with seasonal components was addressed in [37]. To estimate the fractional parameters, they propose several log-periodogram regression estimators with different bandwidths selected around or between the seasonal frequencies. The same methodology was used in [34] for fractionally differenced autoregressive-moving average processes in the stationary time series context. Several contributions have also been made for the MLE of long memory spatial processes (see, for example, [8]). For two-dimensional spatial data the paper [12] introduced a spatial unilateral first-order autoregressive moving average (ARMA) model. To implement MLE they provided a proper treatment to border cell values with a substantial effect in the estimation of the parameters. Gaussian maximum likelihood estimation, in the context of ARMA models, was applied in [44].

Linear and non-linear functionals of the periodogram play a key role in the design of Minimum Contrast Estimation (MCE) techniques (see, for example, [42]). Whittle estimation procedure was treated quite extensively in the literature (see

[27, 28, 36], for the case of random fields). Special attention has been paid to minimum contrast parameter estimation of fractional, fractal and long-range dependence stochastic models (see, for example, [18]). For instance, [43] applied the minimum contrast parameter estimation to approximate the drift parameter of the Ornstein-Uhlenbeck process, when the corresponding stochastic differential equation is driven by the fractional Brownian motion. Consistency and asymptotic normality of the Whittle maximum likelihood estimator for stationary seasonal autoregressive fractionally integrated moving-average (SARFIMA) processes was proved in [26]. Maximization of the Whittle likelihood has also been considered in the papers [17, 16, 32]. A continuous version of the Whittle contrast functional, supplied with a specific weight function, was formulated in [32], where the problem of parameter estimation of continuous stationary processes was addressed, deriving the consistency and the asymptotic normality of the formulated MCE. Modified Whittle estimation of multilateral models on a lattice is considered in [39]. Estimation of the spectral density for aggregated possibly strong-dependent Gaussian random fields, from an expansion in terms of orthogonal Gegenbauer polynomials, was studied in [33]. These fields are constructed from accumulation of i.i.d. short memory fields, via an unknown mixing density (to be estimated).

An alternative to Whittle family of linear functionals was proposed in [30] (for generalizations, see also [2, 3, 4, 5, 6]). In particular, [2] derived consistency and asymptotic normality of a class of MCEs based on Ibragimov functional for fractional Riesz-Bessel motion (see [1]). This functional was considered in [22] for MCE of long-range dependence spatial time series, constructed from fractional difference operators associated with Gegenbauer polynomials. Consistency and asymptotic normality results were derived as well. Minimum contrast estimators, based on the nonlinear objective functions of the periodogram, have been studied in [35] and [42].

In this paper we study Ibragimov minimum contrast estimators based on tapered data. The case of Gaussian stationary random fields on $\mathbb{Z}^d$ is considered.

Note that benefits of tapering data have been widely reported in the literature. It is well known that tapering reduces leakage effects, especially when spectral densities contain high peaks. Even more importantly, the use of tapers leads to the bias reduction, which is a key issue in Spatial Statistics. Namely, tapers can help to fight the so-called "edge effects"; see, e.g., [20, 28].

The main contributions of the paper are two-fold:

1) We state the theorems on consistency and asymptotic normality for Ibragimov minimum contrast estimators for long-range dependent random fields in such a form which was not presented in the previous papers devoted to Ibragimov functional (see, e.g. [2, 3, 4]), namely, in the present paper: (i) conditions are given in the form prescribing behavior of the spectral density and its derivatives at the point of singularity, and (ii) exact conditions on tapers are given. This make the results very convenient to apply for particular models, the form of conditions gives the key to construct the weight function, which is incorporated into the Ibragimov functional in order to compensate singularities of the spectral density.

2) For the case of long-range dependent Gaussian random fields, we state the central limit theorem for spectral functionals (or quadratic forms) based on the tapered data, which is of independent importance and also serves as the main tool to derive the asymptotic properties of estimators in the present paper.

The numerical part contributes to investigations on convergence rates of minimum contrast estimators. It also demonstrates that the obtained theoretical results can be further extended to other classes of fractional models.

For convenience of a reader, the statements of the main results are accompanied by discussion to clarify the conditions used and also some digressions intended to comment the related results existing in the literature.

The paper is organized as follows.

In Section 2, we first explain the general methodology of minimum contrast estimation, and introduce Ibragimov contrast functional based on tapered data. Then, in subsections 2.4–2.5 we state our main results on consistency and asymptotic normality of the corresponding minimum contrast estimators. The conditions cover the case of long-range dependent random fields, with a particular model of spectral densities with singularities which factorize. We also formulate exact conditions on tapers, leading to an asymptotic normality result, with standard normalizing factor, for dimensions $d = 1, 2, 3$.

In Section 3 we revise the results concerning the bias control in minimum contrast estimation based on the Whittle functional: results in [27, 28] and in [36], for the cases of for short- and long-range dependent Gaussian random fields correspondingly.

Section 4 contains the proofs of the results from Section 2 on asymptotic properties of our minimum contrast estimators.

The proofs are based on the central limit theorem for spectral functionals $J_T(\phi) = \int_{\mathbb{T}} \phi(\boldsymbol{\lambda}) I_T^h(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}$ (with $I_T^h(\boldsymbol{\lambda})$ being the periodogram based on tapered data), which we state and prove in Appendix A. This theorem (Theorem A.2) provides an extension of the classical result in [24] to the case of fields and tapered data. Beyond its application here for derivation of asymptotic properties of MCE, it is of separate interest itself.

Since the main part of the paper is devoted to the study of Ibragimov minimum contrast estimators, we placed this central limit theorem with its derivation and related observations in Appendix A, which can be considered, to some extend, as a self-contained part of the paper. With this structure of the paper, the reader can choose to focus firstly on the statistical estimation results and their applications and then consider in detail the central limit theorem for spectral functionals based on tapered data.

Section 5 presents simulation studies to demonstrate performance of the estimations technique developed in the paper. Namely, in the simulation study undertaken, we illustrate consistency and asymptotic normality of the plug-in MCE considered, based on tapered periodogram, in the context of spatial Gaussian fractional autoregressive processes, and spatial Gegenbauer random fields.

2. Minimum contrast estimators based on the Ibragimov contrast function and tapered data

2.1. General definition of minimum contrast estimators

We introduce the definition of minimum contrast estimators following [28], pp. 119–127, where these estimators have been studied for some classes of discrete-time random fields.

Let a random field $Y(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{Z}^d$, be observed on a sequence L_T of increasing finite domains. We will suppose that L_T is a hypercube: $L_T = [-T,T]^d = \{\boldsymbol{t} \in \mathbb{Z}^d : -T \leq t_i \leq T,\ i = 1, ..., d\}$.

Consider a parametric statistical model with a family of distributions $P_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta$, where Θ is a compact subset of $\mathbb{R}^q$, and the true parameter value $\boldsymbol{\theta}_0 \in int\,\Theta$, the interior of Θ. Denote $P_0 = P_{\boldsymbol{\theta}_0}$.

We define:

1) a nonrandom real-valued function $K(\boldsymbol{\theta}_0; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, to be called a *contrast function*, such that $K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) \geq 0$ and it has a unique minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

2) a *contrast field* for a contrast function $K(\boldsymbol{\theta}_0; \boldsymbol{\theta})$, which is a random field $U_T(\boldsymbol{\theta})$, $T \in \mathbb{Z}$, $\boldsymbol{\theta} \in \Theta$, related to observations $Y(t), t \in L_T$, and such that the following relation holds:

$$U_T(\boldsymbol{\theta}) - U_T(\boldsymbol{\theta}_0) \to K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \Theta \tag{2.1}$$

in P_0-probability, as $T \to \infty$.

Then the *minimum contrast estimator* $\hat{\boldsymbol{\theta}}_T$ is defined as a minimum point of the functional $U_T(\boldsymbol{\theta})$, that is,

$$\hat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta} \in \Theta} U_T(\boldsymbol{\theta}).$$

Note that the usual way to realize the above definition in practice is to construct, basing on observations, a function $U_T(\boldsymbol{\theta})$ which converges in P_0-probability to some function $U(\boldsymbol{\theta})$ such that $U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}_{\mathbf{0}}) = K(\boldsymbol{\theta}_0; \boldsymbol{\theta})$.

In this paper we apply minimum contrast technique for parameter estimation in the spectral domain. We consider a measurable stationary zero-mean real-valued Gaussian random field $Y(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{Z}^d$. We suppose that the parametric form of its spectral density is known: $f(\boldsymbol{\lambda}) = f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, $\boldsymbol{\lambda} \in \mathbb{T} = (-\pi, \pi]^d$, $\boldsymbol{\theta} \in \Theta$, with $\Theta \subset \mathbb{R}^q$, $q \geq 1$, being a compact, and we are interested in estimation of θ_0, the true parameter value.

In the present paper we will study the so-called Ibragimov minimum contrast estimators, where $K(\boldsymbol{\theta}_0; \boldsymbol{\theta})$ has the form (2.7), the functions $U_T(\boldsymbol{\theta})$ and $U(\boldsymbol{\theta})$ are defined by (2.6) and (2.8) correspondingly (see Section 2.3 below). Another example of minimum contrast estimatiors is provided by the well known Whittle estimators, for which we present the expressions for contrast field and contrast function in Section 3. We refer to [42] for some other examples of minimum contrast estimators and also to [28] where many useful general results for minimum contrast method can be found.

2.2. Data tapers and tapered periodograms

We will base our analysis on tapered data. The use of tapers leads to the bias reduction, which is important when dealing with spatial data: tapers can help to fight the so-called "edge effects".

Consider the tapered values

$$\{h_T(\boldsymbol{t})\, Y(\boldsymbol{t}),\ \boldsymbol{t} \in L_T\},$$

where $h_T(\boldsymbol{t}) = h(\boldsymbol{t}/T)$, $\boldsymbol{t} = (t^{(1)}, ..., t^{(d)}) \in R^d$, and the taper $h(\boldsymbol{t})$ factorizes as $h(\boldsymbol{t}) = \prod_{i=1}^d \tilde{h}\left(t^{(i)}\right), t^{(i)} \in R^1$, with $\tilde{h}(\cdot)$ satisfying the assumption below.

H1. $\tilde{h}(t), t \in R^1$, is a positive even function of bounded variation with bounded support: $\tilde{h}(t) = 0$ for $|t| > 1$.

Denote

$$\tilde{H}_{k,T}(\lambda) = \sum_{t=-T}^{T} [\tilde{h}_T(t)]^k e^{-i\lambda t},$$

$$H_{k,T}(\boldsymbol{\lambda}) = \sum_{\boldsymbol{t} \in L_T} [h_T(\boldsymbol{t})]^k e^{-i(\boldsymbol{\lambda},\boldsymbol{t})} = \prod_{i=1}^{d} \tilde{H}_{k,T}(\lambda^{(i)}), \tag{2.2}$$

where $\tilde{h}_T(t) = \tilde{h}(t/T)$, $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_d)$, and k is a positive integer number.

Note that evaluation of asymptotic behavior of spectral estimates is based on the properties of functions $\tilde{H}_{k,T}(\lambda)$, which, in its own turn, is based on properties of functions $\tilde{h}(t)$. For example, the assumption that $\tilde{h}(t)$ is of bounded variation allows to write down useful upper bounds for $\tilde{H}_{k,T}(\lambda)$. In what follows we will introduce some further assumptions on tapers.

Define the finite Fourier transform of tapered data $\{h_T(\boldsymbol{t})\, Y(\boldsymbol{t}), \boldsymbol{t} \in L_T\}$:

$$d_T^h(\boldsymbol{\lambda}) = \sum_{\boldsymbol{t} \in L_T} h_T(\boldsymbol{t})\, Y(\boldsymbol{t}) e^{-i(\boldsymbol{\lambda},\boldsymbol{t})},\ \boldsymbol{\lambda} \in \mathbb{T},$$

and the tapered periodogram of the second order (provided that $H_{2,T}(\mathbf{0}) \neq 0$):

$$I_T^h(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^d\, H_{2,T}(\mathbf{0})}\, d_T^h(\boldsymbol{\lambda}) d_T^h(-\boldsymbol{\lambda}).$$

2.3. Ibragimov contrast function and estimators

We begin with the following assumptions concerning our parametric model to be estimated.

B1. Let $Y(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{Z}^d$, be a real-valued measurable stationary Gaussian random field with zero mean and a spectral density $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, where $\boldsymbol{\lambda} \in \mathbb{T} = (-\pi, \pi]^d$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$, and Θ is a compact set. Assume that $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$, where $\boldsymbol{\theta}_0$ is the true value of the parameter vector $\boldsymbol{\theta}$.

B2. If $\boldsymbol{\theta_1} \neq \boldsymbol{\theta_2}$ then $f(\boldsymbol{\lambda}, \boldsymbol{\theta_1}) \neq f(\boldsymbol{\lambda}, \boldsymbol{\theta_2})$ for almost all $\boldsymbol{\lambda} \in \mathbb{T}$ with respect to the Lebesgue measure.

We restrict our study to the fields with $d \leq 3$. Note that the results on consistency of estimators will hold for the general $d \geq 1$, however, for asymptotic normality we will impose the restriction $d \leq 3$, since only for this dimensions we are able to control the bias with the help of tapers.

Remark 2.1. *In what follows, by differentiability with respect to $\boldsymbol{\theta}$, we mean differentiability in the interior of Θ.*

To define the Ibragimov functional, we will use a weight function satisfying the next assumption.

B3. There exists a nonnegative function $w(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in \mathbb{T}$, such that

1. $w(\boldsymbol{\lambda})$ is symmetric, i.e. $w(\boldsymbol{\lambda}) = w(-\boldsymbol{\lambda})$;
2. $w(\boldsymbol{\lambda})f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in L_1(\mathbb{T})$ for all $\boldsymbol{\theta} \in \Theta$.

Under this condition, we set

$$\sigma^2(\boldsymbol{\theta}) = \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda} \tag{2.3}$$

and represent the spectral density in the form:

$$f(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \sigma^2(\boldsymbol{\theta}) \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}). \tag{2.4}$$

For the function $\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ we have

$$\int_{\mathbb{T}} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) d\boldsymbol{\lambda} = 1, \tag{2.5}$$

and we additionally suppose:

B4. The derivatives $\nabla_{\boldsymbol{\theta}} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ exist and it is legitimate to differentiate under the integral sign in equation (2.5), i.e.

$$\nabla_{\boldsymbol{\theta}} \int_{\mathbb{T}} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda} = \int_{\mathbb{T}} \nabla_{\boldsymbol{\theta}} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda} = 0.$$

We consider the following contrast field based on the tapered periodogram defined above:

$$U_T(\boldsymbol{\theta}) = -\int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \, d\boldsymbol{\lambda} \tag{2.6}$$

and the minimum contrast estimator

$$\hat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta} \in \Theta} U_T(\boldsymbol{\theta}).$$

We also define the functions

$$K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) = -\int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \log \frac{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)} \, d\boldsymbol{\lambda} \tag{2.7}$$

and

$$U(\boldsymbol{\theta}) = -\int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda}. \tag{2.8}$$

The minimum contrast property of the function $K(\theta_0, \theta)$ is stated in Theorem 2.1 below, well as the conditions under which $U_T(\boldsymbol{\theta})$ is the contarst field related to $K(\theta_0, \theta)$.

Remark 2.2. *The weight function w used in the contrast functional U_T is aimed to compensate for possible singularities of the spectral density. In some cases, the weight function is not required, for example, when spectral density is bounded.*

2.4. Assumptions

We now formulate conditions needed to state the results on consistency and asymptotic normality of Ibragimov estimators for Gaussian fields .

We will suppose that there exist functions $\alpha_i : \Theta \in (0, 1)$, $i = 1, ..., d$, such that the following conditions are satisfied (with the proper choice of a weight function $w(\boldsymbol{\lambda})$). To simplify notations in formulation of the conditions below we will omit the argument $\boldsymbol{\theta}$ in functions $\alpha_i(\boldsymbol{\theta})$ and write them simply as α_i.

B5. For all $\boldsymbol{\theta} \in \Theta$ $\quad f(\boldsymbol{\lambda}, \boldsymbol{\theta}) = O(\prod_{i=1}^{d} |\lambda_i|^{-\alpha_i})$ as $\lambda_i \to 0$, and $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ is bounded for $\delta \le |\boldsymbol{\lambda}| \le \pi$ for all $\delta > 0$; $w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = O(\prod_{i=1}^{d} |\lambda_i|^{\alpha_i})$ as $\lambda_i \to 0$.

B6. There exists a function $\upsilon(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in \mathbb{T}$, such that

1. the function $h(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \upsilon(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ is uniformly continuous in $\mathbb{T} \times \Theta$;
2. $w(\boldsymbol{\lambda})/\upsilon(\boldsymbol{\lambda}) = O(\prod_{i=1}^{d} |\lambda_i|^{\alpha_i})$ as $\lambda_i \to 0$.

B7. The function $\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ is twice differentiable in Θ and

1. $w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = O(\prod_{i=1}^{d} |\lambda_i|^{\alpha_i})$ as $\lambda_i \to 0$ for all i, j, $\boldsymbol{\theta} \in \Theta$;
2. $w(\boldsymbol{\lambda}) \frac{\partial}{\partial\theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = O(\prod_{i=1}^{d} |\lambda_i|^{\alpha_i})$ as $\lambda_i \to 0$ for all i, $\boldsymbol{\theta} \in \Theta$;
3. the second order derivatives $\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$, $i, j = 1, \ldots, q$ are continuous in $\boldsymbol{\theta}$.

Remark 2.3. *In conditions* **B5-B7** *the behavior of the spectral density and some of its derivatives is prescribed at the point of singularity. These conditions appear quite naturally and are analogous, for example, to those which are used in* [23] *for the case of Whittle estimators, with evident modifications since we use another functional to construct estimators. It is clear from assumptions* **B6-B7** *that the weight function $w(\boldsymbol{\lambda})$ is constructed to compensate the singularity of the spectral density.*

B8. The matrices $\mathbf{S}(\boldsymbol{\theta}) = (s_{ij}(\boldsymbol{\theta}))_{i,j=1,...,q}$ and $\mathbf{A}(\boldsymbol{\theta}) = (a_{ij}(\boldsymbol{\theta}))_{i,j=1,...,q}$ are positive definite, where for $i, j = 1, ..., q$:

$$\begin{aligned}(s_{ij}(\boldsymbol{\theta})) &= \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda} \\ &= \sigma^2(\boldsymbol{\theta}) \int_{\mathbb{T}} w(\boldsymbol{\lambda}) \Big[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \\ &- \frac{1}{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})} \frac{\partial}{\partial\theta_i} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \frac{\partial}{\partial\theta_j} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \Big]\, d\boldsymbol{\lambda}, \\ (a_{ij}(\boldsymbol{\theta})) &= 2(2\pi)^d \int_{\mathbb{T}} f^2(\boldsymbol{\lambda}, \boldsymbol{\theta}) w^2(\boldsymbol{\lambda}) \frac{\partial}{\partial\theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \frac{\partial}{\partial\theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda} \\ &= 2(2\pi)^d (\sigma^2(\boldsymbol{\theta}))^2 \int_{\mathbb{T}} w^2(\boldsymbol{\lambda}) \frac{\partial}{\partial\theta_i} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \frac{\partial}{\partial\theta_j} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda}.\end{aligned}$$

Remark 2.4. *The last two conditions are used to achieve proper rate of convergence of bias to 0.*

H2. The taper $\tilde{h}(t)$ is a Lipschitz-continuous function on $[-1, 1]$ and $\tilde{h}(-1) = \tilde{h}(1) = 0$.

B9. The spectral density $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, the function $w(\boldsymbol{\lambda})$ and the function $\varphi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = w(\boldsymbol{\lambda}) \frac{\partial}{\partial\theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ are such that one of the following conditions holds:

(i) φ is twice boundedly differentiable;

or

(ii) the convolution $g(\boldsymbol{u}) = \int_{\mathbb{T}} f(\boldsymbol{\lambda}) \varphi(\boldsymbol{\lambda} + \boldsymbol{u})\, d\boldsymbol{\lambda}$ is twice boundedly differentiable at zero.

Remark 2.5. *One example of a long-range dependent random field for which all the above conditions are satisfied (with the proper choice of the weight function $w(\boldsymbol{\lambda})$), is the solution to the following equation: $(1 - B_1)^{d_1}(1 - B_2)^{d_2} Y(t_1, t_2) = \varepsilon(t_1, t_2)$, with $\varepsilon(t_1, t_2)$ being a two-dimensional white noise and B_i being a backward shift operator for i-th coordinate, $i = 1, 2$. This example of fractional spacial autoregression, in more general form, is considered in detail in Section 5, together with other two exaples, namely, the Gegenbauer random fields. Appropriate for these models weight functions $w(\boldsymbol{\lambda})$ are presented. Note that the estimation of fields obeying the fractional spacial authoregressive model have been considered in [16] (based on Whittle functional), however, only the consistency of the estimators was stated therein. Within the approach of the present paper, we are able to construct consistent and asymptotically normal estimates for this model (see Section 5.1 below).*

2.5. *Theorems on consistency and asymptotic normality of the estimators*

Theorem 2.1. *Let the conditions* **B1-B3**, **B5-B6**, *and* **H1** *to hold. Then, the function* $K(\boldsymbol{\theta}_0;\boldsymbol{\theta})$ *defined by* (2.7) *is the contrast function for the contrast field* $U_T(\boldsymbol{\theta})$ *defined by* (2.6). *Moreover, the minimum contrast estimator* $\hat{\boldsymbol{\theta}}_T$

$$\hat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}\in\Theta} U_T(\boldsymbol{\theta}) \tag{2.9}$$

is a consistent estimator of the parameter vector $\boldsymbol{\theta}$. *That is, there is a convergence in* P_0 *probability:*

$$\hat{\boldsymbol{\theta}}_T \xrightarrow{P_0} \boldsymbol{\theta_0}, \quad \text{as } T \longrightarrow \infty.$$

The estimator $\hat{\sigma}^2_T \xrightarrow{P_0} \sigma^2(\boldsymbol{\theta_0})$, *as* $T \longrightarrow \infty$, *where* $\hat{\sigma}^2_T$ *is an estimator of the parameter* $\sigma^2(\theta_{\mathbf{0}})$ *given by*

$$\hat{\sigma}^2_T = \int_{\mathbb{T}} I^h_T(\boldsymbol{\lambda}) w(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}.$$

The proof will be given in Section 4.

Theorem 2.2. *Let the conditions* **B1- B9** *and* **H1- H2** *to hold and* $d \leq 3$. *Then the minimum contrast estimator defined by* (2.9) *is asymptotically normal, that is, as* $T \longrightarrow \infty$

$$T^{d/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}_q(0, e(h)\mathbf{S}^{-1}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)\mathbf{S}^{-1}(\boldsymbol{\theta}_0)), \tag{2.10}$$

where the entries of the matrices $\mathbf{S}(\boldsymbol{\theta}) = (s_{ij}(\boldsymbol{\theta}))$ *and* $\mathbf{A}(\boldsymbol{\theta}) = (a_{ij}(\boldsymbol{\theta}))$ *are defined in condition* ***B8***, $e(h)$ *is given by the formula*

$$e(h) = \Big(\int (\tilde{h}(t))^4 dt \Big(\int (\tilde{h}(t))^2 dt\Big)^{-2}\Big)^d, \tag{2.11}$$

and $\mathcal{N}_q(\cdot,\cdot)$, *denotes the q-dimensional Gaussian law.*

The proof will be given in Section 4.

Remark 2.6. *In the discrete-time case the domain over which the field is observed is usually taken to be* $L_T = [1, T]^d$. *Our results remain valid for such a domain as well, we just need to adjust the assumptions on a taper* $\tilde{h}(t)$. *Namely, assumption* **H1** *must be modified as follows:* $\tilde{h}(t)$ *is a positive measurable function of bounded variation with support on* $[0, 1]$ *and* $h(0) = 0$, $h(1 - v) = h(v)$ *for* $0 \leq v \leq \frac{1}{2}$.

Remark 2.7. *An example of a taper* $\tilde{h}(t)$ *satisfying our assumptions is*

$$\tilde{h}(t) = \frac{1}{2}(1 + \cos(4\pi t)), \quad t \in [-1, 1].$$

This is just a modification of the well known cosine bell (or the Tukey-Hanning taper)

$$\tilde{h}(t) = \frac{1}{2}(1 - \cos(2\pi t)), \quad t \in [0,1],$$

suitable for the domain $L_T = [1, T]^d$.

Remark 2.8. *To establish the consistency of a minimum contrast estimator* $\hat{\theta}_T$, *which corresponds to a functional* $U_T(\theta)$, *the following standard reasonings are used: one needs to check that the convergence* (2.1) *holds in probability, and then, due to Theorem* 3.4.1 [28], *it is sufficient to prove that the convergence* (2.1) *holds uniformly with respect to* θ.

The standard approach to state the asymptotic normality of the estimator $\widehat{\boldsymbol{\theta}}_T$ *is to consider the relation:*

$$\nabla_\theta U_T\left(\widehat{\boldsymbol{\theta}}_T\right) = \nabla_\theta U_T\left(\boldsymbol{\theta}_0\right) + \nabla_\theta \nabla'_{\boldsymbol{\theta}} U_T\left(\boldsymbol{\theta}_T^*\right)\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right), \;\; \left|\boldsymbol{\theta}_T^* - \boldsymbol{\theta}_0\right| < \left|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right|,$$

and then evaluate the asymptotic behavior of $\nabla_\theta U_T\left(\boldsymbol{\theta}_0\right)$ *and* $\nabla_\theta \nabla'_\theta U_T\left(\boldsymbol{\theta}_T^*\right)$.

Therefore, one needs the results on large sample properties of the empirical spectral functionals of the form:

$$J_T^h\left(\varphi\right) = \int_{\mathbb{T}} I_T^h\left(\boldsymbol{\lambda}\right) \varphi\left(\boldsymbol{\lambda}\right) w\left(\boldsymbol{\lambda}\right) d\boldsymbol{\lambda},$$

where $I_T^h\left(\boldsymbol{\lambda}\right)$ *is the periodogram based on tapered data. In particular, for derivation of the asymptotic normality results it is important to state the conditions guaranteeing the proper rate of convergence of bias to zero, that is, the following relation to hold:*

$$T^{d/2}\left(EJ_T^h\left(\varphi\right) - \int_{\mathbb{T}} f\left(\boldsymbol{\lambda}\right) \varphi\left(\boldsymbol{\lambda}\right) w\left(\boldsymbol{\lambda}\right) d\boldsymbol{\lambda}\right) \to 0 \quad as \;\; T \to \infty. \tag{2.12}$$

In the above Theorem 2.2 *the conditions which help to control bias are* **H1-H2** *and* **B9**.

Remark 2.9. *Note that the previous results on Ibragimov estimators (see, e.g.* [2], [3], [4], [7], [11]*) where stated under the conditions of integrability of the spectral density* $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ *and function* $w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ *(and integrability of some derivatives of the latter function), and the condition for bias control* (2.12) *was just imposed. We mention also that the investigation of bias for estimators of spectral functionals, in non-parametric setting, was presented in* [5], [40], [41].

In comparison with previous papers on Ibragimov estimators, in the present paper we give all conditions for consistency and asymptotic normality in the different form, prescribing exactly the behavior of the spectral density $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, *the function* $w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ *and some derivatives at the point of singularity; we also give the corresponding conditions needed for the proper rate of convergence of bias (that is for* (2.12) *to hold) in the exact form. The form of conditions* *makes them very operational and easily applicable to concrete models of random fields with singularities (as demonstrated with some examples in Section* 5*).*

We mention that for the case of Whittle estimators for Gaussian random fields with spectral densities possessing singularities of multiplicative form, the results on consistency and asymptotic normality were obtained in [36]*, therein the weight function* $w(\boldsymbol{\lambda})$ *and taper are both related to the order of singularity of the spectral density. We discuss the results from* [36] *in Section* 3.2*, and we also describe in Section* 3.1 *conditions used in* [28] *for the case of weakly dependent random fields.*

Remark 2.10. *As follows from the results by Guyon (see, e.g.,* [28]*) outlined in the next section, to control bias for Ibragimov minimum contrast estimators in the case of short-range dependent random fields the following conditions can be used: the taper* $\tilde{h}(t)$ *is in* $C^2[-1,1]$*; the spectral density* $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ *and the function* $\varphi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = w(\boldsymbol{\lambda})\frac{\partial}{\partial \theta_i}\log\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ *belong to* $C^2(\mathbb{T})$*.*

Remark 2.11. *Concerning the condition* **B8***. If instead of condition* 3 *in* **B7** *we use the following condition* $3'$*: the second order derivatives* $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$*,* $i, j = 1, \ldots, q$ *are continuous in both* $(\boldsymbol{\lambda}, \boldsymbol{\theta})$*,* $\lambda \neq 0$*, then one can use just non-degenerate* $\mathbf{S}(\boldsymbol{\theta})$ *in* **B8***. Indeed, since the point* $\boldsymbol{\theta_0}$ *is the point of maximum for the functional* $-U(\boldsymbol{\theta}) = \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)w(\boldsymbol{\lambda})\log\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})\,d\boldsymbol{\lambda}$*, then under the above 3′ the matrix* $\nabla_\theta\nabla'_\theta(-U_T(\boldsymbol{\theta_0}))$ *have to be negatively semi-definite, and supposing* $\mathbf{S}(\boldsymbol{\theta_0}) = \nabla_\theta\nabla'_\theta(U_T(\boldsymbol{\theta_0}))$ *to be nondegenerate, we conclude that* $\nabla_\theta\nabla'_\theta(-U_T(\boldsymbol{\theta_0}))$ *is negatively definite, and, therefore,* $\mathbf{S}(\boldsymbol{\theta_0}) = \nabla_\theta\nabla'_\theta(U_T(\boldsymbol{\theta_0}))$ *is positively definite. (See, for example,* [31]*, comments after condition N8 therein).*

3. On the Whittle estimators with tapered data

In this section we will give a brief review of some results from the literature on the Whittle estimators based on tapered data, with a particular focus on the conditions on tapers which help to control the bias of estimators.

Recall that the Whittle estimators can be defined as minimum contrast estimators with the following contrast field:

$$U_T(\boldsymbol{\theta}) = \frac{1}{2(2\pi)^d}\int_{\mathbb{T}}\left(\log f(\boldsymbol{\lambda}; \boldsymbol{\theta}) + \frac{I_T(\boldsymbol{\lambda})}{f(\boldsymbol{\lambda}; \boldsymbol{\theta})}\right) w(\boldsymbol{\lambda})\,d\boldsymbol{\lambda}, \quad \boldsymbol{\theta} \in \Theta, \tag{3.1}$$

and corresponding contrast function

$$K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) = \frac{1}{2(2\pi)^d}\int_{\mathbb{T}}\left(\frac{f(\boldsymbol{\lambda}; \boldsymbol{\theta}_0)}{f(\boldsymbol{\lambda}; \boldsymbol{\theta})} - 1 - \log\frac{f(\boldsymbol{\lambda}; \boldsymbol{\theta}_0)}{f(\boldsymbol{\lambda}; \boldsymbol{\theta})}\right) w(\boldsymbol{\lambda})\,d\boldsymbol{\lambda}, \quad \boldsymbol{\theta}_0, \boldsymbol{\theta} \in \Theta.$$

Note that here again the function w is aimed to compensate for possible singularities of the spectral density and in some cases is not required.

3.1. Case of weakly dependent random fields: results from [28]

Apparently, one of the first publications that addressed the problem of edge effects was [27]: considering the usual parametric Whittle estimates for lattice data, it found that such estimates had bias of order $N^{-1/d}$, for fields observed on a rectangle $P_N = \{1, ..., n_1\} \times ... \times \{1, ..., n_d\}$ in Z^d, with N being $\prod_{i=1}^d n_i$, or of order n^{-1} if P_N is a cube of edge n. That is, for $d \geq 2$ the bias is of the same order or a higher order as the standard deviation which is usually $O(N^{-1/2})$. One possible solution to the described edge effects problem is tapering the data at the edges of the observation domain.

We review the results by [27, 28]. There the domain L_T is taken to be of the form $L_T = \prod_{i=1}^d [1, T_i]$, we simplify our exposition and consider the cube, that is, all $T_i = T$ and $L_T = [1, T]^d$.

Instead of the observed data $\{Y(\boldsymbol{t}),\ \boldsymbol{t} \in L_T\}$, one considers the tapered data $\{h_T(\boldsymbol{t})Y(\boldsymbol{t}),\ \boldsymbol{t} \in L_T\}$, where the tapers are of the following form:

$$h_T(\boldsymbol{t}) = h\left(\frac{\boldsymbol{t} - 1/2}{T}\right) = \prod_{i=1}^{d} \tilde{h}\left(\frac{t_i - 1/2}{T}\right),$$

and

$$\tilde{h}(u) = \begin{cases} g\left(\frac{2u}{\rho}\right) & \text{if } 0 \leq u \leq \frac{1}{2}\rho \\ 1 & \text{if } \frac{1}{2}\rho \leq u \leq \frac{1}{2} \\ \tilde{h}(1-u) & \text{if } \frac{1}{2} \leq u \leq 1 \end{cases}$$

where the function $g : [0, 1] \to [0, 1]$ is such that $g(0) = 0,\ g(1) = 1$, and it is increasing and belonging to $\mathcal{C}^2$, and $0 \leq \rho \leq 1$. (In such a way one obtains the so-called *g-taper* which tapers $100(1-\rho)\%$ of the edge values.)

We next formulate the results on bias control for tapered estimators stated by Guyon (see, [28]).

Consider $J(\phi) = \int_{\mathbb{T}} f(\boldsymbol{\lambda})\phi(\boldsymbol{\lambda})\ d\boldsymbol{\lambda}$ and the corresponing estimator based on tapered data $J_T(\phi) = \int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda})\phi(\boldsymbol{\lambda})\ d\boldsymbol{\lambda}$.

Proposition 1. Let $\phi \in \mathcal{C}(\mathbb{T})$ and assume that the taper h and the spectral density f belong to $\mathcal{C}^2$. Then as $T \to \infty$,

$$E\left[J_T(\phi) - J(\phi)\right] = CT^{-2}\left(1 + o(1)\right),$$

where C is a constant.

Therefore, under the above conditions, if ρ is fixed the bias is of order T^{-2}, thus smaller than $T^{-d/2}$, for $d = 1, 2, 3$. If $\rho = \rho_T \to 0$, C behaves like ρ_T^{-1} so that the bias is like $o\left(T^{-2+\frac{1}{4}}\right)$, which is still smaller than $T^{-d/2}$. That is, for $d = 1, 2, 3$,

$$\lim_{T\to\infty} E\left[T^{d/2}(J_T(\phi) - J(\phi))\right] = 0.$$

Moreover, if $\phi_1, \phi_2 \in \mathcal{C}(\mathbb{T})$, the taper h and the spectral density f belong to $\mathcal{C}^2$, then for a Gaussian random field we have the following asymptotic behavior of the covariance:

$$\lim_{T\to\infty} T^d Cov\left(J_T(\phi_1), J_T(\phi_2)\right) = 2(2\pi)^d\, e(h) \int_{\mathbb{T}} \phi_1(\boldsymbol{\lambda})\phi_2(\boldsymbol{\lambda}) f^2(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}$$

where the taper factor $e(h)$ is given by (2.11), $e(h) \geq 1$ and $e(h) = 1$ if there is no tapering. It it possible to choose ρ_T such that the taper factor will tend to 1.

Remark 3.1. *In view of the above Proposition 1, in order to control the bias for the Whittle estimator, the following conditions can be used:*
the taper $\tilde{h}(t)$ is in $C^2[-1,1]$; the spectral density $f(\boldsymbol{\lambda},\boldsymbol{\theta})$ is in $C^2(\mathbb{T})$ and the function $\frac{\partial}{\partial\theta_i} f(\boldsymbol{\lambda},\boldsymbol{\theta})$ is continuous with respect to $(\boldsymbol{\lambda},\boldsymbol{\theta})$.

We also mention that the asymptotic normality of the Whittle estimators was stated in [28] under some mixing assumptions.

3.2. Case of strongly dependent random fields: results from [36]

In the paper [36] the following two models of spectral densities with singularities were considered:

A1 There exist functions $\alpha_i : \Theta \to (0,1)$, $1 \leq i \leq d$, such that we have $f(\boldsymbol{\lambda},\boldsymbol{\theta}) = f_0(\boldsymbol{\lambda},\boldsymbol{\theta}) \prod_{i=1}^d f_i(\lambda_i,\boldsymbol{\theta})$ where we assume for all $\delta > 0$ that $f_i(\lambda_i,\boldsymbol{\theta}) = O(|\lambda_i|^{-\alpha_i(\boldsymbol{\theta})-\delta})$, $|\lambda_i| \to 0$ for each $i = 1,...,d$ and $f_0(\boldsymbol{\lambda},\boldsymbol{\theta})$ is a positive twice continuously differentiable function of $(\boldsymbol{\lambda},\boldsymbol{\theta})$

A2 In this case, we assume there exists a function $\alpha : \Theta \to (0,d)$, such that the spectral density $f(\boldsymbol{\lambda},\boldsymbol{\theta})$ satisfies for all $\delta > 0 : f(\boldsymbol{\lambda},\boldsymbol{\theta}) = \|\boldsymbol{\lambda}\|^{-\alpha(\boldsymbol{\theta})+\delta}(1 + o(1))$ where the $o(1)$ is uniform with respect to $\boldsymbol{\lambda}$. (For more details, see [36]).

The authors study the Whittle estimator $\hat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}\in\Theta} U_T(\boldsymbol{\theta})$ for both models **A1** and **A2** using the tapered data $\{h_{T,\varepsilon}(t)Y(t)\}$, the functional $U_T(\theta)$ is given by (3.1), where the function $w(\boldsymbol{\lambda})$ (called by the authors smoothing or regularizing function) depends on the shape of singularity. Under **A1** $w(\boldsymbol{\lambda}) = \prod_{i=1}^d w_i(\lambda_i)$, $w_i(\lambda_i) = |\lambda_i|^{\nu}$, or $w(\boldsymbol{\lambda}) = \mathbb{1}_{\mathbb{T}}(\boldsymbol{\lambda})$ (the indicator function on the d-torus) under **A2**.

The taper function $h_{T,\varepsilon}(t) = \prod_{i=1}^d \tilde{h}_\varepsilon\left(\frac{t_i}{T}\right)$, where $\tilde{h}_\varepsilon : [0,1] \to [0,1]$ is of the following form:

$$h_\varepsilon(u) = \begin{cases} u/\varepsilon & \text{if } u \leq \varepsilon \\ 1 & \text{if } \varepsilon \leq u \leq 1-\varepsilon \\ h_\varepsilon(1-u) & \text{if } u > 1-\varepsilon \end{cases}$$

with $\varepsilon = T^{-\gamma}$.

In the course of derivation of asymptotic normality for the Whittle estimator the following result on bias was obtained for the functional $J_T(\varphi) = \int_{\mathbb{T}} I_T^h(\lambda)\varphi(\lambda)\,d\lambda$ taken as an estimator for $J(\varphi) = \int_{\mathbb{T}} f(\lambda)\varphi(\lambda)\,d\lambda$ with a particular $\varphi(\lambda) = \varphi_i(\lambda,\theta) = w(\lambda)\frac{\partial}{\partial\theta_i} f^{-1}(\lambda,\theta)$.

Proposition 2. Assume **A1** or **A2**, and let $w(\boldsymbol{\lambda}) = \prod_{i=1}^d w_i(\lambda_i)$, under **A1**, or $w(\boldsymbol{\lambda}) = \mathbb{1}_{\mathbb{T}}(\boldsymbol{\lambda})$, under **A2**. Suppose further that $\gamma < (1 + v - {}^{d}\!/_{2})/v$, under **A1**, or $\gamma < 2 - {}^{d}\!/_{2}$ under **A2**. Then, as $T \to \infty$

$$E[J_T(\varphi) - J(\varphi)] = O(T^{-(1+\kappa(1-\gamma))}),$$

where $\kappa = \nu$ under **A1**, or $\kappa = 1$ under **A2**, and, therefore, the bias is of order $o(T^{-d/2})$ if $d \le 3$.

Note that in the derivation of the above result, the form of φ (as needed for a particular case of the Whittle functional) was essentially used. With this function, and under conditions of Proposition 2, some regularity properties of convolution $\int_{\mathbb{T}} \varphi(\boldsymbol{\lambda} - \boldsymbol{\mu}) f(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}$ were obtained, and used for bias evaluation, as well as an interplay between a taper, and the function $w(\boldsymbol{\lambda})$ under **A1**. Multiplicatives structure of the functions defining $J(\varphi)$ was also essential under **A1**.

We also mention that asymptotic normality of the Whittle estimators was stated in [36] under the assumption **A1** using the CLT analogous to that of [24], and, under the assumption **A2**, the CLT stated in [21] was used.

4. Proofs

The proofs are based on the results on asymptotic properties of the functionals $J_T(\varphi)$ obtained via evaluation of their cumulants. These cumulants can be represented in the form of some integrals involving spectral densities, weight functions and Fejér type kernels. Although these kernels are different for non-tapered and tapered cases, all technique for the proofs works similarly in both cases (see more details in Appendix).

The main difference from the previous papers on Ibragimov estimators is that here, using the tapers, we are able to formulate the exact conditions to achieve proper rate of convergence of bias to state asymptotic normality. We present here principal steps for the proofs and state Lemma 4.1, which gives the possibility to control bias as needed for Theorem 2.2.

The main tool to obtain asymptotic normality of the estimators is Theorem A.2 (see Appendix). This theorem gives sufficient conditions for asymptotic normality of $J_T(\varphi)$. We will also need conditions for convergence in probability of functionals $J_T(\varphi)$ which we discuss in the next remark.

Remark 4.1. *To state the convergence in probability*

$$J_T(\varphi) = \int_{\mathbb{T}} I_T^h(\lambda)\varphi(\lambda)\,d\lambda \xrightarrow{P} J(\varphi) = \int_{\mathbb{T}} f(\lambda)\varphi(\lambda)\,d\lambda, \tag{4.1}$$

it is sufficient to show that:

(i) $\int_{\mathbb{T}}(EI_T^h(\lambda) - f(\lambda))\varphi(\lambda)\, d\lambda \to 0$ *;*

(ii) $\int_{\mathbb{T}}(I_T^h(\lambda) - EI_T^h(\lambda))\varphi(\lambda)\, d\lambda \xrightarrow{P} 0.$

Convergence (ii) will hold if $VarJ_T(\varphi) = E(\int_{\mathbb{T}}(I_T^h(\lambda) - EI_T^h(\lambda))\varphi(\lambda)\, d\lambda)^2 \to 0$*. Therefore, for* (4.1) *to hold we can use conditions guaranteeing the convergence (i) and convergence of the variance of the normalized functional* $T^{d/2}(J_T(\varphi) - EJ_T(\varphi))$ *to a finite limit. In particular, convergence* (4.1) *holds under the conditions on* $f(\lambda)$ *and* $\varphi(\lambda)$ *imposed in Theorem* A.2*.*

Proof of Theorem 2.1*.* In view of Remark 4.1, taking into account the expression for $U_T(\boldsymbol{\theta})$, we conclude that convergence

$$\lim_{T\to\infty} U_T(\boldsymbol{\theta}) - U_T(\boldsymbol{\theta}_0) \xrightarrow{P} U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}_0) = K(\boldsymbol{\theta}_0; \boldsymbol{\theta})$$

holds under conditions **B5**.

Moreover, under conditions **B5** and **B6**, the above convergence holds uniformly with respect to $\boldsymbol{\theta} \in \Theta$. Indeed, denoting by $\eta(\varepsilon)$ the modulus of continuity of the function $h(\lambda, \theta)$, we can write:

$$\sup\left\{|U_T(\theta_1) - U_T(\theta_2)|, \theta_1, \theta_2 \in \Theta, |\theta_1 - \theta_2| \le \varepsilon\right\} \le \eta(\varepsilon) \int_{\mathbb{T}} I_T^h(\lambda) \frac{w(\lambda)}{v(\lambda)} d\lambda,$$

and the integral in the r.h.s. is asymptotically bounded in probability under conditions **B5** and **B6**. Therefore, in view of Theorem 3.4.1 from Guyon [28], we conclude that the estimator $\hat{\boldsymbol{\theta}}_T$ is consistent: $\hat{\boldsymbol{\theta}}_T \xrightarrow{P_0} \boldsymbol{\theta_0}$, as $T \to \infty$.

Minimum contrast property for $K(\boldsymbol{\theta}_0; \boldsymbol{\theta})$ follows from Jensen's inequality:

$$\begin{aligned} -K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) &= \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \log \frac{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)}\, d\boldsymbol{\lambda} \\ &= \sigma^2(\boldsymbol{\theta}_0) \int_{\mathbb{T}} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \log \frac{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)}\, d\boldsymbol{\lambda} \\ &\le \sigma^2(\boldsymbol{\theta}_0) \log \int_{\mathbb{T}} w(\boldsymbol{\lambda}) \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda} = 0, \end{aligned}$$

therefore, $K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) \ge 0$, and, moreover, $K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) > 0$ if $\psi(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) \not\equiv \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ for $\boldsymbol{\theta}_0 \ne \boldsymbol{\theta}$ almost everywhere with respect to the Lebesgue measure.

Proof of Theorem 2.2*.* Applying the mean value theorem, we have

$$\nabla_{\boldsymbol{\theta}} U_T(\hat{\boldsymbol{\theta}}_T) = \nabla_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_0) + \nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_T^*)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0),$$

where $\left|\boldsymbol{\theta}_T^* - \boldsymbol{\theta}_0\right| < \left|\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right|$ and

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}) &= -\int_{\mathbb{T}} I_T(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \nabla_{\boldsymbol{\theta}} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda} \\
&= \Big\{ -\int_{\mathbb{T}} I_T(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda} \Big\}_{i=1,\dots,q}, \\
\nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}) &= -\int_{\mathbb{T}} I_T(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda} \\
&= \Big\{ -\int_{\mathbb{T}} I_T(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda} \Big\}_{i,j=1,\dots,q}.
\end{aligned}$$

It follows from the definition of minimum contrast estimators that for T sufficiently large

$$\nabla_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_0) = -\nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_T^*)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0). \tag{4.2}$$

So, we need to show that

(i) $\nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}) \to \mathbf{S}(\boldsymbol{\theta_0})$ in P_0-probability,

(ii) $T^{d/2} \nabla_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N_q\left(0, e(h)\mathbf{A}(\boldsymbol{\theta_0})\right)$),

where $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{S}(\boldsymbol{\theta})$ are given in condition **B8**. Then, by Slutsky's lemma, relation (2.10) is a consequence of (4.2) and (i)–(ii).

Note that in view of **B4**

$$\int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta_0}) w(\boldsymbol{\lambda}) \nabla_{\boldsymbol{\theta}} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta_0}) d\boldsymbol{\lambda} = 0,$$

therefore, $\nabla_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_0) = \big(J_T(\varphi_i) - J(\varphi_i)\big)_{i=1,\dots,q}$, and convergence (ii) will hold if we show that

$$T^{d/2}\big(J_T(\varphi_i) - EJ_T(\varphi_i)\big)_{i=1,\dots,q} \xrightarrow{\mathcal{D}} N_q\left(0, e(h)\mathbf{A}(\boldsymbol{\theta_0})\right) \tag{4.3}$$

and

$$T^{d/2}\big(EJ_T(\varphi_i) - J(\varphi_i)\big) \to 0,\ i = 1, \dots, q, \tag{4.4}$$

where $\varphi_i(\boldsymbol{\lambda}) = w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta_0})$. Convergence (4.3) holds under conditions **B5** and **B7**(2) in view of Theorem A.2; convergence (i) holds under **B5**, **B6**, **B7**(1), **B7**(3) in view of Remark 4.1 (and taking into account consistency of $\hat{\boldsymbol{\theta}}_T$). Convergence (4.4) will hold under conditions **H1-H2** and **B9** in view of Lemma 4.1 stated below. Note that in the previous papers convergence (4.4) was imposed as assumption (see, for example, [7, 10]).

Convergence

$$\hat{\sigma}_T^2 = \int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda}) w(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \xrightarrow{P_0} \sigma^2(\boldsymbol{\theta_0}) = \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta_0}) w(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \tag{4.5}$$

can be stated with the use of the same arguments as those for convergence $U_T(\boldsymbol{\theta}) \xrightarrow{P_0} U(\boldsymbol{\theta})$, which was stated above. Namely, we can apply again our observation form Remark 4.1, that is to state (4.5) we can use Theorem A.2. One can see that under condition **B5**, $w(\boldsymbol{\lambda})$ compensates the singularity of $f(\boldsymbol{\lambda}, \boldsymbol{\theta_0})$, and therefore, the convergence (4.5) follows.

Lemma 4.1. *Let the taper $\tilde{h}(t)$ satisfy the assumptions **H1** and **H2**. Suppose further that one of the following conditions holds:*

(i) f is twice boundedly differentiable and $\varphi \in L_1(\mathbb{T})$;
(ii) φ is twice boundedly differentiable;
(iii) the convolution $g(u) = \int f(\boldsymbol{\lambda})\varphi(\boldsymbol{\lambda} + \mathbf{u})\, d\boldsymbol{\lambda}$ is twice boundedly differentiable at zero.

Then, as $T \to \infty$,

$$EJ_T(\varphi) - J(\varphi) = O\left(T^{-2}\right). \tag{4.6}$$

Proof. The analog of this lemma for the continuous-parameter fields was obtained in [40], the results on bias evaluation for the spectral functionals of higher-orders in continuous and discrete contexts were stated in [7, 10]. The present proof uses ideas from these papers. With direct calculations, the bias of $J_T(\varphi)$ can be represented as follows:

$$\begin{aligned} E\left[J_T(\varphi) - J(\varphi)\right] &= \int\int \varphi(\boldsymbol{\lambda}) f(\mathbf{u})\, \Phi^h_{2,T}(\mathbf{u} - \boldsymbol{\lambda})\, d\mathbf{u} d\boldsymbol{\lambda} \\ &\quad - \int \varphi(\boldsymbol{\lambda}) f(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \\ &= \int\int \varphi(\boldsymbol{\lambda})\left(f(\boldsymbol{\lambda} + \mathbf{u}) - f(\boldsymbol{\lambda})\right) \Phi^h_{2,T}(\mathbf{u})\, d\mathbf{u} d\boldsymbol{\lambda} \quad (4.7) \\ &= \int\int f(\boldsymbol{\lambda})\left(\varphi(\boldsymbol{\lambda} + \mathbf{u}) - \varphi(\boldsymbol{\lambda})\right) \Phi^h_{2,T}(\boldsymbol{\lambda})\, d\mathbf{u} d\boldsymbol{\lambda} \quad (4.8) \\ &= \int \left(g(\mathbf{u}) - g(\mathbf{0})\right) \Phi^h_{2,T}(\mathbf{u})\, d\mathbf{u}, \quad (4.9) \end{aligned}$$

where $\Phi^h_{2,T}(\mathbf{u}) = \left[(2\pi)^d\ H_{2,T}(\mathbf{0})\right]^{-1} \left|H_{1,T}(\mathbf{u})\right|^2$ is an even function which integrates to 1 (more precisely, this is the kernel of Fejér type, see Appendix), with $H_{1,T}(\mathbf{u})$ being given in (2.2) for $k = 1$, and we have denoted $g(\mathbf{u}) = \int f(\boldsymbol{\lambda})\varphi(\boldsymbol{\lambda} + \mathbf{u})\, d\boldsymbol{\lambda}$.

Therefore, to evaluate the bias we need to analyze the asymptotic behavior of the expressions (4.7)-(4.9). Here we can apply standard arguments if we impose conditions of regularity on the functions f, φ, or, more generally, on their convolution g.

Consider the expression (4.7) and conditions (i).

By Taylor's theorem and assumption that $f(\lambda)$ is twice boundedly differentiable on $\mathbb{T}$,

$$\left| f(\boldsymbol{\lambda} + \mathbf{u}) - f(\boldsymbol{\lambda}) - \sum_{i=1}^{d} u_i \frac{\partial f(\boldsymbol{\lambda})}{\partial \lambda_i} \right| \leq const \cdot \sum_{i=1}^{d} |u_i|^2 .$$

Consider the expression

$$\sum_{i=1}^{d}\int_{\mathbb{T}}\varphi\left(\boldsymbol{\lambda}\right)\frac{\partial f\left(\boldsymbol{\lambda}\right)}{\partial\lambda_i}\int_{\mathbb{T}}u_i\frac{\left|H_{1,T}(\mathbf{u})\right|^2}{\left(2\pi\right)^d H_{2,T}\left(\mathbf{0}\right)}d\mathbf{u}\,d\boldsymbol{\lambda}$$
$$=\sum_{i=1}^{d}\int_{\mathbb{T}}\varphi\left(\boldsymbol{\lambda}\right)\frac{\partial f\left(\boldsymbol{\lambda}\right)}{\partial\lambda_i}\int_{\mathbb{T}}u_i\frac{\prod_{i=1}^{d}\left|\tilde{H}_{1,T}(u_i)\right|^2}{\left(2\pi\right)^d H_{2,T}\left(\mathbf{0}\right)}d\mathbf{u}\,d\boldsymbol{\lambda}\,. \tag{4.10}$$

Since $\left|\tilde{H}_{1,T}(u_i)\right|^2\cdot(2\pi\tilde{H}_{2,T}\left(0\right))^{-1}$ is a kernel of Fejér type and integrates to 1 (see, for example, [19]), the inner integrals in (4.10) reduce to the expressions

$$\int_{-\pi}^{\pi}u_i\frac{\left|\tilde{H}_{1,T}(u_i)\right|^2}{2\pi\tilde{H}_{2,T}\left(\mathbf{0}\right)}du_i\,,$$

which is equal to zero since $\left|\tilde{H}_{1,T}(u_i)\right|^2$ is an even function. Therefore, (4.10) is equal to zero and

$$\left|EJ_T(\varphi)-J\left(\varphi\right)\right|\le const\cdot\sum_{i=1}^{d}\int_{\mathbb{T}}\left|\varphi\left(\boldsymbol{\lambda}\right)\right|\int_{\mathbb{T}}\left|u_i\right|^2\prod_{i=1}^{d}\frac{\left|\tilde{H}_{1,T}(u_i)\right|^2}{2\pi\,\tilde{H}_{2,T}\left(0\right)}d\mathbf{u}\,d\boldsymbol{\lambda}. \tag{4.11}$$

The inner integrals in (4.11) reduce to the expressions

$$\int_{-\pi}^{\pi}\left|u_i\right|^2\frac{\left|\tilde{H}_{1,T}(u_i)\right|^2}{2\pi\tilde{H}_{2,T}\left(0\right)}du_i\,.$$

We note that

$$\tilde{H}_{2,T}\left(0\right)\sim T\int\tilde{h}^2(t)dt \tag{4.12}$$

as $T\to\infty$, since $\tilde{H}_{2,T}\left(0\right)$ supplied with the factor $\frac{1}{T}$ gives, in fact, the partial sum for the integral appearing in the r.h.s. of (4.12).

Next, as in Brillinger [15] we can write:

$$\tilde{H}_{1,T}\left(u\right)=\sum_{t}\tilde{h}\left(t/T\right)e^{-iut}=-\sum_{t}\Delta^t(u)\left(\tilde{h}\left((t+1)/T\right)-\tilde{h}\left(t/T\right)\right), \tag{4.13}$$

where $\Delta^t(u)$ has the form of a product of $(\sin(u/2))^{-1}$ with some bounded with respect to T and u function, $\Delta^0(u)\equiv 1$. For more details we refer, e.g., to [15] (see the proof of Corollary 7.2.1), [19]. Now, taking into account that the taper function $\tilde{h}(t)$ is Lipschitz-continuous, we can write the following estimate:

$$\int_{-\pi}^{\pi}u^2\left|\tilde{H}_{1,T}\left(u\right)\right|^2du\le const\sum_{t}\left|\tilde{h}\left((t+1)/T\right)-\tilde{h}\left(t/T\right)\right|^2\le const\frac{1}{T}. \tag{4.14}$$

Similar arguments are used also in [38] (see the proof of Theorem 3 and formula (3.21) therein).

Therefore, each term at the right-hand side of (4.11) is bounded by $const{\cdot}T^{-2}$, which gives the asymptotics (4.6) under the condition (i). Analogously we can deduce (4.6) from the expressions (4.8) or (4.9) applying the conditions (ii) or (iii) respectively.

5. Numerical examples

This section illustrates the obtained results and numerically investigates properties of the considered minimum contrast estimators. Numerical results fully support the theoretical findings. First, we consider the family of spatial fractional autoregressive processes introduced in [16]. The simulation studies also suggest that similar results are valid for Gegenbauer random fields, see [22], and the approach is applicable to more general models.

In the following examples we use random field models with spectral densities of the form

$$f(\lambda,\theta)=\frac{\sigma_\varepsilon^2}{(2\pi)^2}\frac{|1-2u_1e^{-i\lambda_1}+e^{-2i\lambda_1}|^{-2d_1}|1-2u_2e^{-i\lambda_2}+e^{-2i\lambda_2}|^{-2d_2}}{|\phi(\exp(-i\lambda_1),\exp(-i\lambda_2),\alpha,\beta)|^2}, \quad (5.1)$$

where $\phi(z_1,z_2,\alpha,\beta)=(1-\alpha z_1)(1-\beta z_2)$.

In Example 1 the values of parameters u_1 and u_2 are known and equal $u_1=u_2=1$. In example 3 $\alpha=\beta=0$ and $\phi(z_1,z_2,0,0)\equiv 1$. Example 2 considers the general case when singularities are located outside of the origin and we suppose again $\alpha=\beta=0$.

We show that in all cases the minimum contrast approach gives very good convergence rates and asymptotic normality.

In all examples below we assume that random fields $Y(\boldsymbol{t})$ are observed on the grid $\{(t_1,t_2):t_1,t_2=1,...,T\}$. Thus we use the observation window from Remark 2.6. Note that in our case $\mathbb{T}=[-\pi,\pi]$.

We chose the taper of the form $h_T(t_1,t_2)=\tilde{h}\,(t_1/T)\,\tilde{h}\,(t_2/T)$, with $\tilde{h}(t)=\frac{1}{2}(1-\cos(2\pi t))$, $t\in[0,1]$, and $w(\boldsymbol{\lambda})=w(\lambda_1,\lambda_2)=|\lambda_1|^{2a_1}|\lambda_2|^{2a_2}$, $a_i>1/2$, $i=1,2$. Then, $w(\boldsymbol{\lambda})$, $\boldsymbol{\lambda}=(\lambda_1,\lambda_2)\in[-\pi,\pi]^2$ is a function satisfying Assumption **B3**. The taper $h(\cdot)$ also satisfies the conditions in Remark 2.6.

5.1. Spatial fractional autoregressive processes

First we start with the family of spatial fractional autoregressive processes introduced in [16].

Let $Y_{\boldsymbol{t}}$, $\boldsymbol{t}=(t_1,t_2)\in\mathbb{Z}^2$, be the spatial process satisfying the following fractional autoregressive model:

$$\phi(B_1,B_2,\alpha,\beta)\nabla_1^{d_1}\nabla_2^{d_2}Y_{t_1,t_2}=\epsilon_{t_1,t_2},\ (t_1,t_2)\in\mathbb{Z}^2, \quad (5.2)$$

where $d_i\in(-1/2,1/2)$, $i=1,2$, $\phi(B_1,B_2,\alpha,\beta)Y_{t_1,t_2}=Y_{t_1,t_2}-\alpha Y_{t_1-1,t_2}-\beta Y_{t_1,t_2-1}+\alpha\beta Y_{t_1-1,t_2-1}$, and $\nabla_1^{d_1}\nabla_2^{d_2}=(1-B_1)^{d_1}(1-B_2)^{d_2}$, with B_i denoting the backward-shift operator for the coordinate t_i, $i=1,2$, i.e., $B_1Y_{t_1,t_2}=Y_{t_1-1,t_2}$, and $B_2Y_{t_1,t_2}=Y_{t_1,t_2-1}$.

The spectral density of the process defined by (5.2) is given by

$$f_Y(\boldsymbol{\lambda}) = \frac{\sigma^2}{4\pi^2}\frac{|1-\exp(-i\lambda_1)|^{-2d_1}|1-\exp(-i\lambda_2)|^{-2d_2}}{|\phi(\exp(-i\lambda_1),\exp(-i\lambda_2),\alpha,\beta)|^2}, \quad \boldsymbol{\lambda}=(\lambda_1,\lambda_2)\in[-\pi,\pi]^2,$$

see [16].

Assume that the values of the parameters α and β are known, or they have been estimated before. Then we are interested to estimate the parameter vector $(d_1, d_2) \in (0, 1/2)^2$. It means that $\boldsymbol{\theta} = (\theta_1, \theta_2) = (d_1, d_2)$ and $\boldsymbol{\theta} \in \Theta = (0, 1/2)^2$.

By direct calculations it is easy to check that all conditions for consistency and asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}_T$ are fulfilled. In particular, for condition **B6** we can use the function $v(\boldsymbol{\lambda}) = v(\lambda_1, \lambda_2) = |\lambda_1|^{2\beta}|\lambda_2|^{2\beta}$, $\beta \in (0, 1/2)$. To check the positive definiteness of matrices $S(\theta)$ and $A(\theta)$ (as required in condition **B8**), we can argue analogously to the very detailed consideration in [22] (see verification of condition **A8** therein). As the matter of fact, the character of singularities in both models, here, and in [22] is the same, with the only difference that here we have singularity of the spectral density at the origin $(0, 0)$, and, in [22], singularity can be shifted from the origin to some another point.

So, we have consistency of the estimator and also asymptotic normality

$$T^{d/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}_q(0, e(h)\mathbf{S}^{-1}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)\mathbf{S}^{-1}(\boldsymbol{\theta}_0)), \tag{5.3}$$

where for the chosen taper we have $e(h) \approx 3.78$. The elements of the matrices $\mathbf{S}(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})$ are calculated in the explicit form, quite similarly to the corresponding calculations in [22], see Appendix B.

Now we provide numerical results for the spatial fractional autoregressive model (5.2). 100 realizations of Y_{t_1,t_2} were generated, for each T in the increasing sample size sequence T=10, 30, 50, 70, 90, 110, 130, 150. The following values of parameters were used in simulations: d_1=0.2, d_2=0.3, α=0.1, β=0.2, and σ_ε^2=1.

The operator $(1-B_1)^{d_1}(1-B_2)^{d_2}$ in (5.2) was expanded in a double power series with respect to each $B_i,\ i = 1, 2$. Realizations of Y were approximated using the truncated sums with powers not exceeding 30 for each B_i. The periodogram I_T^h was computed with $h_T(t_1, t_2)$ and $\omega(\lambda_1, \lambda_2)$ given above.

The minimizing arguments $\widehat{\boldsymbol{\theta}}_T$ of the functional $U_T(\boldsymbol{\theta})$ in (2.6) were found numerically for each simulation.

For each T Table 1 shows estimated values $\hat{d}_1$ and $\hat{d}_2$ of the parameters and the mean square errors (MSE) of $\hat{d}_1$ and $\hat{d}_2$. For each value of T Figure 1 produces a boxplot of estimated values $\hat{d}_1$ and $\hat{d}_2$ based on 100 realizations. It is clear that the estimated values are centered at the true values d_1=0.2 and d_2=0.3. Moreover, Figure 1 and Table 1 demonstrate that $\widehat{\boldsymbol{\theta}}_T = (\hat{d}_1, \hat{d}_2)$ converges to $\boldsymbol{\theta}_0 = (0.2, 0.3)$ as T increases.

Normal Q-Q plots of $\hat{d}_1$ and $\hat{d}_2$ were built for large values of T to verify asymptotic normality of the estimators. Figure 2 demonstrates that the empirical distribution of $\hat{d}_1$ and $\hat{d}_2$ match the theoretical normal distribution.

TABLE 1
Mean, standard deviation and MSE of $\hat{d}_1$ and $\hat{d}_2$.

T	$\hat{d}_1$	$\hat{d}_2$	MSE of $\hat{d}_1$	MSE of $\hat{d}_2$
10	0.227 (0.122)	0.403 (0.141)	0.02	0. 02
30	0.219 (0.054)	0.361 (0.072)	0.003	0.01
50	0.206 (0.034)	0.346 (0.054)	0.0012	0.03
70	0.199 (0.027)	0.326 (0.034)	0.001	0.0012
90	0.205 (0.029)	0.314 (0.025)	0.001	0.001
110	0.199 (0.026)	0.311 (0.022)	0.001	0.001
130	0.197 (0.027)	0.310 (0.025)	0.001	0.001
150	0.199 (0.022)	0.307 (0.022)	0.0005	0.001

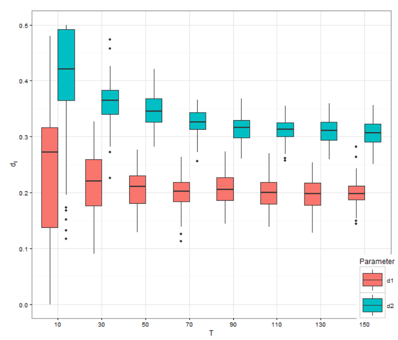

Fig 1: Boxplots of sampled values of $\hat{d}_1$ and $\hat{d}_2$.

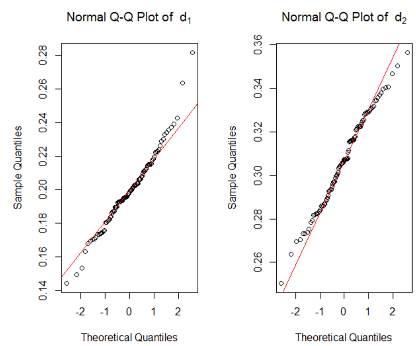

Fig 2: Normal Q-Q plots of $\hat{d}_1$ and $\hat{d}_2$.

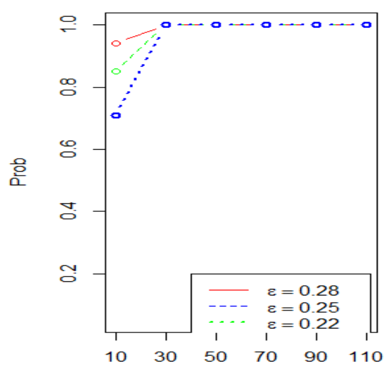

Fig 3: Sample probabilities $P_0(|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$.

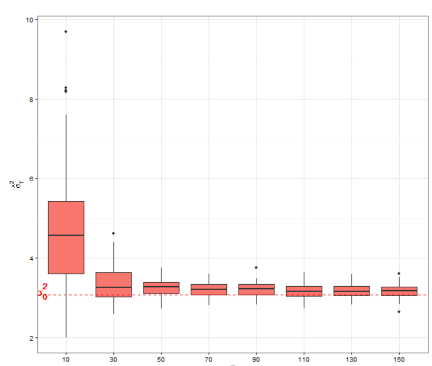

Fig 4: Boxplots of sampled values of $\hat{\sigma}_T^2$.

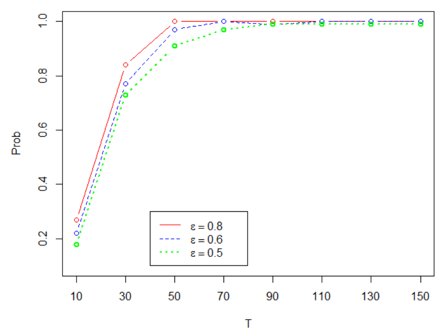

Fig 5: Sampled probabilities $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon)$.

For each T 100 simulated values were used to estimate $P_0(|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$. The plot of the sample probabilities $P_0(|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$ in Figure 3 confirms convergence of $\widehat{\boldsymbol{\theta}}_T = (\hat{d}_1, \hat{d}_2)$ to $\boldsymbol{\theta}_0 = (0.2, 0.3)$ in probability.

Finally, we investigated asymptotic properties of $\hat{\sigma}_T$. Similarly to the above cases, for each T we built boxplots and plotted sample probabilities for 100 simulations. Figures 4 and 5 support convergence in probability $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon) \to 1$, when T increases.

5.2. *Spatial Gegenbauer random fields: singularities at non-zero locations*

This section illustrates that the proposed minimum contrast estimation technique works even for the case of Gegenbauer random fields, see [22].

Let Y_{t_1,t_2} be a discrete random field satisfying

$$\Delta_{u_1}^{d_1} \Delta_{u_2}^{d_2} Y_{t_1,t_2} = (I - 2u_1 B_1 + B_1^2)^{d_1} (I - 2u_2 B_2 + B_2^2)^{d_2} Y_{t_1,t_2} = \epsilon_{t_1,t_2},$$

where ϵ_{t_1,t_2}, $(t_1, t_2) \in \mathbb{Z}^2$, is a zero-mean white noise field with the common variance $E[\epsilon_{t_1,t_2}^2] = \sigma_\epsilon^2$, where, for $j = 1, 2$,

$$\begin{aligned} \Delta_{u_j}^{d_j} &= (I - 2u_j B_j + B_j^2)^{d_j} \\ &= (1 - 2\cos\nu_j B_j + B_j^2)^{d_j} \\ &= [(1 - \exp(i\nu_j) B_j)(1 - \exp(-i\nu_j) B_j)]^{d_j}. \end{aligned}$$

As before B_j, $j = 1, 2$, denotes the backward-shift operator for each spatial coordinate and $u = \cos\nu$, i.e. $\nu = arccos(u)$.

There exists the following representation of a spatial stationary Gegenbauer random field, see [22],

$$Y_{t_1,t_2} = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} C_{n_1}^{(d_1)}(u_1) C_{n_2}^{(d_2)}(u_2) \epsilon_{t_1-n_1,t_2-n_2}, \tag{5.4}$$

where $d_i \neq 0$, and $|u_i| \leq 1$, $i = 1, 2$.

The Gegenbauer polynomial $C_n^{(d)}(u)$ is given by

$$C_n^{(d)}(u) = \sum_{k=0}^{[n/2]} (-1)^k \frac{(2u)^{n-2k}\Gamma(d-k+n)}{k!(n-2k)!\Gamma(d)}.$$

The spectral density f of the spatial Gegenbauer random field is defined as

$$\begin{aligned} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) &= \frac{\sigma_\epsilon^2}{(2\pi)^2} |1 - 2u_1 e^{-i\lambda_1} + e^{-2i\lambda_1}|^{-2d_1} |1 - 2u_2 e^{-i\lambda_2} + e^{-2i\lambda_2}|^{-2d_2} \\ &= \frac{\sigma_\epsilon^2}{(2\pi)^2} \left|2\cos(\lambda_1) - 2u_1\right|^{-2d_1} \left|2\cos(\lambda_2) - 2u_2\right|^{-2d_2}. \end{aligned}$$

We generated 200 replications of random field given in (5.4), for the increasing sequence of sample sizes T=10, 30, 50, 70, 90, 110, using the parameter values d_1=0.2, d_2 =0.3, u_1=0.4, u_2=0.3, and σ_ϵ^2=1.

Analogously to the example in section 5.1 realizations of Y_{t_1,t_2} were approximated by the truncated sums with 100 terms in (5.4). The periodogram I_T^h was computed with $h_T(t_1, t_2)$=$\tilde{h}(t_1/T)\tilde{h}(t_2/T)$, $\tilde{h}(t) = \frac{1}{2}(1 - cos(2\pi t))$, $t \in [0, 1]$. We used the weight function $\omega(\lambda_1, \lambda_2) = |2\cos(\lambda_1) - 2u_1|^2 |2\cos(\lambda_2) - 2u_2|^2$.

The analysis, plots and explanations below are analogous to the example in section 5.1. The minimizing arguments $\widehat{\boldsymbol{\theta}}_T$ of the functional $U_T(\boldsymbol{\theta})$ in (2.6) were found numerically for each simulation. Figure 6 and Table 2 demonstrate that $\widehat{\boldsymbol{\theta}}_T$ converges to $\boldsymbol{\theta}_0$ as T increases. The plot of the sample probabilities $P_0(|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$ in Figure 8 also confirms convergence in probability. The normal Q-Q plots of $\hat{d}_1$ and $\hat{d}_2$ in Figure 7 match with the theoretical normal distribution. Table 2 shows MSEs of $\hat{d}_1$ and $\hat{d}_2$, computed for different T values. Figures 10 and 9 support convergence in probability $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon) \to 1$ when T increases.

Similar results were also obtained using the weight function $\omega(\lambda_1, \lambda_2) = (|\lambda_1| - \arccos(u_1))^2 (|\lambda_2| - \arccos(u_2))^2$.

Table 2
Mean, standard deviation and MSE of $\hat{d}_1$ and $\hat{d}_2$.

T	$\hat{d}_1$	$\hat{d}_2$	MSE of $\hat{d}_1$	MSE of $\hat{d}_2$
10	0.264 (0.186)	0.334 (0.181)	0.04	0. 035
30	0.220 (0.115)	0.346 (0.130)	0.013	0.02
50	0.208 (0.078)	0.314 (0.085)	0.01	0.01
70	0.205 (0.063)	0.313 (0.085)	0.004	0.005
90	0.204 (0.063)	0.298 (0.064)	0.003	0.003
110	0.193 (0.059)	0.304 (0.060)	0.003	0.003

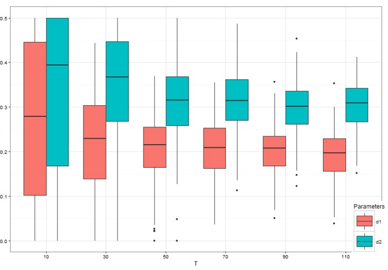

Fig 6: Boxplots of sampled values of $\hat{d}_1$ and $\hat{d}_2$.

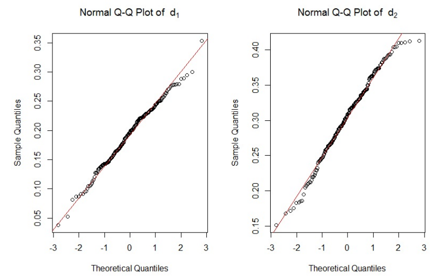

Fig 7: The Normal Q-Q plots of $\hat{d}_1$

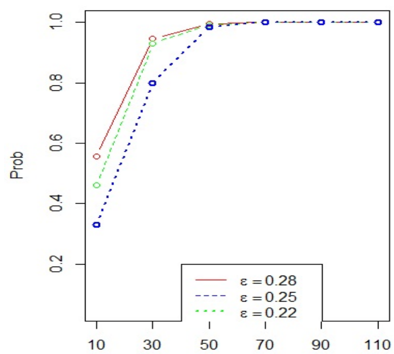

Fig 8: Sample probabilities $P_0(|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$.

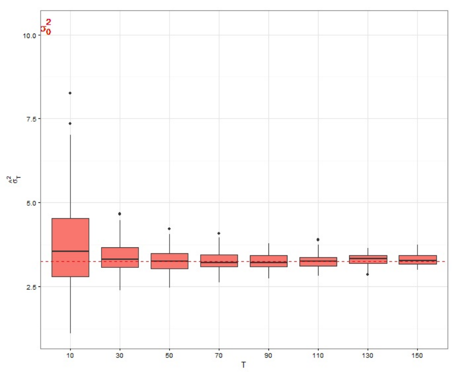

Fig 9: Boxplots of sampled values of $\hat{\sigma}_T^2$.

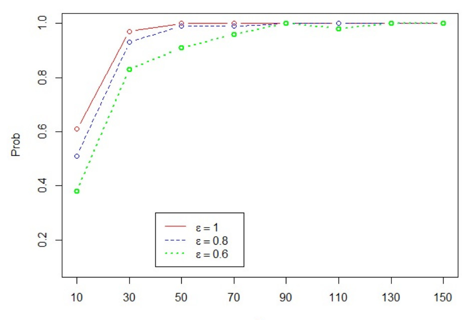

Fig 10: Sampled probabilities $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon)$.

5.3. Spatial Gegenbauer random fields: singularities at the origin

Finally in the third example we consider the case of spectral densities with singularities at the origin and the degenerated $\phi(B_1, B_2, \alpha, \beta) = I$, i.e. $\alpha = \beta = 0$. The method shows very good performance in this case as well.

In this example we also used random fields Y satisfying (5.4). Their realisations were simulated for the parameter values u_i=1, $i = 1, 2$, σ_ϵ^2=1, $d_1 = 0.2$ and $d_2 = 0.3$. Realizations of Y were simulated using the truncated sums with 100 terms in (5.4). The periodogram I_T^h was computed with $h_T(t_1, t_2)$=$\tilde{h}(t_1/T)\tilde{h}(t_2/T)$, $\tilde{h}(t) = \frac{1}{2}(1 - cos(2\pi t))$, $t \in [0, 1]$. We used the weight function $\omega(\lambda_1, \lambda_2) = (|\lambda_1| - \arccos(u_1))^2(|\lambda_2| - \arccos(u_2))^2 = |\lambda_1|^2|\lambda_2|^2$. Similar results were also obtained for $\omega(\lambda_1, \lambda_2) = |2\cos(\lambda_1) - 2u_1|^2|2\cos(\lambda_2) - 2u_2|^2$.

The analysis, plots and explanations below are analogous to the example in section 5.1. The minimizing arguments $\widehat{\boldsymbol{\theta}}_T$ of functional $U_T(\boldsymbol{\theta})$ in (2.6) were found numerically for each simulation. Figure 11 and Table 3 show that $\hat{\boldsymbol{\theta}}_T$ converges to $\boldsymbol{\theta}_T$ as T increases.

TABLE 3
Mean, standard deviation and MSE of $\hat{d}_1$ and $\hat{d}_2$.

T	$\hat{d}_1$	$\hat{d}_2$	MSE of $\hat{d}_1$	MSE of $\hat{d}_2$
10	0.284 (0.211)	0.361 (0.184)	0.04	0. 034
30	0.219 (0.141)	0.356 (0.130)	0.02	0.02
50	0.218 (0.105)	0.323 (0.102)	0.01	0.01
70	0.203 (0.092)	0.301 (0.098)	0.01	0.01
90	0.206 (0.091)	0.304 (0.089)	0.01	0.01
110	0.190 (0.096)	0.299 (0.082)	0.01	0.01

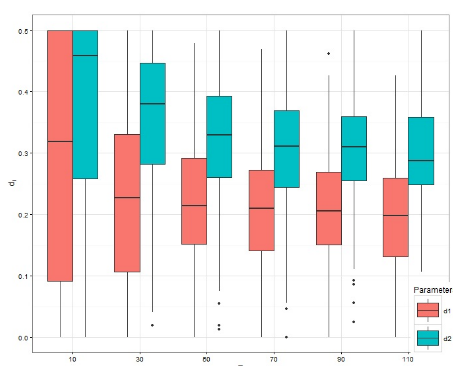

Fig 11: Boxplots of sampled values of $\hat{d}_1$ and $\hat{d}_2$ with $u_i = 0$, $i = 1, 2$.

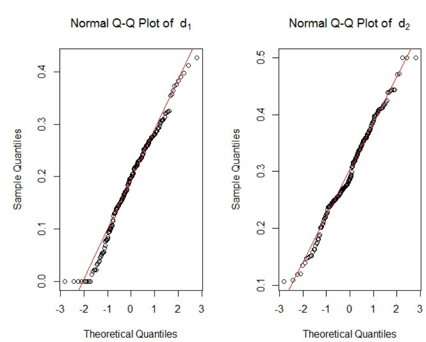

Fig 12: Normal Q-Q plots of $\hat{d}_1$ and $\hat{d}_2$ with $u_i = 0$, $i = 1, 2$.

The Normal Q-Q plots of $\hat{d}_1$ and $\hat{d}_2$ in Figure 12 match with the theoretical normal distribution. The plot of the sample probabilities $P_0(|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$ in Figure 13 also confirms convergence in probability. Table 3 shows MSEs of $\hat{d}_1$

and $\hat{d}_2$, computed for different T values. Figures 14 and 15 support convergence in probability $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon) \to 1$, when T increases.

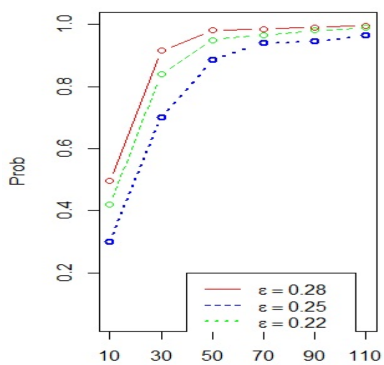

Fig 13: Sample probabilities $P_0(|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$.

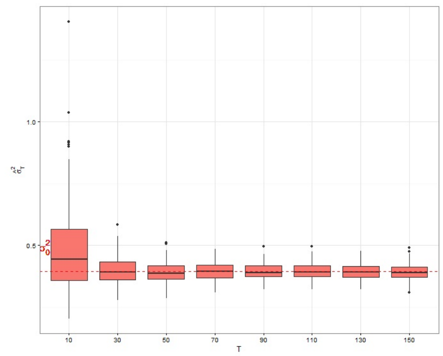

Fig 14: Boxplots of sampled values of $\hat{\sigma}_T^2$.

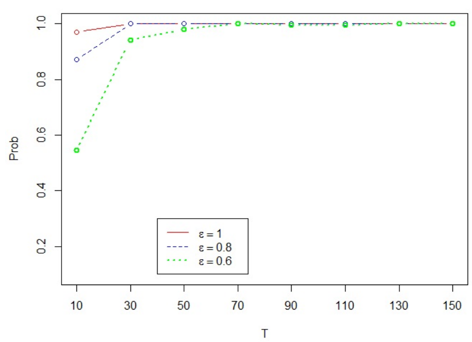

Fig 15: Sampled probabilities $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon)$.

Appendix A: Central limit theorems for spectral functionals (or quadratics forms) of Gaussian random fields, from tapered data

One of the classical approaches to derive central limit theorem for spectral functionals $J_T(\varphi) = \int_{\mathbb{T}} I_T(\boldsymbol{\lambda}) \varphi(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda}$ consists of calculating and evaluating their cumulants. We present here some details of this approach, and the corresponding results for the case of Gaussian fields, and tapered data, as needed for consideration in the present paper. To state CLT for the normalized functional $\tilde{J}_T(\varphi) = T^{d/2}(J_T(\varphi) - EJ_T(\varphi))$ it is enough to provide conditions for convergence to the finite limit of the second order cumulant of $\tilde{J}_T(\varphi)$, and for convergence to zero of all cumulants of higher orders.

Note. Throughout this section we will omit the superscript 'h' and write simply $I_T(\boldsymbol{\lambda})$ to denote the tapered periodogram and $J_T(\varphi)$ to denote the corresponding spectral functional. Only Proposition A.1. here below concerns the non-tapered case.

The cumulant of the general k-th order can be represented in the form:

$$\begin{aligned} c_k(\tilde{J}_T(\varphi)) &= T^{kd/2}2^{k-1}(k-1)![H_{2,T}(\mathbf{0})]^{-k}\int_{\boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_{2k}\in\mathbb{T}^{2k}} f(\boldsymbol{\lambda}_1)\varphi(\boldsymbol{\lambda}_2)f(\boldsymbol{\lambda}_3)...\varphi(\boldsymbol{\lambda}_{2k}) \\ &\quad \times H_{1,T}(\boldsymbol{\lambda}_2-\boldsymbol{\lambda}_1)H_{1,T}(\boldsymbol{\lambda}_3-\boldsymbol{\lambda}_2)... \\ &\quad \times H_{1,T}(\boldsymbol{\lambda}_{2k}-\boldsymbol{\lambda}_{2k-1})H_{1,T}(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_{2k})d\boldsymbol{\lambda}_1\ldots d\boldsymbol{\lambda}_{2k} \end{aligned}$$

$$\begin{aligned} &= T^{kd/2}2^{k-1}(k-1)![H_{2,T}(\mathbf{0})]^{-k}\int_{\mathbf{u}_1,\ldots,\mathbf{u}_{2k-1}\in\mathbb{T}^{2k-1}}\int_{\boldsymbol{\lambda}\in\mathbb{T}} f(\boldsymbol{\lambda})\varphi(\boldsymbol{\lambda}+\mathbf{u}_1) \\ &\quad \times f(\boldsymbol{\lambda}+\mathbf{u}_1+\mathbf{u}_2)\varphi(\boldsymbol{\lambda}+\mathbf{u}_1+\mathbf{u}_2+\mathbf{u}_3)\ldots\varphi\left(\boldsymbol{\lambda}+\sum_{i=1}^{2k-1}\mathbf{u}_i\right) \\ &\quad \times\prod_{i=1}^{2k-1}H_{1,T}(\mathbf{u}_i)H_{1,T}\left(-\sum_{i=1}^{2k-1}\mathbf{u}_i\right)d\boldsymbol{\lambda}d\mathbf{u}_1\ldots d\mathbf{u}_{2k-1}, \end{aligned}$$

where $H_{k,T}(\boldsymbol{\lambda})=\sum_{\mathbf{t}\in L_T}h_T^k(\mathbf{t})e^{-i(\boldsymbol{\lambda},\mathbf{t})}$, and $h_T(\mathbf{t})=h(\mathbf{t}/T)$. Details of calculations of the cumulants of spectral functionals can be found, for example, in [2], [10], [14] for the nontapered case, and in [3], [19], [20] for the tapered case. The calculations are based on the so-called product formula for cumulants which gives the expression for cumulants of products of random variables in terms of cumulants of the individual variables, the mentioned formula reduces to a particular simple form in the Gaussian case.

Note that the functions

$$\Phi_{k,T}^h(\boldsymbol{\lambda}_1,...,\boldsymbol{\lambda}_{k-1}):=\frac{1}{(2\pi)^{d(k-1)}H_{k,T}(\mathbf{0})}\prod_{j=1}^{k-1}H_{1,T}(\boldsymbol{\lambda}_j)H_{1,T}\left(-\sum_{j=1}^{k-1}\boldsymbol{\lambda}_j\right) \tag{A.1}$$

are multidimensional kernels of Fejér type over $\mathbb{T}^{k-1}$, or approximate identities for convolution.

In the non-tapered case, when $h(t)\equiv 1$, this was shown in [13, 14]. In the case under consideration, when the taper factorizes (as defined in Section 2.2), and the domain of observation is a cube $L_T=[-T,T]^d$, this fact follows as a straightforward generalization of the corresponding result by [19] for dimension $d=1$.

The kernel property of $\Phi^h_{k,T}(\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_{k-1})$ implies

$$\lim_{T\to\infty} \int\limits_{\mathbb{T}^{k-1}} G\left(\mathbf{u}_1 - \mathbf{v}_1, ..., \mathbf{u}_{k-1} - \mathbf{v}_{k-1}\right) \Phi^h_{k,T}\left(\mathbf{u}_1, ..., \mathbf{u}_{k-1}\right) d\mathbf{u}_1 ... d\mathbf{u}_{k-1}$$
$$= G\left(\mathbf{v}_1, ..., \mathbf{v}_{k-1}\right), \tag{A.2}$$

provided that the function G is bounded and continuous at the point $(\mathbf{v}_1, ..., \mathbf{v}_{k-1})$.

We have, in particular,

$$Var(\tilde{J}_T(\varphi)) = 2T^d (2\pi)^{3d} H_{4,T}(\mathbf{0})[H_{2,T}(\mathbf{0})]^{-2} \int_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3 \in \mathbb{T}^3} \int_{\boldsymbol{\lambda}\in\mathbb{T}} f(\boldsymbol{\lambda})\varphi(\boldsymbol{\lambda} + \mathbf{u}_1) \tag{A.3}$$
$$\times f(\boldsymbol{\lambda} + \mathbf{u}_1 + \mathbf{u}_2)\varphi(\boldsymbol{\lambda} + \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3)\, \Phi^h_{4,T}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)\, d\boldsymbol{\lambda} d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3.$$

$$= \quad 2T^d (2\pi)^{3d} H_{4,T}(\mathbf{0})[H_{2,T}(\mathbf{0})]^{-2} \int_{(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\boldsymbol{\lambda}_3,\boldsymbol{\lambda}_4)\in\mathbb{T}^4} f(\boldsymbol{\lambda}_1)\varphi(\boldsymbol{\lambda}_2) f(\boldsymbol{\lambda}_3)\varphi(\boldsymbol{\lambda}_4)$$
$$\times \Phi^h_{4,T}(\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_3 - \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_4 - \boldsymbol{\lambda}_3) d\boldsymbol{\lambda}_1 d\boldsymbol{\lambda}_2 d\boldsymbol{\lambda}_3 d\boldsymbol{\lambda}_4.$$

Asymptotic analysis of expressions for cumulants, based on the property (A.2), allows to state the following asymptotic normality result for the functional $J_T(\varphi) = J^h_T(\varphi) = \int_{\mathbb{T}} I^h_T(\boldsymbol{\lambda})\, \varphi(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}$ in the case of tapered data.

Theorem A.1. *Let $X(\mathbf{t})$, $\mathbf{t} \in Z^d$, be a zero-mean Gaussian random field with spectral density $f(\boldsymbol{\lambda}) \in L_p$ and $\varphi(\boldsymbol{\lambda}) \in L_q$, where $\frac{1}{p} + \frac{1}{q} \le \frac{1}{2}$. Then*

$$T^{d/2}(J_T(\varphi) - EJ_T(\varphi)) \xrightarrow{D} N(0, \sigma^2) \quad \text{as } T \to \infty, \tag{A.4}$$

where

$$\sigma^2 = 2(2\pi)^{3d} e(h) \int_{\mathbb{T}} f^2(\boldsymbol{\lambda})\varphi^2(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \tag{A.5}$$

where $e(h)$ is defined in (2.11).

The convergence of the variance (A.3) to σ^2 can be obtained by the following arguments:

(i) Under the integrability conditions imposed in the theorem on f and φ, the inner integral over $\boldsymbol{\lambda}$ in expression (A.3) is a bounded and continuous function of $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, say $G(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, therefore, the overall integral in (A.3) can be written as $\int G(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)\Phi^h_{4,T}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) d\mathbf{u}_1 d\mathbf{u}_3 d\mathbf{u}_3$; then, due to the kernel property (A.2), the limit of this integral is equal to $G(\mathbf{0}, \mathbf{0}, \mathbf{0}) = \int_{\mathbb{T}} f^2(\boldsymbol{\lambda})\varphi^2(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$;

(ii) The normalizing factor $T^d H_{4,T}(\mathbf{0})[H_{2,T}(\mathbf{0})]^{-2}$ converges to $e(h)$ due to the asymptotic behavior $H_{k,T}(0) \sim T^d \left(\int \tilde{h}^k(t) dt\right)^d$ as $T \to \infty$.

Following the same arguments as in [10, 11], the integrability conditions on f and φ imply also convergence to zero of all cumulants of orders $k \geq 3$. Note that, under the assumptions considered on the taper function, the following estimates for the norms of $H_{k,T}(\boldsymbol{\lambda})$ hold: $\|H_{k,T}(\boldsymbol{\lambda})\|_p \leq CT^{d(1-\frac{1}{p})}$, $p > 1$, which is a consequence of Lemma 1 by [19], and which can be used on the place of analogous estimates in the non-tapered case. Therefore all the proofs can be preserved.

From Theorem A.1 and Lemma 4.1 (see Section 4), we can deduce the following corollary.

Corollary 1. *Let the conditions of Theorem A.1 and Lemma 4.1 be satisfied. Then*

$$T^{d/2}(J_T(\varphi) - J(\varphi)) \xrightarrow{D} N(0, \sigma^2) \quad \text{as } T \to \infty, \tag{A.6}$$

where σ^2 is given by (A.5).

Note that the recent paper by [25] (Section 4.2.1) presents a collection of classical sufficient conditions for CLT, for quadratic forms $Q_T = T^d J_T(\varphi)$, in the case of Gaussian processes in discrete and continuous time.

Theorem A.1 is a generalization of the statements (C) of Theorem 4.7 in [25] (see also [9]) to the case of fields (in Z^d) and for tapered data.

For the case of Gaussian processes with discrete time, and spectral densities with possible singularities, the following results were obtained in [24] (without tapering, that is, $h(t) \equiv 1$).

Proposition A.1. *Let $X(t)$, $t \in Z^1$, be a zero-mean Gaussian process with spectral density $f(\lambda)$.*

I. *Suppose that the following conditions hold:*

(i) The sets of discontinuities of functions $f(\lambda)$ and $\varphi(\lambda)$ have Lebesgue measure zero, and these functions are bounded on the interval $[\delta, \pi]$ for all $\delta > 0$;

(ii) There exist $\alpha < 1$ and $\beta < 1$ such that for $\alpha + \beta < 1/2$, and for each $\delta > 0$, $f(\lambda) = O(|\lambda|^{-\alpha-\delta})$ and $\varphi(\lambda) = O(|\lambda|^{-\beta-\delta})$ as $\lambda \to 0$.

Then

$$T^{1/2}(J_T(\varphi) - EJ_T(\varphi)) \xrightarrow{D} N(0, \sigma^2) \quad \text{as } T \to \infty, \tag{A.7}$$

where $\sigma^2 = 16\pi^3 \int_{-\pi}^{\pi} \varphi^2(\lambda) f^2(\lambda) d\lambda$.

II. *Assumption (ii) holds, if there exist $\alpha < 1$ and $\beta < 1$ such that $\alpha + \beta < 1/2$, and $f(\lambda) = |\lambda|^{-\alpha} L_1(\lambda)$ and $\varphi(\lambda) = |\lambda|^{-\beta} L_2(\lambda)$ as $\lambda \to 0$, where L_1 and L_2 are slowly varying functions at zero.*

We present here an extension of Proposition A.1. for the case of Gaussian fields and tapered periodogram in the following form.

Theorem A.2. *Let $X(\mathbf{t})$, $\mathbf{t} \in Z^d$, be a zero-mean Gaussian random field with spectral density $f(\boldsymbol{\lambda})$ such that for some $0 < \alpha_i < 1$, $i = 1, .., d$, $f(\boldsymbol{\lambda}) = O(\prod_{i=1}^d |\lambda_i|^{-\alpha_i})$ as $\lambda_i \to 0$, and $\varphi(\boldsymbol{\lambda}) = O(\prod_{i=1}^d |\lambda_i|^{\alpha_i})$ as $\lambda_i \to 0$. The sets of discontinuities of functions $f(\boldsymbol{\lambda})$ and $\varphi(\boldsymbol{\lambda})$ have Lebesgue measure zero, and these functions are bounded for $\delta \leq |\boldsymbol{\lambda}| \leq \pi$ for all $\delta > 0$. Then*

$$T^{d/2}(J_T(\varphi) - EJ_T(\varphi)) \xrightarrow{D} N(0, \sigma^2) \quad \text{as } T \to \infty, \tag{A.8}$$

where σ^2 is the same as in Theorem A.1

From Theorem A.2 and Lemma 4.1 (see Section 4) we can deduce the following corollary.

Corollary 2. *Let the conditions of Theorem A.2 and Lemma 4.1 be satisfied. Then*

$$T^{d/2}(J_T(\varphi) - J(\varphi)) \overset{D}{\to} N(0, \sigma^2) \quad \text{as } T \to \infty, \tag{A.9}$$

where σ^2 is given by (A.5).

Remark A.1. *Proposition A.1 is stated for the functions $f(\lambda)$ and $\varphi(\lambda)$ which have singularities at the point $\lambda = 0$. However, it can be shown (see, [25]) that the choice $\lambda = 0$ is not essential and Proposition A.1 holds if the singularity is located at any other point $\lambda_0 \in [-\pi, \pi]$. The analogous observation is true for the result stated for the case of random fields (with multiplicative form of singularities) in Theorem A.2. In fact, Theorem A.2 remains valid if, instead of the origin ($\lambda_i = 0$, $i = 1, ..., d$), the singularity of the form prescribed in the theorem takes place at any other point.*

Proof of Theorem A.2 For the proof we use an idea from the paper [29]. Consider firstly the case $d = 1$. Introduce the filtered process

$$Y(t) = \nabla^{\alpha/2} X(t),$$

where $\nabla = 1 - B$, B is the backward shift operator ($BX(t) = X(t-1)$), and $\nabla^{\alpha/2} = (1-B)^{\alpha/2} := \sum_{j=0}^{\infty} C_j^{\alpha/2}(-B)^j$. Then the process $Y(t)$ has the spectral density $f_Y(\lambda) = (2\sin|\frac{\lambda}{2}|)^{\alpha} f_X(\lambda)$, since $Y(t)$ is obtained from $X(t)$ using the filter with transfer function $D(i\lambda) = (1 - e^{i\lambda})^{\alpha/2}$ and $|D(i\lambda)|^2 = (2\sin|\frac{1}{2}\lambda|)^{\alpha}$.

Let $\psi(\lambda) = \varphi(\lambda)/(2\sin|\frac{1}{2}\lambda|)^d$ and consider the functional

$$\tilde{J}_T^Y(\psi) = \int_{-\pi}^{\pi} \psi(\lambda) I_T^Y(\lambda) d\lambda - E\int_{-\pi}^{\pi} \psi(\lambda) I_T^Y(\lambda) d\lambda,$$

where $I_T^Y(\lambda) = \frac{1}{2\pi T}|\sum_{t\in L_T} h_T(t)e^{i\lambda t}Y(t)|^2$ is the tapered periodogram which corresponds to $\{Y(t), t \in L_T\}$.

Since spectral density $f_Y(\lambda)$ of the process $Y(t)$ and function $\psi(\lambda)$ satisfy conditions of Theorem A.1, for the functional $\tilde{J}_T^Y(\psi)$ we have the convergence as $T \to \infty$

$$T^{1/2}\tilde{J}_T^Y(\psi) \xrightarrow{D} N(0, \sigma^2), \tag{A.10}$$

where

$$\sigma^2 = 16\pi^3 e(h)\int_{-\pi}^{\pi} \psi^2(\lambda)f_Y^2(\lambda)d\lambda = 16\pi^3 e(h)\int_{-\pi}^{\pi} \varphi^2(\lambda)f_X^2(\lambda)d\lambda. \tag{A.11}$$

Therefore, in order to prove the statement of the theorem, it is sufficient to show that

$$\lim_{T\to\infty} TE|\tilde{J}_T^X(\varphi) - \tilde{J}_T^Y(\psi)|^2 = 0, \tag{A.12}$$

where

$$\tilde{J}_T^X(\varphi) = \int_{-\pi}^{\pi} \varphi(\lambda)I_T^X(\lambda)d\lambda - E\int_{-\pi}^{\pi} \varphi(\lambda)I_T^X(\lambda)d\lambda.$$

Consider

$$TE|\tilde{J}_T^X(\varphi) - \tilde{J}_T^Y(\psi)|^2 = TE|\tilde{J}_T^X(\varphi)|^2 + TE|\tilde{J}_T^Y(\psi)|^2 - 2TE\tilde{J}_T^X(\varphi)\tilde{J}_T^Y(\psi). \tag{A.13}$$

For the functional which corresponds to the process $Y(t)$ we readily have the convergence $TE|\tilde{J}_T^Y(\psi)|^2 \to \sigma^2$ as $T \to \infty$.

We will show that under the conditions of the theorem $TE|\tilde{J}_T^X(\varphi)|^2$ and $TE\tilde{J}_T^X(\varphi)\tilde{J}_T^Y(\psi)$ also tend to σ^2 as $T \to \infty$, and, therefore convergence (A.12) holds.

Using (A.3) we can write:

$$\begin{aligned} TE|\tilde{J}_T^X(\varphi)|^2 &= 2T^d (2\pi)^{3d} H_{4,T}(0)[H_{2,T}(0)]^{-2} \int_{[-\pi,\pi]^4} f_X(\lambda_1)\varphi(\lambda_2)f_X(\lambda_3)\varphi(\lambda_4) \\ &\quad \times\Phi_{4,T}^h(\lambda_2-\lambda_1, \lambda_3-\lambda_2, \lambda_4-\lambda_3)d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4. \end{aligned} \tag{A.14}$$

$$\begin{aligned} TE|\tilde{J}_T^Y(\psi)|^2 &= 2T^d (2\pi)^{3d} H_{4,T}(0)[H_{2,T}(0)]^{-2} \int_{[-\pi,\pi]^4} f_Y(\lambda_1)\psi(\lambda_2)f_Y(\lambda_3)\psi(\lambda_4) \\ &\quad \times\Phi_{4,T}^h(\lambda_2-\lambda_1, \lambda_3-\lambda_2, \lambda_4-\lambda_3)d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4. \end{aligned} \tag{A.15}$$

$$\begin{aligned} TE|\tilde{J}_T^X(\varphi)\tilde{J}_T^Y(\psi)| &= 2T^d (2\pi)^{3d} H_{4,T}(0)[H_{2,T}(0)]^{-2} \\ &\quad \times\int_{[-\pi,\pi]^4} f_{XY}(\lambda_1)\varphi(\lambda_2)f_{XY}(\lambda_3)\psi(\lambda_4) \\ &\quad \times\Phi_{4,T}^h(\lambda_2-\lambda_1, \lambda_3-\lambda_2, \lambda_4-\lambda_3)d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4, \end{aligned} \tag{A.16}$$

where by f_{XY} we have denoted cross spectral density of processes $X(t)$ and $Y(t)$.

Since $Y(t)$ is obtained from $X(t)$ with the use of filter with transfer function $D(i\lambda)$, then $f_{XY}(\lambda) = D(-i\lambda)f_X(\lambda)$, or $f_{XY}(\lambda) = D(-i\lambda)\frac{f_Y(\lambda)}{|D(i\lambda)|^2}$, recall also that $\varphi(\lambda) = \psi(\lambda)|D(i\lambda)|^2$, $f_X(\lambda) = \frac{f_Y(\lambda)}{|D(i\lambda)|^2}$.

If we define the measure μ_T on $[-\pi,\pi]^4$ by

$$\mu_T(E) = \int_E \Phi^h_{4,T}(\lambda_2-\lambda_1, \lambda_3-\lambda_2, \lambda_4-\lambda_3)d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4, \text{ for } E \subset [-\pi,\pi]^4$$

then μ_T converges weakly to the measure μ which is concentrated on the diagonal $D = \{y_1 = y_2 = y_3 = y_4\}$ and satisfies: $\mu(\{y : a \leq y_1 = y_2 = y_3 = y_4 \leq b\}) = b - a$. For non-tapered case this was shown in [24]. The result holds for the tapered case as well under the conditions on taper introduced. This can be shown by the same arguments as in [24], namely, by considering the Fourier coefficients of measures μ and μ_T. Convergence of Fourier coefficients of μ_T to those of μ can be deduced from the kernel property (A.2).

Next we note that the following estimate holds for $H_{1,T}$: $|H_{1,T}(\lambda)| \leq const \cdot l_T(\lambda)$, where $l_T(u)$ denotes 2π-periodic extension of the function $l^*_T(u)$, which is defined as: $l^*_T(u) = T$ for $|u| \leq \frac{1}{T}$, and $l^*_T(u) = \frac{1}{|u|}$ for $\frac{1}{T} < |u| \leq \pi$. From this point convergence of (A.14), (A.15), (A.16) to σ^2 defined by (A.11) can be obtained following the corresponding lines of the proof in [24], more precisely, we do not need their arguments in full generality, but just use those arguments for the second order cumulant, and parts a) of their Propositions 6.1 and 6.2 work in our case.

The proof for the case $d = 1$ can be directly extended for $d > 1$ in the case under consideration, when the singularities of spectral density factorize as described in the formulation of theorem, and for the taper which factorizes, so that the integrals can be split as d-tuple of integrals, which appear when $d = 1$, and, therefore, corresponding reasonings can be preserved. Note that the filtered field is introduced as

$$\begin{aligned} Y(\boldsymbol{t}) = Y(t_1,...,t_d) &= \nabla_1^{\alpha_1/2}...\nabla_d^{\alpha_d/2}X(\boldsymbol{t}) \\ &= \sum_{k_1=0}^{\infty}...\sum_{k_d=0}^{\infty}\prod_{i=1}^{d} C_{k_i}^{\alpha_i/2}X(t_1-k_1,...,t_d-k_d), \end{aligned}$$

and has the spectral density

$$f_Y(\lambda_1,...,\lambda_d) = \left(\prod_{i=1}^{d}\left|2\sin\frac{\lambda}{2}\right|^{\alpha_i}\right) f_X(\lambda_1,...,\lambda_d).$$

Appendix B: Expression for the elements of $\mathbf{S}(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})$

To write down the expression for the elements of the matrices $\mathbf{S}(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})$, we will use the following derivatives:

$$\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta}) = -2\log\left|2\sin\frac{\lambda_i}{2}\right| f(\boldsymbol{\lambda},\boldsymbol{\theta}),$$

$$\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta}) = -2\int_{[-\pi,\pi]^2} \log\left|2\sin\frac{\lambda_i}{2}\right| w(\boldsymbol{\lambda})f(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda}.$$

We obtain, see pages 671 and 672 in [22],

$$\begin{aligned}
s_{ij} &= 3\int_{[-\pi,\pi]^2} \frac{w(\boldsymbol{\lambda})}{\sigma^2(\boldsymbol{\theta})}\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]d\boldsymbol{\lambda} \\
&\quad - \int_{[-\pi,\pi]^2} \frac{w(\boldsymbol{\lambda})}{f(\boldsymbol{\lambda},\boldsymbol{\theta})}\left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]d\boldsymbol{\lambda} \\
&= \frac{3}{\sigma^2(\boldsymbol{\theta})}\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right] \\
&\quad -4\int_{[-\pi,\pi]^2} \log\left|2\sin\frac{\lambda_i}{2}\right|\log\left|2\sin\frac{\lambda_j}{2}\right| w(\boldsymbol{\lambda})f(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda},
\end{aligned}$$

and

$$\begin{aligned}
a_{ij} &= 8\pi^2\sigma^4(\boldsymbol{\theta})\int_{[-\pi,\pi]^2} w^2(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}\psi(\boldsymbol{\lambda},\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\psi(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda} \\
&= 8\pi^2\int_{[-\pi,\pi]^2} w^2(\boldsymbol{\lambda})\left(\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})\right) \\
&\quad \times\left(\left[\frac{\partial}{\partial\theta_j}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})\right)d\boldsymbol{\lambda}.
\end{aligned}$$

Therefore, we can write $a_{ij} = S_1 - S_2(i,j) - S_2(j,i) + S_3$, where

$$S_1 = 32\pi^2\sigma^4(\boldsymbol{\theta})\int_{[-\pi,\pi]^2} \log\left|2\sin\frac{\lambda_i}{2}\right|\log\left|2\sin\frac{\lambda_j}{2}\right| w^2(\boldsymbol{\lambda})f^2(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda},$$

$$S_2(i,j) = 16\pi^2\sigma^2(\boldsymbol{\theta})\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\int_{[-\pi,\pi]^2} \log\left|2\sin\frac{\lambda_i}{2}\right| w^2(\boldsymbol{\lambda})f^2(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda},$$

$$S_3 = 8\pi^2\left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\int_{[-\pi,\pi]^2} w^2(\boldsymbol{\lambda})f^2(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda}.$$

Acknowledgments

N. Leonenko was supported in particular by Cardiff Incoming Visiting Fellowship Scheme, International Collaboration Seedcorn Fund and Australian Research Council's Discovery Projects funding scheme (project DP160101366). This work has been supported in part by project MTM2015-71839-P of MINECO, Spain (co-funded with FEDER funds). Huda Alomari was partly supported by Australian Research Council's Discovery Projects funding scheme (project number DP160101366). Huda Alomari is also grateful to Dr A.Olenko, her PhD supervisor, for many useful suggestions and discussions to improve the paper. The authors are also thankful to the referees for careful reading of the manuscript and many detailed comments and suggestions that helped to improve it.

References

[1] Anh, V. V., Angulo, J. M. and Ruiz-Medina, M. D. (1999). Possible long-range dependence in fractional random fields. *J. Stat. Plan. Infer.* **80** 95-110.

[2] Anh, V. V., Leonenko N. N. and Sakhno, L. M. (2004a). On a Class of Minimum Contrast Estimators for Fractional Stochastic Processes and Fields. *J. Stat. Plan. Infer.* **123** 161–185.

[3] Anh, V. V., Leonenko N. N. and Sakhno L. M. (2004b). Quasi-likelihood-based higher-order spectral estimation of random fields with possible long-range dependence. Stochastic methods and their applications. *J. Appl. Probab.* **41A** 35–53.

[4] Anh, V. V., Leonenko, N. N. and Sakhno, L. M. (2007a). Statistical inference based on the information of the second and third order. *J. Stat. Plan. Infer.* **137** 1302–1331.

[5] Anh, V. V., Leonenko, N. N. and Sakhno, L. M. (2007b). Evaluation of bias in higher-order spectral estimation. *Theory Probab. Math. Stat.* **80** 1–14.

[6] Anh, V. V., Leonenko, N. N. and Sakhno, L. M. (2007c). Statistical inference using higher-order information. *J. Multivariate Anal.* **98(4)** 706–742.

[7] Anh, V. V., Leonenko, N. N. and Sakhno, L. M. (2007d). Minimum contrast estimation of random processes based on information of second and third orders. *J Stat Plan Infer* **137(4)** 1302–1331.

[8] Anh, V. V. and Lunney, K. E. (1995). Parameter estimation of random fields with longrange depedence. *Math. Comput. Model.* **21** 67–77.

[9] Avram, F. (1988). On Bilinear Forms in Gaussian Random Variables and Toeplitz Matrices. *Probab. Theory Relat. Field* **79** 37–45.

[10] Avram, F., Leonenko, N. N. and Sakhno, L. M. (2010a). On a Szegö Type Limit Theorem, the Hölder-Young-Brascamp-Lieb Inequality, and the Asymptotic Theory of Integrals and Quadratic Forms of Stationary Fields. *ESAIM: PS.* **14** 210–255.

[11] Avram, F., Leonenko, N. N. and Sakhno, L. M. (2010b). Harmonic Analysis Tools for Statistical Inference in the Spectral domain. *Lecture Notes in Statistics, Springer-Verlag.* **200** 59–70.

[12] Basu, S. and Reinsel, G. C. (1993). Properties of the spatial unilateral first-order ARMA model. *Adv. Appl. Probab.* **25(3)** 631–648.

[13] Bentkus, R. (1972). The Error in Estimating the Spectral Function of a Stationary Process. *Lietuvos Matematikos Rinkinys* **12(1)** 55–71.

[14] Bentkus, R. (1976). Cumulants of Estimates of the Spectrum of a Stationary Time Series. *Litovskii Matematicheskii Sbornik* **16(4)** 37–61.

[15] Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory.* Holden Day, San Francisco.

[16] Boissy, Y., Bhattacharyya, B. B., Li, X. and Richardson, G. D. (2005). Parameter estimates for fractional autoregressive spatial process. *Ann. Stat.* **33** 2553–2567.

[17] Chan, K. S. and Tsai, H. (2012). Inference of seasonal long-memory aggregate time series. *Bernoulli* **4(18)** 1448–1464.

[18] Chung, C. F. (1996). Estimating a generalized long memory process. *J. Econom.* **73** 237–259.

[19] Dahlhaus, R. (1983). Spectral Analysis with Tapered Data. *J. Time Ser. Anal.* **4(3)** 163–175.

[20] Dahlhaus, R. and Künsch, H. (1987). Edge Effects and Efficient Parameter Estimation for Stationary Random Fields. *Biometrika* **73(4)** 877–882.

[21] Doukhan, P., León, J. and Soulier, P. (1996). Central and non central limit theorems for strongly dependent stationary Gaussian field. *Rebrape* **10** 205–223.

[22] Espejo, R., Leonenko, N. N., Olenko, A. and Ruiz-Medina, M. D. (2015). On a Class of Minimum Contrast Estimators for Gegenbauer Random Fields. *TEST* **24** 657–680.

[23] Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist* **14** 517–532.

[24] Fox, R. and Taqqu, M. S. (1987). Central Limit Theorems for Quadratic Forms in Random Variables Having Long-Range Dependence. *Probab. Theory Relat. Field* **74** 213–240.

[25] Ginovyan, M. S., Sahakyan, A. A. and Taqqu, M. S. (2014). The trace problem for Toeplitz matrices and operators and its impact in probability. *Probab. Surv.* **11** 393–440.

[26] Guo, H., Lim, C. Y. and Meerschaert, M. M. (2009). Local Whittle estimator for anisotropic random field. *J. Multivariate Anal.* **100** 993–1028.

[27] Guyon, X. (1982). Parameter Estimation for a Stationary Process on a d-Dimensional Lattice. *Biometrika* **69** 95–105.

[28] Guyon, X. (1995). *Random Fields on a Network*. Springer-Verlag, New York.

[29] Heyde, C. C. and Gay, R. (1993). Smoothed Periodogram Asymptotics and Estimation for Processes and Fields with Possible Long-Range Dependence. *Stoch. Process. Their Appl.* **45** 169–182.

[30] Ibragimov, I. A. (1967). On Maximum Likelihood Estimation of Parameters of the Spectral Density of Stationary Time Series. *Theory Probab. Appl.* **12(1)** 115–119.

[31] Ivanov, A.V. and Prikhodko, V. (2015). Asymptotic properties of Ibaragimov's estimate of parameter of spectral density of noise in the nonlinear regression method. *Theory Probab. Math.Stat.* **93** 50–66.

[32] Leonenko, N. N. and Sakhno, L. M. (2006). On the Whittle estimators for some classes of continuous-parameter random processes and fields. *Stat. Probab. Lett.* **76** 781–795.

[33] Leonenko, N. N. and Taufer, E. (2013). Disaggregation of spatial autoregressive processes. *Spatial Statistics* **3** 1–20.

[34] Li, W. K. and McLeod, A. I. (1986). Fractional time series modelling. *Biometrika* **73** 217–221.

[35] Ludeña C. (2000). Parametric estimation for Gaussian long-range dependent processes based on the log-periodogram. *Bernoulli* **6** 709–728.

[36] Ludeña, C. and Lavielle, M. (1999). The Whittle Estimator for Strongly De-

pendent Stationary Gaussian Fields. *Board of the Foundation of the Scand. J. Stat.* **26** 433–450.

[37] Reisen, V., Rodrigues, A. L. and Palma, W. (2006). Estimation of seasonal fractionally integrated processes. *Comput. Stat. Data Anal.* **50** 568–582.

[38] Robinson, P. M. (2007). Nonparametric spectrum estimation for spatial data. *J. Stat. Plan. Infer.* **137** 1024–1034.

[39] Robinson, P. M. and Sanz, J. V. (2006). Modified Whittle estimation of multilateral models on a lattice. *J. Multivariate Anal.* **97** 1090–1120.

[40] Sakhno, L. (2007). Bias control in the estimation of spectral functionals. *Theory Stoch. Process.* **13** 225–233.

[41] Sakhno, L. (2014). Minimum Contrast Method for Parameter Estimation in the Spectral Domain. In: Modern Stochastics and Applications, Springer Optimization and Its Applications V. 90, 319 − 336.

[42] Taniguchi, M. (1987). Minimum contrast estimation for spectral densities of stationary processes. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **49** 315–325.

[43] WeiLin, X., WeiGuo, Z. and XiLi, Z. (2012). Minimum contrast estimator for fractional Ornstein-Uhlenbeck processes. *Sci. China-Math.* **55(7)** 1497–1511.

[44] Yao, Q. and Brockwell. P. J. (2006). Gaussian maximum likelihood estimation for ARMA models II: Spatial processes. *Bernoulli* **12(3)** 403–429.