Testing normality of spatially indexed functional data

Siegfried HÖRMANN\textsuperscript{1}, Piotr KOKOSZKA\textsuperscript{2*}, and Thomas KUENZER\textsuperscript{1}

\textsuperscript{1}Institute of Statistics, Graz University of Technology, Graz, Austria
\textsuperscript{2}Department of Statistics, Colorado State University, Fort Collins, CO, USA

Key words and phrases: Functional data; normality test; principal components; spatial statistics.

MSC 2010: Primary 62R10; secondary 62M30.

Abstract: We develop a test of normality for spatially indexed functions. The assumption of normality is common in spatial statistics, yet no significance tests, or other means of assessment, have been available for functional data. This article aims at filling this gap in the case of functional observations on a spatial grid. Our test compares the moments of the spatial (frequency domain) principal component scores to those of a suitable Gaussian distribution. Critical values can be readily obtained from a chi-squared distribution. We provide rigorous theoretical justification for a broad class of weakly stationary functional random fields. We perform simulation studies to assess the power of the test against various alternatives. An application to surface incoming shortwave radiation illustrates the practical value of this procedure. 

1. INTRODUCTION

Over the last two decades, there has been increasing interest in functional data, where observations are regarded as elements of a suitable function space. Several monographs and textbooks give accounts of various aspects of this field, e.g., Bosq (2000), Ramsay & Silverman (2005), Shi & Choi (2011), Horváth & Kokoszka (2012), Hsing & Eubank (2015), and Kokoszka & Reimherr (2017). While studies of random samples of functions continue to dominate, the last decade has seen growing development of functional data analysis for dependent data. Contributions are becoming more and more numerous, so we list just a handful of them, without any claim on their relative importance. Within the field of time-series analysis, the emphasis has been on forecasting and inference for temporal dependence, see e.g., Hyndman & Shang (2009),

\* Corresponding author: piotr.kokoszka@colostate.edu

© 2021 Statistical Society of Canada
Hörmann & Kokoszka (2010), Liebl (2013), Horváth, Kokoszka & Rice (2014), Aue, Norinho & Hörmann (2015), Zhang (2016). In the ambit of spatial statistics, chief research directions have been kriging and inference for the spatio-temporal dependence structure. Several review articles and collections are available, e.g., Delicado et al. (2010), Mateu & Giraldo (2020), and Martinez-Hernádez & Genton (2020). Spatial functional data can be regarded as a type of spatio-temporal data: at each location, we observe a function, generally defined on a time domain. Temperature or precipitation curves at spatial locations offer well-known examples, but there are many more. Most data observed from satellite measurements and outputs of computer climate models can be treated either as a temporal sequence of spatial fields or as a field of temporal functions. We adopt the latter modelling approach. In either case, the observations are available on a regular spatial grid.

The Gaussian assumption has been utilized much more extensively and profoundly in spatial statistics than in any other field of statistics. This is chiefly due to the well-established use of covariance function modelling since covariances determine distribution only for Gaussian data. (See Gelfand & Schliep (2016) for a broader perspective.) Still, non-Gaussian data seem to be widespread in many contexts. An overview of non-Gaussianity in climatology was provided by Perron & Sura (2013). They investigated key atmospheric variables observed over several decades and came to the conclusion that Gaussianity is quite rare in the atmosphere.

Somewhat surprisingly, tests for normality of spatial data have been absent. Even the application of exploratory tools, such as QQ-plots, is questionable because these are justified only if the observations form a random sample. Horváth, Kokoszka & Wang (2020) derived a normality test for a scalar spatial field, which falls into a broad Jarque–Bera family of tests. This test is based on the asymptotic distribution of suitably defined skewness and kurtosis. In the case of spatially dependent data, these statistics must be defined differently than for random samples. (Earlier related contributions include Shenton & Bowman (1977), Jarque & Bera (1980), Jarque & Bera (1987). Lobato & Velasco (2004), Bai & Ng (2005), and Doornik & Hansen (2008).)

This article is concerned with observations that are functions with domain $\mathbb{U}$, collected at spatial locations $s \in \mathbb{Z}^d$, one function at each location. The domain $\mathbb{U}$ can be a time domain (as in our real data example), but could also be some other continuous domain, like altitude. Then, $X_s(u)$ may be, for example, the air temperature at location $s$ at time (or altitude) $u$. No normality tests are currently available for such data, to the best of our knowledge. Our objective is to fill this gap. The need for such a test arises in many contexts. For example, Liu, Ray & Hooker (2017) developed tests of spatio-temporal separability and isotropy of spatial functional data, which rely on the assumption that these data are normal. The same is true of the separability test of Constantinou, Kokoszka & Reimherr (2017). Gromenko, Kokoszka & Sojka (2017) assumed normality to derive a test for the presence of a common temporal trend in a sample of spatially indexed functions. Tests for normality of functional random samples were derived and compared in Górecki, Horváth & Kokoszka (2020), and those for functional time series in Górecki et al. (2018).

Our approach is based on the decomposition of a functional spatial field recently derived by Kuenzer, Hörmann & Kokoszka (2020). It uses spatial (frequency domain) functional principal components analysis (SFPCA) to decompose the functional spatial random field into $p$ fields of SFPC scores that are orthogonal at all spatial lags. Under normality, this orthogonality implies that the functional data are actually decomposed into $p$ layers of independent scalar random fields, each of which is again Gaussian. Therefore, the testing procedure breaks down the infinite-dimensional concept of Gaussianity of functions into testing $p$ independent scalar random fields for Gaussianity. The number $p$ represents the level of dimension reduction. It is generally a small single-digit number, often 2, 3, or 4. At each of these levels, we apply a test of normality based on the skewness and kurtosis of suitably defined spatial fields. This is the simplest and most commonly used approach, which turns out to work well.
The remainder of the article is organized as follows. After presenting the required background in Section 2, we provide a self-contained description of our test in Section 3. Section 4 is dedicated to its asymptotic justification, with the proofs collected in the Appendix. Finite-sample performance is investigated in Section 5. The article concludes with an application to surface incoming shortwave (SIS) radiation in Section 6.

2. PRELIMINARIES

Before we formulate the test, we need to introduce the notation and framework in which we operate. We consider functions defined on a spatial grid in a Euclidean space of dimension $d$. The functions live in the space $H = L^2([0, 1])$, the set of square-integrable, real-valued functions on the interval $[0, 1]$, with the usual inner product and norm. The interval $[0, 1]$ is considered only for the convenience of notation; it can be replaced by any other interval. A functional random field is then an infinite collection of random functions, $(X_s)_{s \in \mathbb{Z}^d}$, where for each $s \in \mathbb{Z}^d$, $X_s \in H$. This means that at each grid point $s$ we have a curve $X_s(u), u \in [0, 1]$. In most applications, the variable $u$ is rescaled time. We assume throughout the article that each function is square integrable, i.e.

$$
\mathbb{E} \left\| X_s \right\|^2 = \mathbb{E} \int_0^1 X_s^2(u) du < \infty.
$$

Under this assumption, we define a Gaussian functional random field as follows:

**Definition 1.** A functional random field $(X_s)_{s \in \mathbb{Z}^d}$ is called Gaussian if for all $n \in \mathbb{N}$, any deterministic elements $\{v_1, \ldots, v_n\} =: V \subset H$ and any grid points $\{s_1, \ldots, s_n\} =: S \subset \mathbb{Z}^d$, the projections $(X_{s_i} v_i)$ are jointly normally distributed, i.e.

$$
\left( \langle X_{s_1}, v_1 \rangle, \ldots, \langle X_{s_n}, v_n \rangle \right)^\top \sim \mathcal{N}_n(\mu_{S,V}, \Sigma_{S,V}),
$$

where $\mu_{S,V}$ and $\Sigma_{S,V}$ depend on the sets $S$ and $V$.

We note that there are several equivalent definitions of normality in a Hilbert space. See e.g., Chapter 7 of Laha & Rohatgi (1979).

We now define a stationary functional random field.

**Definition 2.** A functional random field $(X_s)_{s \in \mathbb{Z}^d}$ is called weakly stationary if

(i) for all $s \in \mathbb{Z}^d$, $\mathbb{E} X_s = \mathbb{E} X_0$;

(ii) for all $s, h \in \mathbb{Z}^d$ and $u, v \in [0, 1]$

$$
c_h(u, v) := \text{Cov}(X_{h}(u), X_0(v)) = \text{Cov}(X_{s+h}(u), X_s(v)).
$$

Observe that the kernel $c_h$ is Hilbert–Schmidt, i.e., $\int c_h^2(u, v) du dv < \infty$. For any Hilbert–Schmidt kernel $\psi$, we define the corresponding operator $\Psi$ on $H$ by $\Psi(y) = \int \psi(u, v)y(v) dv$, $y \in H$. The integral operator defined by the autocovariance kernel $c_h$ is thus denoted by $C_h$.

Next we turn to the concept of SFPCA, which relies on frequency domain concepts. In particular, we need the so-called spectral density operator. For a weakly stationary functional random field, the integral operator $F^X_{\theta}$ with the kernel

$$
f^X_{\theta}(u, v) := \frac{1}{(2\pi)^d} \sum_{h \in \mathbb{Z}^d} c_h(u, v) e^{-ih^\top \theta}, \quad \theta \in (-\pi, \pi]^d
$$

Published online by Cambridge University Press.

DOI: 10.1002/cjs.11662
is called the spectral density operator of \((X_s)\) at the spatial frequency \(\theta\).

To ensure convergence of the infinite series in (2), we impose the following assumption.

**Assumption 1.** The field \((X_s)_{s \in \mathbb{Z}^d}\) is weakly stationary with mean zero and absolutely summable autocovariances in the sense that

\[
\sum_{k \in \mathbb{Z}^d} \text{Tr} \left( C^X_k \right) < \infty,
\]

where \(\text{Tr}(\cdot)\) denotes the trace norm, defined as the sum of the singular values of the operator.

Exponentially decaying autocovariances satisfy (3), but admit slower decay. Under Assumption 1, the theory of Kuenzer, Hörmann & Kokoszka (2020) is applicable. We now outline its elements, which we need for the development of the normality test.

The SFPC scores \(Y_{m,s}\) and the filter functions \(\phi_{m,k}\) are defined using the eigensystem of the spectral density operator. Let \(\lambda_1(\theta) > \lambda_2(\theta) > \cdots > \lambda_m(\theta) > \cdots > 0\) be the ordered eigenvalues of \(F^X_\theta\) and \(\varphi_m(\cdot|\theta)\) be the eigenvector corresponding to \(\lambda_m(\theta)\). The level \(m\) SFPC scores are defined by

\[
Y_{m,s} := \sum_{k \in \mathbb{Z}^d} \langle X_{s-k}, \phi_{m,k} \rangle, \quad s \in \mathbb{Z}^d,
\]

where \(\phi_{m,k}\) is defined by

\[
\phi_{m,k}(u) := \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \varphi_m(u|\theta)e^{-ik\cdot\theta} \, d\theta.
\]

The SFPC score fields \((Y_{m,s} : s \in \mathbb{Z}^d), 1 \leq m \leq p,\) are then orthogonal in the sense that

\[
\text{Cov} \left( Y_{m,s}, Y_{n,s'} \right) = 0, \quad \text{if} \ m \neq n, \ \forall s, s' \in \mathbb{Z}^d.
\]

The usual functional principal component (FPC) scores are defined by \(\xi_{m,s} = \langle X_s, v_m \rangle\), where \(v_m\) is the \(m\)th FPC. They are uncorrelated at each location, i.e., \(\text{Cov}(\xi_{m,s}, \xi_{n,s}) = 0, \) if \(m \neq n\). The analogue of (6) does not hold. As we will see in Section 3, it is property (6) that allows us to construct our normality test. The scores \(Y_{m,s}, s\), obtained with the SFPCA, define a spatial field for each \(m\). They take into account data from neighbouring spatial locations \(s\) (in theory all \(s\)) via a spatial filter \((\phi_{m,k}(u) : |k| \in \mathbb{Z}^d)\). The filters are chosen in such a way that scores \((Y_{m,s} : s \in \mathbb{Z}^d)\) from different “layers” \(m\) become mutually orthogonal (uncorrelated). Under Gaussianity, this implies that \((Y_{m,s} : s \in \mathbb{Z}^d)\) and \((Y_{m',s} : s \in \mathbb{Z}^d)\) are independent fields for \(m \neq m'\). One should contrast this property with the scores in the usual Karhunen–Loève expansion, for which we could conclude only that \(\xi_{m,s}\) is independent of \(\xi_{m',s}\) but not necessarily from \(\xi_{m',s'}\) at a different location \(s'\).

The population scores \(Y_{m,s}\) can be approximated by their sample counterparts \(\hat{Y}_{m,s}\). The construction of the sample scores \(\hat{Y}_{m,s}\) involves several steps, which are explained in Sections 3.2 and 3.3 of Kuenzer, Hörmann & Kokoszka (2020). In a nutshell, the spectral density estimator \(\hat{F}^X_\theta\) is estimated by a suitably constructed estimator \(\hat{P}^X_\theta\), and the steps listed above are applied to \(\hat{P}^X_\theta\) in place of \(F^X_\theta\), with infinite sums replaced by truncated sums. The estimated scores are thus defined by

\[
\hat{Y}_{m,s} := \sum_{\|k\| \leq L} \langle X_{s-k}, \hat{\phi}_{m,k} \rangle, \quad 1 \leq m \leq p,
\]
where $\hat{\phi}_{m,k}$ are the Fourier expansion coefficients of the eigenvectors $\hat{\phi}_m(\theta)$ of $\hat{\mathbb{F}}^{X}$. While the exact choice of this estimator is not crucial to our method, we define the estimated spectral density by

$$
\hat{\mathbb{F}}^{X}_\theta := \frac{1}{(2\pi)^r} \sum_{h} w_q(h) \hat{C}_h e^{-ih^\top \theta},
$$

where $\hat{C}_h$ are the usual sample autocovariance operators at lag $h$, $w_q$ is a weight function, and $q$ is a vector of positive window sizes. There are different possibilities for the choice of the weight function. For our calculations, we used the Bartlett kernel $w_q(z) = (1 - \|z/q\|)^+$.  

### 3. DESCRIPTION OF THE TEST

We will use the sample moments of a generic scalar field $(Z_s)_{s \in \mathbb{Z}^d}$ observed on a domain $R_n \subset \mathbb{Z}^d$ with cardinality $|R_n| = N$. We defined them by

$$
\hat{m}_k^Z := \frac{1}{N} \sum_{s \in R_n} Z^k_s.
$$

The Jarque–Bera test compares sample skewness and kurtosis of a distribution with the corresponding values of a normal distribution. Let $(Z_s)_{s \in \mathbb{Z}^d}$ be a stationary scalar random field on a grid. The mean of $Z_s$ is denoted by $\mu = \mathbb{E}Z_s$, and the $k$th central moment is $\mu_k = \mathbb{E}[(Z_s - \mu)^k]$. The variance, as a special case, is denoted by $\sigma^2 = \mu_2$. Skewness and kurtosis are defined by

$$
\tau = \frac{\mu_3}{\sigma^3} \quad \text{and} \quad \kappa = \frac{\mu_4}{\sigma^4}.
$$

These parameters can be estimated by the corresponding sample moments. We call the resulting estimators, $\hat{\tau}$ and $\hat{\kappa}$, the sample skewness and kurtosis. If the $Z_s$ are i.i.d., then the standard Jarque–Bera test is based on the convergence

$$
\text{JB}_N := N \left( \frac{\hat{\tau}^2}{6} + \frac{(\hat{\kappa} - 3)^2}{24} \right) \xrightarrow{d} \chi^2_d,
$$

which holds under the null hypothesis of normality. Under spatial dependence, convergence (8) no longer holds, as explained in Horváth, Kokoszka & Wang (2020).

For the observed functional field $(X_s)_{s \in R_n}$, we proceed as follows: For $1 \leq m \leq p$, we compute the estimated score fields $(\hat{Y}_{m,s})$. We centre each field and obtain

$$
\hat{Z}_{m,s} = \hat{Y}_{m,s} - \hat{m}_1^y_m, \quad s \in R_n.
$$

Next, we compute for the levels $m \in \{1, \ldots, p\}$ the statistics related to sample skewness and kurtosis, which are defined by

$$
\hat{S}_n^{(m)} := \sqrt{N} \hat{m}_3^y_m, \quad \hat{K}_n^{(m)} := \sqrt{N} \left( \hat{m}_4^y_m - 3 \left( \hat{m}_2^y_m \right)^2 \right).
$$

Finally, we define the test statistic by

$$
\hat{T}_p := \sum_{m=1}^p \hat{J}_m, \quad \text{with} \quad \hat{J}_m := \frac{\left( \hat{S}_n^{(m)} \right)^2}{6 \hat{\sigma}_{S,m}^2} + \frac{\left( \hat{K}_n^{(m)} \right)^2}{24 \hat{\sigma}_{K,m}^2}.
$$
The variance estimators $\hat{\sigma}^2_{S,m}$ and $\hat{\sigma}^2_{K,m}$ are defined by

$$\hat{\sigma}^2_{S,m} = \sum_{\|k\|_{\infty} \leq L'} \hat{\gamma}^3_{m,k^*} \quad \text{and} \quad \hat{\sigma}^2_{K,m} = \sum_{\|k\|_{\infty} \leq L'} \hat{\gamma}^4_{m,k^*}$$  \hspace{1cm} (11)

with the sample autocovariances defined by

$$\hat{\gamma}_{m,h} = \frac{1}{N} \sum_{s \in M_{h,n}} (\hat{Y}_{m,s+h} - \bar{Y}_m^{M}) (\hat{Y}_{m,s} - \bar{Y}_m^{M})$$.

The set $M_{h,n}$ is defined as the set of the locations for which $s \in R_n$ and $s + h \in R_n$. The truncation parameter $L'$ is discussed below. We will see in Section 4 that the test statistic $\hat{T}_p$ is asymptotically $\chi^2_p$-distributed under the null.

We conclude this section with an algorithmic description of the test, which contains guidance on the choice of the tuning parameters and suitable $R$ functions. We use our package `fsd.fd` that is available on Github. The following recommendations reflect our experience based on extensive numerical experiments.

1. In a first step, the spectral density is estimated (`fsd.spectral.density`) on a suitable equidistant integration grid. The window size parameter $q \in \mathbb{N}^d$ can be selected according to a suitable rule of thumb, such as $q_i = \sqrt{n_i}$. However, we recommend using the automatic, data-driven procedure described in Section III of Kuenzer, Hörmann & Kokoszka (2020).

2. Using the function `fsd.spca.var`, we obtain an estimate of the portion of variability that the single layers of SFPC scores explain. We then choose $p$ such that the first $p$ SFPCs explain at least 85% of the total variability of the functional data.

3. For the computation of the filter functions, we use `fsd.spca.filters` with the parameters $N_{pc} = p$, and the maximum lag $L$ to calculate the SFPC filters and scores is chosen as the smallest integer $L$ such that the filter functions reach at least 95% of the total weight, i.e., $\sum_k \|\phi_{1,k}\|^2 \geq 0.95$.

4. We apply the filter functions to the functional random field (`fsd.spca.scores`) to obtain the SFPC scores.

5. Finally, we can use `fsd.jb.test` to conduct the test. For this, we supply as argument `X.spca` a list with the entry `scores` and set `var.method = "direct"`. Regarding the choice of $L'$, i.e., the maximum lag of the SFPC score autocovariances for the estimation of $\hat{\sigma}^2_S$ and $\hat{\sigma}^2_K$, we recommend setting the argument $L = \|q\|_{\infty}$ in order to capture enough covariance. This selection criterion stems from the fact that the spatial dependence structure of the original data ($X_s$) is mostly reflected in the SFPC scores.

It is possible to shorten the procedure by using the function `fsd.spca` with suitably preselected arguments and then following up with step (5). The levels of 85% in step (2) and of 95% in step (3) are unrelated and somewhat arbitrary. The first is related to the variability of the data explained by SFPCs, and the second to the numerical accuracy of the approximation of the filters. The level of 85% in step (2) is fairly standard in FDA and works well in the context of this article. The level of 95% in step (3) is also typical and turns out to work well in our simulations and applications.

4. ASYMPTOTIC JUSTIFICATION

As noted at the end of Section 2, the starting point to the implementation of the test is an estimation of the spectral density operator. We need only the following weak assumption, which

\[ DOI: 10.1002/cjs.11662 \]
is satisfied by the estimator proposed by Kuenzer, Hörmann & Kokoszka (2020) for broad classes of functional random fields:

**Assumption 2.** The estimator $\hat{F}_θ^X$ satisfies

$$\int_{[-π,π]^d} \mathbb{E} \left\| \hat{F}_θ^X - F_θ^X \right\| dθ \to 0,$$

where $\| \cdot \|_C$ is the operator norm.

The next assumption, also used in Kuenzer, Hörmann & Kokoszka (2020), is needed to ensure the identifiability and the convergence of the SFPC estimators.

**Assumption 3.** Let $α_m(θ)$ be the spectral gaps, i.e.

$$α_1(θ) := \lambda_1(θ) - \lambda_2(θ),$$
$$α_m(θ) := \min\{\lambda_m(θ) - \lambda_{m+1}(θ), \lambda_{m-1}(θ) - \lambda_m(θ)\}, \quad ∀ m ≥ 2.$$

We assume that for all $1 ≤ m ≤ p$, the spectral gaps are bounded from below, such that

$$\inf_{θ ∈ [-π,π]^d} α_m(θ) =: β_m > 0.$$

The final assumption on the population quantities refers to the summability of the filter functions.

**Assumption 4.** For all $1 ≤ m ≤ p$, the filter functions of the SFPCs are absolutely summable in the sense that

$$\sum_{k ∈ ℤ^d} \| φ_{m,k} \| < ∞.$$

Next we turn to the assumption on the sampling region.

**Assumption 5.** The sampling region is the rectangle

$$R_n = \{ s ∈ ℤ^d : 1 ≤ s_i ≤ n_i, ∀ 1 ≤ i ≤ d \}$$

such that $\min_{1 ≤ i ≤ d} n_i → ∞$.

Recall that the number of locations in $R_n$ is denoted by $N$, and the index $n$ is used to identify this expanding spatial domain. Assumption 5 could be replaced by a more complex technical assumption, but it simplifies arguments and is generally satisfied in applications.

Asymptotic results are stated in terms of the following quantities:

$$G(N) := \int_{[-π,π]^d} \left\| \hat{F}_θ^X - F_θ^X \right\| dθ, \quad H_m(L) := \left( \sum_{\|k\|_∞ > L} \| φ_{m,k} \|^2 \right)^{1/4}.$$

Our main asymptotic result, Theorem 1, states that the asymptotic null distribution of the test statistic $\hat{T}_p$ defined by (10) is chi-squared with $2p$ degrees of freedom. Recall that $p$ is the number of levels in the SFPCA used to construct the statistic. At each level, the asymptotic distribution is chi-squared with two degrees of freedom, and by utilizing the asymptotic independence between the levels we obtain the desired result. The proof is presented in the Appendix. It is quite

The Canadian Journal of Statistics / La revue canadienne de statistique

DOI: 10.1002/cjs.11662
complex because independence properties hold only at the population level. At the sample level, independence is only asymptotic.

To understand Theorem 1, we first consider analogues of the statistics \( \hat{S}_n^{(m)} \) and \( \hat{K}_n^{(m)} \) defined in (9) in terms of the population scores \( Y_{m,s} \) rather than the estimated scores \( \hat{Y}_{m,s} \). We thus set

\[
\hat{S}_n^{(m)} := \sqrt{N} \hat{m}_3 Z_m, \quad \hat{K}_n^{(m)} := \sqrt{N} \left( \hat{m}_4 - 3 \left( \hat{m}_2 \right)^2 \right), \quad Z_{m,s} = Y_{m,s} - m_1 Y_m. \tag{13}
\]

Set

\[
\gamma_{m,h} = \text{Cov}(Y_{m,s+h}, Y_{m,s}).
\]

By Lemma 1

\[
\left( \begin{array}{c}
\hat{S}_n^{(m)} \\
\hat{K}_n^{(m)}
\end{array} \right) \xrightarrow{d} N_2 \left( \begin{array}{cc}
0 & 6 \sigma_{S,m}^2 \\
0 & 24 \sigma_{K,m}^2
\end{array} \right), \tag{14}
\]

where

\[
\sigma_{S,m}^2 = \sum_{h \in \mathbb{Z}^d} \gamma_{m,h}^3 \quad \text{and} \quad \sigma_{K,m}^2 = \sum_{s \in \mathbb{Z}^d} \gamma_{m,h}^4.
\tag{15}
\]

Theorem 1 states that under technical assumptions, the null distribution is \( \chi^2_{2p} \), as long as the asymptotic variances \( \sigma_{S,m}^2 \) and \( \sigma_{K,m}^2 \) can be consistently estimated.

**Theorem 1.** Suppose Assumptions 1–5 are satisfied and \( (X_s) \) is a Gaussian process. Suppose that \( L = L(N) \to \infty \) such that \( L^d G(N) \to 0 \), and \( L' = L'(N) \to \infty \), and for \( 1 \leq m \leq p \), \( \hat{\sigma}_{S,m}^2 \) and \( \hat{\sigma}_{K,m}^2 \) in (10) are consistent estimators of \( \sigma_{S,m}^2 \) and \( \sigma_{K,m}^2 \) in (15). Then \( \hat{T}_p \xrightarrow{d} \chi^2_{2p} \).

The remaining question that must be addressed is whether the estimators considered in Section 3 are consistent. This is indeed the case, as stated in the following proposition:

**Proposition 1.** Suppose Assumptions 1–5 hold and \( (X_s) \) is a Gaussian process. Assume that, as \( N \to \infty \),

\[
L, \quad L' \to \infty, \quad \text{such that} \quad L' = o \left( \min_{1 \leq i \leq d} n_i \right), \quad (L')^{d/3} H_m(L) \xrightarrow{P} 0.
\]

Then, \( \hat{\sigma}_{S,m}^2 \) and \( \hat{\sigma}_{K,m}^2 \) defined by (11) are consistent for the asymptotic variances (15).

Proofs of both Theorem 1 and Proposition 1 are developed in the Appendix.

### 5. FINITE-SAMPLE PERFORMANCE

The purpose of this section is to assess the performance of our test by means of a simulation study. To evaluate empirical size and power, we simulated samples

\[
X_{s,t}(u), \quad s, t \in \{1, 2, \ldots, n\}, \quad u \in [0, 1],
\]

according to the following autoregressive scheme:

\[
X_{s,t} = AX_{s-1,t} + BX_{s,t-1} + \epsilon_{s,t},
\]

DOI: 10.1002/cjs.11662

The Canadian Journal of Statistics / La revue canadienne de statistique
where the $\epsilon_{s,t}$ are i.i.d. errors and $A$ and $B$ are two operators. We explain the details below, but the idea is that $\epsilon_{s,t}$ are Gaussian curves under the null hypothesis and have different distributions under alternatives.

We simulated samples in the finite-dimensional space spanned by 15 Fourier basis functions $(v_i)_{1 \leq i \leq 15}$. The $X_s$ are determined by their coefficient vectors. The operators are represented by $15 \times 15$ dimensional coefficient matrices whose $(i,j)$th entries are simulated as independent normal random variables with mean 0 and variance $(i^2 + j^2)^{-1/2}$. The operators are then scaled to the operator norm $\|A\|_F = 0.6$ and $\|B\|_F = 0.35$, which ensures convergence and stationarity (Kuenzer, Hörmann & Kokoszka, 2020).

Under the null, we simulate $\epsilon_{s,t}$ such that the Fourier coefficients $\epsilon_{s,t,i} = \langle \epsilon_{s,t}, v_i \rangle$ are independent and normal with mean zero and variance $2^{-i}$. For these data, three SFPCs explain 85% of the variance. Under the alternative, we simulate the $\epsilon_{s,t,i}$ from Johnson’s $S_U$ distribution (Johnson, 1949), which is a leptokurtic distribution family with four parameters that allows fixing the first four moments. It is defined as a transformation of the normal distribution by

$$X = \xi + \lambda \sinh \left( \frac{Z - Y}{\delta} \right), \quad \text{with} \quad Z \sim N(0,1).$$

For this distribution, all moments exist. We chose the parameters such that the mean and variance of $\epsilon_{s,t,i}$ are the same as under the null setting, and specify skewness $\tau$ and kurtosis $\kappa$ separately. We denote such a distribution as $S_U(\tau, \kappa)$.

We simulated these data-generating processes for several sample sizes. Empirical rejection rates can be found in Table 1. The critical values are the quantiles of the $\chi^2_{2p}$ distribution.

The nominal size is attained in the simulation settings we considered. The empirical size is stable with respect to $p$, while the empirical power increases with $p$. (The number of SFPCs we would use according to our goal of explaining 85% of the variability in the data is $p = 3$.) This can be explained by the simulation setting we have chosen, where all principal components are similarly non-Gaussian. For other settings, it may well be possible that the deviation from Gaussianity manifests only in certain principal components. In such cases, choosing a large $p$ might drown out this signal and, in fact, lower the power of the test.

Our results suggest that the power of the test is similar to that of the functional time-series case considered in Górecki et al. (2018) and the scalar spatial case studied in Horváth, Kokoszka & Wang (2020).

Our conclusion is that the test is able to reliably detect moderate deviations from normality even when the sample size is small ($12 \times 12$). Starting from medium sample sizes ($25 \times 25$), most practically relevant departures from normality can be detected. The case closest to Gaussianity that we considered is a kurtosis of $\kappa = 3.2$, which is the kurtosis of a Student’s $t$-distribution with 34 degrees of freedom. The $S_U$-distribution with this kurtosis is almost identical to the corresponding Student’s $t$-distribution except for the tails that permit all moments to be finite. In scalar data, this kind of distribution is visually indistinguishable from the normal distribution. For many practical applications, this deviation from normality can even be neglected.

6. APPLICATION TO SURFACE INCOMING SHORTWAVE RADIATION

Research by Perron & Sura (2013) on many aspects of the atmosphere suggests that when taking daily mean values, many variables do not follow a Gaussian distribution. Horváth, Kokoszka & Wang (2020) analyzed mean monthly sea surface temperature data from various regions of the world and came to the conclusion that one needs to be cautious when assuming Gaussianity of spatial data. We want to explore what is revealed by the application of our test to an important dataset. We studied SIS radiation data, also called solar surface irradiance, provided by EUMETSAT (Schulz et al., 2009). These data are available for free download.
| $N$     | 12 x 12 | 25 x 25 | 50 x 50 |
|---------|---------|---------|---------|
| $p$     | 1       | 2       | 3       | 4       | 1       | 2       | 3       | 4       | 1       | 2       | 3       | 4       |
| Gaussian| 0.05    | 0.06    | 0.06    | 0.07    | 0.05    | 0.05    | 0.06    | 0.06    | 0.06    | 0.06    | 0.06    | 0.06    |
| $S_U(0, 3.2)$ | 0.08    | 0.10    | 0.11    | 0.12    | 0.10    | 0.12    | 0.15    | 0.17    | 0.15    | 0.19    | 0.26    | 0.35    |
| $S_U(0.1, 3.2)$ | 0.08    | 0.09    | 0.10    | 0.11    | 0.12    | 0.15    | 0.18    | 0.22    | 0.24    | 0.30    | 0.45    | 0.58    |
| $S_U(0, 3.5)$ | 0.12    | 0.14    | 0.17    | 0.18    | 0.20    | 0.27    | 0.35    | 0.44    | 0.45    | 0.59    | 0.78    | 0.91    |
| $S_U(0.25, 3.5)$ | 0.14    | 0.17    | 0.20    | 0.22    | 0.32    | 0.42    | 0.56    | 0.66    | 0.70    | 0.85    | 0.96    | 0.99    |
| $S_U(0, 4)$ | 0.19    | 0.25    | 0.29    | 0.32    | 0.42    | 0.55    | 0.69    | 0.81    | 0.78    | 0.90    | 0.98    | 1.00    |
| $S_U(0.5, 4)$ | 0.26    | 0.32    | 0.39    | 0.45    | 0.63    | 0.80    | 0.91    | 0.97    | 0.94    | 0.99    | 1.00    | 1.00    |
| $S_U(0, 5)$ | 0.30    | 0.39    | 0.48    | 0.55    | 0.66    | 0.83    | 0.94    | 0.98    | 0.97    | 0.99    | 1.00    | 1.00    |
| $S_U(1, 5)$ | 0.51    | 0.62    | 0.72    | 0.79    | 0.91    | 0.99    | 1.00    | 1.00    | 0.99    | 1.00    | 1.00    | 1.00    |
| $S_U(0, 6)$ | 0.41    | 0.52    | 0.60    | 0.67    | 0.81    | 0.92    | 0.98    | 1.00    | 0.99    | 1.00    | 1.00    | 1.00    |
| $S_U(1, 6)$ | 0.55    | 0.65    | 0.76    | 0.81    | 0.93    | 0.98    | 1.00    | 1.00    | 1.00    | 1.00    | 1.00    | 1.00    |
Longitude | Latitude
---|---
-45 | 15
-35 | 20
-25 | 25
15 | 30
20 | 35

**Figure 1:** SIS radiation anomalies as viewed over different time scales. Daily mean (left) versus 3-month’s mean (right). Note that the regional mean of the SIS radiation anomaly is positive in 2015. For orientation, the archipelagos of the Azores and Cape Verde are added to the map.

The SIS data measure how much solar radiation reaches a certain point of the earth’s surface on average on a given day. It is measured in W/m². The SIS radiation has as a natural upper limit, the solar radiation that reaches the top of the atmosphere, which varies naturally according to the angle of the sun at a certain latitude on a given day of the year. The amount of radiation that reaches the surface is then reduced by the absorption in the atmosphere, mainly due to clouds. For our purposes, we extracted a region in the Northern Atlantic Ocean (10°–40°N, 50°–20°E) at a spatial resolution of 0.25°. This yields 120 × 120 spatial measurement points with a temporal time span of 33 years (1983–2015) of daily observations. The region we chose has a homogeneous surface with few topographical features. Also, it does not comprise high latitudes, which would induce distortion due to the projection onto the angular grid. Hence, we can assume that after suitably de-meaning the data, the resulting field is stationary.

We took the daily measurements of the SIS radiation data for the summer months June, July, and August (92 days), and fitted smooth curves for each year and point on the grid. We used a 30-dimensional B-spline base, notably smoothing out the rather volatile data. The resulting curves can be interpreted as a moving average over time, which indicates how much solar radiation reaches the surface around a given point in time. Because we have 33 years of data, we centre the data by subtracting the long-term mean from each point on the grid and day of the year. The centred data can be referred to as anomalies. Snapshots of the resulting data can be seen in Figures 1 and 2.

From the nature of the daily mean data and from the left plot in Figure 1, one can already guess that it will probably exhibit notable skewness, which excludes Gaussianity. But if we average over a longer time period, say the summer season, it is conceivable that the mean values approach a Gaussian distribution. While the data may still deviate from normality, it will be close enough to justify the application of algorithms that require Gaussianity. We will illustrate this process of smoothing out the non-Gaussianity by employing the scalar test of Horváth, Kokoszka & Wang (2020) to verify that the marginal distributions (data on a day-by-day level) clearly do not follow a Gaussian distribution, while the mean values from a summer season do. These are the two extrema of smoothing. The smoothed B-spline curves will naturally lie somewhere in between. In a strict sense, the curves will not be Gaussian, but our test helps to assess whether the assumption of Gaussianity is reasonable or not.

In the following, we used \( L' = 20 \) to capture enough of the covariance of the SFPC scores. We set \( q_i = 15 \) and followed the other steps of the algorithm at the end of Section 3. For comparison, we use the test of Horváth, Kokoszka & Wang (2020) with the power estimator of
the covariances. In order to guarantee comparability, we use the same truncation parameter as for our functional test.

Results are presented in Table 2. We do not report the results of the univariate tests on the single days, as aggregation of the marginal $P$-values is a delicate matter and it suffices to say that most $P$-values were in the range of $<10^{-4}$, indicating a strong departure from Gaussianity. We note that for the samples of the 3-month’s mean SIS radiation and a significance level of $\alpha = 0.05$, the test by Horváth, Kokoszka & Wang (2020) rejects the null hypothesis of a Gaussian distribution in merely 6 of the 33 years, marked by the $P$-values in italics. This means that for most years, the 3-month’s mean of the summer months can actually be assumed to be Gaussian, which is a consequence of a general central limit theorem principle. The number of rejections rises to 26 if we use the fitted B-spline curves. The most general conclusion is that the fitted curve data cannot be assumed to be Gaussian for most years. This means that by smoothing out the daily fluctuations, we do not approach the Gaussian distribution as fast as expected. However,

![Figure 2: The SIS radiation anomalies (W/m²) at 39.875°N 49.875°E in summer 2015. The time (June – August) is rescaled to [0, 1].](image)

| Year | B-spline | 3-m mean | Year | B-spline | 3-m mean | Year | B-spline | 3-m mean |
|------|----------|----------|------|----------|----------|------|----------|----------|
| 1983 | 0.0132   | 0.1190   | 1994 | 0.7971   | 0.8681   | 2005 | 0.4951   | 0.0437   |
| 1984 | 0.0455   | 0.1783   | 1995 | 0.0000   | 0.3721   | 2006 | 0.0010   | 0.0100   |
| 1985 | 0.0000   | 0.4274   | 1996 | 0.0498   | 0.3472   | 2007 | 0.0010   | 0.8964   |
| 1986 | 0.5611   | 0.0989   | 1997 | 0.0002   | 0.1038   | 2008 | 0.1465   | 0.5577   |
| 1987 | 0.0003   | 0.0828   | 1998 | 0.0193   | 0.7030   | 2009 | 0.0005   | 0.2197   |
| 1988 | 0.0001   | 0.0039   | 1999 | 0.0000   | 0.0281   | 2010 | 0.0000   | 0.4231   |
| 1989 | 0.0000   | 0.9229   | 2000 | 0.0000   | 0.6011   | 2011 | 0.6868   | 0.1196   |
| 1990 | 0.0000   | 0.1030   | 2001 | 0.0083   | 0.9889   | 2012 | 0.0425   | 0.1164   |
| 1991 | 0.0000   | 0.0002   | 2002 | 0.0000   | 0.3769   | 2013 | 0.0018   | 0.1620   |
| 1992 | 0.0322   | 0.1509   | 2003 | 0.8110   | 0.2778   | 2014 | 0.0000   | 0.2479   |
| 1993 | 0.0000   | 0.0000   | 2004 | 0.0000   | 0.2282   | 2015 | 0.0971   | 0.4810   |

Table 2: $P$-values of the normality test applied to the SIS radiation anomaly curves on a $120 \times 120$ grid. The number $p$ of SFPCs is different in each year’s sample, depending on the fraction of variance explained by the SFPCs. Rejections of the null hypothesis on a level $\alpha = 0.05$ are marked in italics.
this result is subject to changes depending on the amount of smoothing that is applied on the data. After sufficient smoothing of the daily data, the application of tools of spatio-temporal statistics that are based on the assumption of Gaussianity might be justified.

Depending on the year, our test uses 11–13 SFPCs to cover 85% of the variance. For some of the years in which we do not obtain a rejection, further inspection shows that strong evidence of non-Gaussianity appears only in the 15th or 16th SFPC. Naturally, this is not captured by our test. This problem would not occur in multivariate data, where the same procedure can be applied with the number of principal components set to the dimension of the multivariate data.

ACKNOWLEDGEMENT

This work was partially supported by NSF grants DMS-1914882 and DMS-1923142.

REFERENCES

Arcones, M. A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *The Annals of Probability*, 22, 2242–2274.

Aue, A., Norinho, D. D., & Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110, 378–392.

Bai, J. & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, 23, 49–60.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, Berlin.

Constantinou, P., Kokoszka, P., & Reimherr, M. (2017). Testing separability of space-time functional processes. *Biometrika*, 104, 425–437.

Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics*, 21, 224–239.

Doornik, J. A. & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70, 927–939.

Gelfand, A. & Schliep, E. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spatial Statistics*, 18, 86–104.

Górecki, T., Hörmann, S., Horváth, L., & Kokoszka, P. (2018). Testing normality of functional time series. *Journal of Time Series Analysis*, 39, 471–487.

Górecki, T., Horváth, L., & Kokoszka, P. (2020). Tests of normality of functional data. *International Statistical Review*, 88, 677–697.

Gromenko, O., Kokoszka, P., & Sojka, J. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics*, 11, 898–918.

Hörmann, S. & Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38, 1845–1884.

Horváth, L. & Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, Berlin.

Horváth, L., Kokoszka, P., & Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179, 66–82.

Horváth, L., Kokoszka, P., & Wang, S. (2020). Testing normality of data on a multivariate grid. *Journal of Multivariate Analysis*, 179, 104640.

Hsing, T. & Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, New York.

Hyndman, R. J. & Shang, H. L. (2009). Forecasting functional time series. *Journal of the Korean Statistical Society*, 38, 199–211.

Jarque, C. M. & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6, 255–259.

Jarque, C. M. & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55, 163–172.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36, 149–176.

Kokoszka, P. & Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press.
Kuenzer, T., Hörmann, S., & Kokoszka, P. (2020). Principal component analysis of spatially indexed functions. *Journal of the American Statistical Association*, https://doi.org/10.1080/01621459.2020.1732395.

Laha, R. G. & Roghatgi, V. K. (1979). *Probability Theory*. Wiley, New York.

Liebl, D. (2013). Modeling and forecasting electricity prices: A functional data perspective. *The Annals of Applied Statistics*, 27, 1639–1654.

Liu, C., Ray, S., & Hooker, G. (2017). Functional principal components analysis of spatially correlated data. *Statistics and Computing*, 27, 1639–1654.

Lobato, I. & Velasco, C. (2004). A simple test of normality for time series. *Econometric Theory*, 20, 671–689.

Martinez-Hernández, I. & Genton, M. G. (2020). Recent developments in complex and spatially correlated functional data. *Brazilian Journal of Probability and Statistics*, 34, 204–229.

Mateu, J. & Giraldo, R. (Eds.) (2020). *Geostatistical Functional Data Analysis: Theory and Methods*. Wiley, New York. (Forthcoming).

Perron, M. & Sura, P. (2013). Climatology of non-Gaussian atmospheric statistics. *Journal of Climate*, 26, 1063–1083.

Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*. Springer, Berlin.

Schulz, J., Albert, P., Behr, H.-D., Caprion, D., Deneke, H., Dewitte, S., Dürr, B., et al. (2009). Operational climate monitoring from space: The EUMETSAT satellite application facility on climate monitoring (CM-SAF). *Atmospheric Chemistry and Physics*, 9, 1687–1709.

Shenton, L. R. & Bowman, K. O. (1977). A bivariate model for the distribution of $\sqrt{b_1}$ and $b_2$. *Journal of the American Statistical Association*, 72, 206–211.

Shi, J. Q. & Choi, T. (2011). *Gaussian Process Regression Analysis for Functional Data*. CRC Press.

Zhang, X. (2016). White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics*, 194, 76–95.

**APPENDIX: PROOFS OF THEOREM 1 AND PROPOSITION 1**

We begin with two known results, which play an important role in our arguments. Theorem 2 is well known. Theorem 3 was essentially established by Arcones (1994). Although the title of the article suggests otherwise, the results are also true for Gaussian fields, not only for sequences. Theorem 3 is thus the spatial version of Theorem 2 in Arcones (1994).

**Theorem 2 (Isserlis).** If $(W_1, \ldots, W_n)$ is a multivariate normal vector with zero mean, then $\mathbb{E}[\prod W_i] = 0$ if $n$ is odd and $\mathbb{E}[\prod W_i] = \sum \prod \mathbb{E}[W_i W_j]$ if $n$ is even, where the sum is over all partitions of \( \{1, \ldots, n\} \) into pairs \( \{i, j\} \).

**Theorem 3 (Arcones).** Let $(X_s)_{s \in \mathbb{Z}^d}$ be a stationary Gaussian random field of mean-zero $\mathbb{R}^r$-valued vectors denoted by $X_s = \left( X_s^{(1)}, \ldots, X_s^{(r)} \right)^T$, and let $f : \mathbb{R}^d \to \mathbb{R}$ be a function with $\mathbb{E} \left[ f^2 (X_1) \right] < \infty$.

Define

$$ r^{(p, q)}(k) = \mathbb{E} \left[ X_s^{(p)} X_{s+k}^{(q)} \right] $$

for $k \in \mathbb{Z}^d$ and $1 \leq p, q \leq r$. Suppose that

$$ \lim_{N \to \infty} \frac{1}{N} \sum_{k \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d} r^{(p, q)}(k - j) \quad \text{and} \quad \lim_{N \to \infty} \frac{1}{N} \sum_{k \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d} (r^{(p, q)}(k - j))^2 $$

exist for $1 \leq p, q \leq r$. Then

$$ \frac{1}{\sqrt{N}} \sum_{s \in \mathbb{Z}^d} \left( f(X_s) - \mathbb{E} [f(X_s)] \right) \overset{d}{\to} N(0, \sigma^2), \quad (A1) $$

DOI: 10.1002/cjs.11662
where
\[ \sigma^2 := \sum_{h \in \mathbb{Z}^d} \text{Cov} \left( f \left( X_h \right), f \left( X_0 \right) \right). \] (A2)

Moreover, there exists a constant \( c \) depending only on the field of covariances such that
\[ \mathbb{E} \left( \frac{1}{\sqrt{N}} \sum_{s \in \mathbb{R}_n} \left( f \left( X_s \right) - \mathbb{E} \left[ f \left( X_s \right) \right] \right) \right)^2 \leq c \text{Var} \left( f \left( X_0 \right) \right) \] (A3)
for each \( n \).

We define the truncated version of the population scores in (4) by
\[ \tilde{Y}_{m,s} = \sum_{\|k\| \leq L} \langle X_s - k, \phi_{m,k} \rangle, \quad s \in \mathbb{Z}^d. \]

If \( (X_s)_{s \in \mathbb{Z}^d} \) is a Gaussian process, then \( (Y_{m,s})_{s \in \mathbb{Z}^d} \) and \( (\tilde{Y}_{m,s})_{s \in \mathbb{Z}^d} \) are also Gaussian processes. This is not the case for the \( \hat{Y}_{m,s} \), as the filters \( \hat{\phi}_{m,k} \) are estimated, so the \( \hat{Y}_{m,s} \) are nonlinear functions of the observations \( X_s \).

In the following, we assume that the value of \( p \) has been selected. To simplify the formulas, we will assume from now on that the data are centred, i.e., \( \mathbb{E} X_s = 0 \). This has no impact on the arguments. To further lighten the notation, we omit the index \( m \) in the subscripts of the objects related to the level \( m \) SFPCs and simply write \( Y_s \) for \( Y_{m,s} \), \( \phi_k \) for \( \phi_{m,k} \), and so on. We will use
\[ S_n := \sqrt{N} \hat{m}^2_1, \quad \mathcal{K}_n := \sqrt{N} \left( \hat{m}^2_4 - 3 \left( \hat{m}^2_2 \right)^2 \right), \quad \text{with } Z_s := Y_s - \hat{m}_1. \]
\[ \tilde{S}_n := \sqrt{N} \tilde{m}^2_3, \quad \tilde{\mathcal{K}}_n := \sqrt{N} \left( \tilde{m}^2_4 - 3 \left( \tilde{m}^2_2 \right)^2 \right), \quad \text{with } \tilde{Z}_s := \tilde{Y}_s - \tilde{m}_1. \]

Our first lemma establishes asymptotic distributions of population quantities, which approximate the corresponding statistics at a level \( m \).

**Lemma 1.** Suppose \( (X_s)_{s \in \mathbb{Z}^d} \) is a Gaussian process satisfying Assumptions 1 and 4 and \( \min_{1 \leq i \leq d} n_i \to \infty \). Then the sums
\[ \frac{1}{\sqrt{N}} \sum_{s \in \mathbb{R}_n} Y_s, \quad \frac{1}{\sqrt{N}} \sum_{s \in \mathbb{R}_n} \left( y_s^2 - \mathbb{E} \left[ y_s^2 \right] \right), \quad \frac{1}{\sqrt{N}} \sum_{s \in \mathbb{R}_n} y_s^3, \quad \frac{1}{\sqrt{N}} \sum_{s \in \mathbb{R}_n} \left( y_s^4 - \mathbb{E} \left[ y_s^4 \right] \right) \] (A4)
are asymptotically normal and
\[ \left( S_n \mathcal{K}_n \right) \overset{d}{\to} N_2 \left( \left( 0 \right), \left( \begin{array}{cc} 6 \sigma_y^2 & 0 \\ 0 & 24 \sigma_k^2 \end{array} \right) \right), \] (A5)
where
\[ \sigma_y^2 = \sum_{h \in \mathbb{Z}^d} \gamma_h^3, \quad \text{and} \quad \sigma_k^2 = \sum_{s \in \mathbb{Z}^d} \gamma_h^4. \] (A6)
Furthermore, it holds for \( k \leq 4 \) that

\[
\frac{1}{\sqrt{N}} \sum_{s \in \mathbb{R}_n} \left( Y^k_s - \tilde{Y}^k_s - \mathbb{E} \left[ Y^k_s - \tilde{Y}^k_s \right] \right)^2 \to 0, \quad \text{if } N \to \infty.
\]  \hspace{1cm} (A7)

**Proof.** We prove this Lemma by using Theorem 3. Its conditions can be easily verified, as they mostly concern the summability of covariances, which follows from our assumptions. The asymptotic normality of the sums in (A4) is a simple application of Theorem 3. Let us denote \( \gamma = \mathbb{E}[Y^3_0] \). Gaussianity implies \( \mathbb{E}[Y^3_0] = 0 \) and \( \mathbb{E}[Y^4_0] = 3 \gamma^2 \). As for statement (A5), we first show that

\[
\sqrt{N} \left( \hat{m}_3^Y - 3 \gamma \hat{m}_1^Y \right)^d \to N_2 \left( (0,0), \left( \begin{array}{cc} 6\sigma^2_S & 0 \\ 0 & 24 \sigma^2_K \end{array} \right) \right)
\]

by employing the Cramér–Wold device as follows: Let \( \lambda_1, \lambda_2 \in \mathbb{R} \), and again using Theorem 3, we have

\[
\frac{1}{\sqrt{N}} \sum_{s \in \mathbb{R}_n} \left( \lambda_1 \left( Y^3_0 - 3 \gamma Y_0 \right) + \lambda_2 \left( Y^4_0 - 6 \gamma Y^2_0 + 3 \gamma^2 \right) \right)^d \to N (0, \tau^2),
\]

where

\[
\tau^2 = \sum_{h \in \mathbb{Z}^d} \mathbb{E} \left[ \left( \lambda_1 \left( Y^3_h - 3 \gamma Y_h \right) + \lambda_2 \left( Y^4_h - 6 \gamma Y^2_h + 3 \gamma^2 \right) \right) \right] \times \left( \lambda_1 \left( Y^3_h - 3 \gamma Y_h \right) + \lambda_2 \left( Y^4_h - 6 \gamma Y^2_h + 3 \gamma^2 \right) \right).
\]

Under Gaussianity of \( (X_s)_s \), \( (Y_s)_s \) is also a Gaussian process and the cross-covariance between the two summands is evidently zero. Isserlis’s theorem (Theorem 2) then yields the simple forms of 6 \( \sigma^2_S \) and 24 \( \sigma^2_K \), since all the lower order terms cancel.

Now we can see from (A4) that \( \hat{m}_k^Y \), for \( k \leq 4 \), are \( O_p(N^{-1/2}) \). Expanding the sums, we can deduce that

\[
S_n = \sqrt{N} \hat{m}_3^Z = \sqrt{N} \left( \hat{m}_3^Y - 3 \gamma \hat{m}_1^Y \right) + O_p \left( N^{-1/2} \right),
\]

\[
K_n = \sqrt{N} \left( \hat{m}_3^Y - 3 \left( \hat{m}_2^Y \right)^2 \right) = \sqrt{N} \left( \hat{m}_3^Y - 6 \gamma \hat{m}_2^Y + 3 \gamma^2 \right) + O_p \left( N^{-1/2} \right),
\]

which amounts to convergence in probability. This proves (A5).

We prove (A7) by using the last statement of Theorem 3, which yields that there exists a constant \( c \) such that

\[
\frac{1}{N} \mathbb{E} \left( \sum_{s \in \mathbb{R}_n} \left( Y^k_s - \tilde{Y}^k_s - \mathbb{E} \left[ Y^k_s - \tilde{Y}^k_s \right] \right)^2 \right) \leq c \text{ Var} \left( Y^k_0 - \tilde{Y}^k_0 \right).
\]

This constant \( c \) does not depend on \( N \), but only on the sequence of covariances \( \text{Cov} \left( Y^k_h - \tilde{Y}^k_h, Y^k_0 - \tilde{Y}^k_0 \right) \).

What is still left to show is that it is possible to choose \( c \) such that it is also not dependent on \( k \) or \( L \). For this, we follow the argument of Górecki et al. (2018).

The convergence rate of \( Y_s - \tilde{Y}_s \) is

\[
|\tilde{Y}_{m,s} - Y_{m,s}| = O_p(H_m(L)),
\]  \hspace{1cm} (A8)
and it holds that
\[
\text{Var}(Y_s - \tilde{Y}_s) = O\left( H_m(L)^2 \right).
\] (A9)

Since \( Y_0 \) and \( \tilde{Y}_0 \) are Gaussian, \( \text{Var} \left( Y^k_0 - \tilde{Y}^k_0 \right) \xrightarrow{N \to \infty} 0 \) follows easily.

Before we continue, we define the notation \( v_p(X) = (E\|X\|^p)^{1/p} \), which we will need in the following proofs. The next lemma shows that the statistics at each level \( m \) are close to their population counterparts.

**Lemma 2.** Suppose \((X_s)_{s \in \mathbb{Z}_d}\) is a Gaussian process, which satisfies Assumptions 1 and 4, and that Assumption 5 holds. Then, the following bounds hold as \( N \to \infty \). If \( L \to \infty \), then
\[
S_n - \tilde{S}_n = o_p(1), \quad K_n - \tilde{K}_n = o_p(1).
\] (A10)

If, in addition, Assumptions 2 and 3 are satisfied, then we can choose \( L = L(N) \to \infty \) such that \( L^d G(N) \overset{P}{\to} 0 \), and it holds that
\[
\tilde{S}_n - S_n = o_p(1), \quad \tilde{K}_n - K_n = o_p(1).
\] (A11)

**Proof.** We begin by verifying the first relation in (A10). Observe that
\[
S_n - \tilde{S}_n = \sqrt{N} \left( \hat{m}_3^Y - \hat{m}_3^\tilde{Y} \right) - 3\sqrt{N} \hat{m}_2^Y \left( \hat{m}_1^Y - \hat{\tilde{m}}_1^\tilde{Y} \right) + 3\sqrt{N} \hat{m}_2^\tilde{Y} \left( \hat{m}_1^\tilde{Y} - \hat{m}_1^Y \right) + 2\sqrt{N} \left( \hat{m}_1^Y \right)^3 - \left( \hat{m}_1^\tilde{Y} \right)^3.
\]

By Lemma 1, the first three summands converge to zero in probability. The last summand can be bounded by \( 6\sqrt{N} |\hat{m}_1^Y - \hat{\tilde{m}}_1^\tilde{Y}| \left( |\hat{m}_1^Y| + |\hat{m}_1^\tilde{Y}| \right)^2 \), which converges to zero in probability, too.

The second difference in (A10) can be handled similarly:
\[
\begin{align*}
K_n - \tilde{K}_n & = \sqrt{N} \left( \hat{m}_4^Y - \hat{\tilde{m}}_4^\tilde{Y} \right) - 4\sqrt{N} \hat{m}_3^Y \left( \hat{m}_1^Y - \hat{\tilde{m}}_1^\tilde{Y} \right) - 4\sqrt{N} \hat{m}_1^Y \left( \hat{m}_3^Y - \hat{m}_3^\tilde{Y} \right) \\
& \quad + 12\sqrt{N} \left( \hat{m}_1^Y \right)^2 \left( \hat{m}_2^Y - \hat{\tilde{m}}_2^\tilde{Y} \right) + 12\sqrt{N} \hat{m}_2^\tilde{Y} \left( \hat{m}_1^\tilde{Y} \right)^2 - \left( \hat{m}_1^\tilde{Y} \right)^2 \\
& \quad - 3\sqrt{N} \left( \hat{m}_2^\tilde{Y} - \hat{\tilde{m}}_2^\tilde{Y} \right) \left( \hat{m}_2^\tilde{Y} + \hat{\tilde{m}}_2^\tilde{Y} \right) - 6\sqrt{N} \left( \hat{m}_1^Y \right)^4 - \left( \hat{m}_1^\tilde{Y} \right)^4.
\end{align*}
\]

Here again, each summand converges to zero in probability.

Next we turn to the first relation in (A11). For this, we first recall the convergence rate for the estimators of the filter functions. Under Assumptions 2 and 3, it holds that
\[
\max_{k \in \mathbb{Z}_d} \left\| \phi_k - \hat{\phi}_k \right\| = O(G(N)).
\] (A12)

For an exact proof of this, we refer to Kuenzer, Hörmann & Kokoszka (2020). Now we can establish the required convergence rates of \( \hat{m}_k^\tilde{Y} - \hat{m}_k^\tilde{Y} \) for \( k \leq 3 \).
For $k = 1$

$$\sqrt{N}(\hat{m}_1^y - \hat{m}_1^\gamma) = \frac{1}{\sqrt{N}} \sum_{s \in R_n} \sum_{\|k\|_{\infty} \leq L} \langle X_{s-k}, \hat{\phi}_k - \phi_k \rangle$$

$$\mathbb{E} \left\| \sqrt{N}(\hat{m}_1^y - \hat{m}_1^\gamma) \right\| \leq \mathbb{E} \left[ \left\| \frac{1}{\sqrt{N}} \sum_{s \in R_n} X_s \right\|^2 \right]^{1/2} \sum_{\|k\|_{\infty} \leq L} \mathbb{E} \left[ \|\hat{\phi}_k - \phi_k\|^2 \right]^{1/2}$$

$$\leq \left( \sum_{h \in \mathbb{Z}^d} \text{Tr}(C_h^X) \right)^{1/2} \sum_{\|k\|_{\infty} \leq L} \nu_2(\hat{\phi}_k - \phi_k)$$

$$= O(\mathbb{E}[L^d G(N)]) .$$

By assumption, the last expression converges to zero.

For $k = 2$

$$\left\| \hat{m}_2^y - \hat{m}_2^\gamma \right\| = \left\| \frac{1}{N} \sum_{s \in R_n} \left( \hat{Y}_s - \tilde{Y}_s \right) \left( \hat{Y}_s + \tilde{Y}_s \right) \right\|$$

$$\leq \sup_{k \in \mathbb{Z}^d} \|\hat{\phi}_k - \phi_k\| \left\| \sum_{\|k\|_{\infty} \leq L} \left( \frac{1}{N} \sum_{s \in R_n} \left( \hat{Y}_s + \tilde{Y}_s \right) X_{s-k} \right) \hat{\phi}_k - \phi_k \right\| .$$

It is easy to see that

$$\mathbb{E} \left\| \sum_{s \in R_n} \left( \hat{Y}_s + \tilde{Y}_s \right) X_{s-k} \right\| \leq \frac{1}{N} \sum_{s \in R_n} \nu_2(\hat{Y}_s + \tilde{Y}_s) \nu_2(X_{s-k})$$

$$= \left( \sup_{s \in R_n} \nu_2(\hat{Y}_s) + \nu_2(\tilde{Y}_s) \right) \nu_2(X_0) .$$

The last expression is uniformly bounded because $L \to \infty$ and $L^d G(N) \overset{p}{\to} 0$. Therefore

$$\left\| \hat{m}_2^y - \hat{m}_2^\gamma \right\| = O_p(L^d G(N)) ,$$

which converges to zero in probability.

For $k = 3$

$$\sqrt{N}(\hat{m}_3^y - \hat{m}_3^\gamma) = \frac{1}{\sqrt{N}} \sum_{s \in R_n} \left( \hat{Y}_s - \tilde{Y}_s \right)^3 + \frac{3}{\sqrt{N}} \sum_{s \in R_n} \left( \hat{Y}_s - \tilde{Y}_s \right)^2 \tilde{Y}_s$$

$$+ \frac{3}{\sqrt{N}} \sum_{s \in R_n} \left( \hat{Y}_s - \tilde{Y}_s \right) \tilde{Y}_s^2 .$$

DOI: 10.1002/cjs.11662
Following the earlier calculations, we now need to show only the uniform boundedness in probability of the following sums:

$$\left\| \frac{1}{\sqrt{N}} \sum_{s \in \mathcal{R}_n} \bar{Y}_s^2 X_{s-k} \right\|, \quad \left\| \frac{1}{\sqrt{N}} \sum_{s \in \mathcal{R}_n} \bar{Y}_s X_{s-j} \otimes X_{s-k} \right\|, \quad \left\| \frac{1}{\sqrt{N}} \sum_{s \in \mathcal{R}_n} (\hat{Y}_s - \bar{Y}_s) X_{s-j} \otimes X_{s-k} \right\|_S,$$

where $\| \cdot \|_S$ denotes the Hilbert–Schmidt norm defined by $\| \Psi \|_S = \int \psi^2(u, v) \, du \, dv$.

The boundedness can be shown along the lines of Lemma B.3 in Górecki et al. (2018). All that is needed for this is stationarity and Isserlis’s theorem.

We can now plug in these properties into

$$\hat{S}_n - \tilde{S}_n = \sqrt{N} \left( \hat{m}_3 - \hat{m}_3^\gamma \right) - 3\sqrt{N} \hat{m}_2 \left( \hat{m}_1^\gamma - \hat{m}_1 \right)$$

$$+ 3\sqrt{N} \hat{m}_1^\gamma \left( \hat{m}_2^\gamma - \hat{m}_2 \right) + 2\sqrt{N} \left( \left( \hat{m}_1^\gamma \right)^3 - \left( \hat{m}_1 \right)^3 \right)$$

and see that it converges to zero in probability. This concludes the verification of the first bound in (A11).

It remains to verify the second bound in (A11), i.e., $\hat{K}_n - \tilde{K}_n = o_P(1)$. Observe that

$$\hat{K}_n - \tilde{K}_n = \sqrt{N} \left( \hat{m}_4^\gamma - \hat{m}_4 \right) - 4\sqrt{N} \hat{m}_3 \left( \hat{m}_1^\gamma - \hat{m}_1 \right) - 4\sqrt{N} \hat{m}_1 \left( \hat{m}_3^\gamma - \hat{m}_3 \right)$$

$$+ 12\sqrt{N} \left( \hat{m}_1^\gamma \right)^2 \left( \hat{m}_2^\gamma - \hat{m}_2 \right) + 12\sqrt{N} \hat{m}_2 \left( \left( \hat{m}_1^\gamma \right)^2 - \left( \hat{m}_1 \right)^2 \right)$$

$$- 3\sqrt{N} \left( \hat{m}_2^\gamma - \hat{m}_2 \right) \left( \hat{m}_2 + \hat{m}_2 \right) - 6\sqrt{N} \left( \left( \hat{m}_1^\gamma \right)^4 - \left( \hat{m}_1 \right)^4 \right).$$

We have shown that

$$\sqrt{N} \left( \hat{m}_1^\gamma - \hat{m}_1 \right) \overset{p}{\to} 0, \quad \hat{m}_2^\gamma - \hat{m}_2 \overset{p}{\to} 0, \quad \sqrt{N} \left( \hat{m}_3^\gamma - \hat{m}_3 \right) \overset{p}{\to} 0.$$

This implies that

$$\hat{K}_n - \tilde{K}_n = \sqrt{N} \left( \hat{m}_4^\gamma - \hat{m}_4 - 3(\hat{m}_2^\gamma)^2 + 3(\hat{m}_2^\gamma)^2 \right) + o_P(1).$$

If we denote $\hat{D}_s = \hat{Y}_s - \bar{Y}_s$, the first term of this can be decomposed into the following terms:

$$\frac{1}{\sqrt{N}} \sum_{s \in \mathcal{R}_n} \hat{D}_s^2 \left( \hat{D}_s^2 - 3\hat{m}_2^\gamma \right), \quad (A13)$$

$$\frac{4}{\sqrt{N}} \sum_{s \in \mathcal{R}_n} \hat{D}_s \bar{Y}_s \left( \hat{D}_s^2 - 3\hat{m}_2^\gamma \right), \quad (A14)$$

$$\frac{6}{\sqrt{N}} \sum_{s \in \mathcal{R}_n} \left( \hat{D}_s^2 \bar{Y}_s^2 - \hat{D}_s^2 \hat{m}_2^\gamma - 2\hat{D}_s \bar{Y}_s \hat{m}_1^\gamma \right), \quad (A15)$$

$$\frac{4}{\sqrt{N}} \sum_{s \in \mathcal{R}_n} \hat{D}_s \bar{Y}_s \left( \bar{Y}_s^2 - 3\hat{m}_2^\gamma \right). \quad (A16)$$
We will show the convergence to zero in probability of the first term. The others will follow by analogy. The sums can be uniformly bounded in probability by first isolating the estimated filter functions from the rest and then taking the expected value of the remaining sums.

To write out the proof concisely, we will rely on tensor notation of higher order. We rewrite (A13) as follows, isolating the estimated filter functions from the random field:

\[ \sum_{k_1,k_2,k_3,k_4} \left\langle A_{k_1,k_2,k_3,k_4} \left( (\hat{\phi}_{k_2} - \phi_{k_2}) \otimes (\hat{\phi}_{k_3} - \phi_{k_3}) \otimes (\hat{\phi}_{k_4} - \phi_{k_4}) \right) \right\rangle, \quad (A17) \]

where the operator at the centre of this expression is defined by

\[ A_{k_1,k_2,k_3,k_4} = \frac{1}{N^{3/2}} \sum_{s,t \in \mathbb{R}^n} \left( X_{s-k_1} \otimes X_{s-k_2} \otimes X_{s-k_3} - X_{s-k_1} \otimes X_{s-k_2} \otimes X_{t-k_1} \otimes X_{t-k_2} - X_{s-k_1} \otimes X_{t-k_2} \otimes X_{s-k_3} \otimes X_{t-k_4} \right). \]

From (A12), it follows that if \( \|A_{k_1,k_2,k_3,k_4}\| = O_P(1) \), then (A13) is \( o_P(1) \).

Isserlis’s theorem implies that for all jointly Gaussian elements \( U_1, U_2, U_3, U_4 \), we have

\[ \mathbb{E} \left[ U_1 \otimes U_2 \otimes U_3 \otimes U_4 \right] = C_{U_1,U_2} \otimes C_{U_3,U_4} + C_{U_1,U_3} \widetilde{\otimes} C_{U_4,U_2} + C_{U_1,U_4} \widetilde{\otimes}_\tau C_{U_3,U_2}, \]

where \( \widetilde{\otimes} \) denotes the Kronecker product and \( \widetilde{\otimes}_\tau \) denotes the transposed Kronecker product on the operator space.

\[ (A \otimes B) \; C := ACB^*, \quad (A \otimes_\tau B) \; C := AC^*B^*, \]

On simple tensors, they rearrange the order of the tensor product:

\[ (a \otimes b) \otimes (c \otimes d) = (a \otimes c) \widetilde{\otimes} (b \otimes d) = (a \otimes d) \widetilde{\otimes}_\tau (b \otimes c). \]

For trace-class operators, \( \text{Tr} \left( A \otimes B \right) = \text{Tr}(A) \text{Tr}(B) \).

We can now take the mean of the operator \( A_{k_1,k_2,k_3,k_4} \). Because of stationarity, the summation indices \( s \) and \( t \) can be replaced by the spatial lag \( h \). All summands without this lag \( h \) in the index cancel out.

\[ \mathbb{E} \left[ A_{k_1,k_2,k_3,k_4} \right] = \frac{-1}{\sqrt{N}} \prod_{i=1}^{d} \left( 1 - \frac{|h_i|}{n_i} \right) C_{k_2-k_1+h} \otimes C_{k_2-k_1-h} + C_{k_2-k_1+h} \widetilde{\otimes}_\tau C_{k_2-k_1-h} \]

\[ + C_{k_1-k_2+h} \otimes C_{k_1-k_2-h} + C_{k_1-k_2+h} \widetilde{\otimes}_\tau C_{k_1-k_2-h} \]

\[ + C_{k_2-k_1+h} \otimes C_{k_1-k_2-h} + C_{k_2-k_1+h} \widetilde{\otimes}_\tau C_{k_2-k_1-h} \].

From the summability of the covariances, we then know that \( \|\mathbb{E}[A_{k_1,k_2,k_3,k_4}]\| = O(N^{-1/2}) \), independently from \( k_1, k_2, k_3, k_4 \).

Using the same argument, we also show that \( \mathbb{E}\left[\|A_{k_1,k_2,k_3,k_4}\|^2\right] \) is bounded. Therefore, (A13) is \( o_P(1) \).
Conveniently, the convergence of (A14) follows easily from the argument on (A13), by noting that (A14) can be rewritten as
\[ 4 \sum_{k_1,k_2,k_3,k_4} \left\langle A_{k_1,k_2,k_3,k_4} \left( \left( \hat{\phi}_{k_2} - \phi_{k_2} \right) \otimes \left( \hat{\phi}_{k_3} - \phi_{k_3} \right) \otimes \left( \hat{\phi}_{k_4} - \phi_{k_4} \right), \phi_{k_1} \right) \rightangle, \]
which has the same structure as (A17). Because of the summability of the filter functions \( \phi_{k_1}, \ldots, \phi_{k_4} \), (A14) is then \( o_P(1) \). For (A15) and (A16), this approach can be iterated.

\[ \text{Proof of Theorem 1.} \] Because of Assumption 2, we can estimate the spectral density in a way such that \( G(N) \xrightarrow{P} 0 \). To achieve convergence of the estimated scores, we choose the truncation index \( L = \mathbb{L}(N) \xrightarrow{N \to \infty} \infty \) in a way such that \( \mathbb{L}^d G(N) \xrightarrow{P} 0 \). It is then clear that the estimators of the scores are consistent.

Under the assumption that \( (X_s) \) is a Gaussian random field, the population scores are also Gaussian. This means that \( Y_{m,s} \) and \( Y_{m',s} \) (for \( m \neq m' \)) are not only uncorrelated but independent. This independence also holds for the vectors \( \left( S_n^{(m)} \ K_n^{(m)} \right)^\top \). Combining this property with Lemma 1, we see that the vector \( \left( S_n^{(1)} \ K_n^{(1)} \ldots S_n^{(p)} \ K_n^{(p)} \right)^\top \) converges in distribution to a vector of independent normal random variables. Scaling with the asymptotic variances, it follows that
\[ T_p := \sum_{m=1}^{p} \frac{(S_n^{(m)})^2}{6 \sigma_{S,m}^2} + \frac{(K_n^{(m)})^2}{24 \sigma_{K,m}^2} \xrightarrow{d} \chi_{2p}^2. \]
Because Lemma 2 is also applicable, \( \hat{S}_n^{(m)} - S_n^{(m)} \xrightarrow{P} 0 \) and \( \hat{K}_n^{(m)} - K_n^{(m)} \xrightarrow{P} 0 \). The estimators \( \hat{\sigma}_{S,m}^2 \) and \( \hat{\sigma}_{K,m}^2 \) of the variances are assumed to be consistent. We can thus replace the population values with their estimators, incurring only an asymptotically negligible error, i.e.
\[ |\hat{T}_p - T_p| \xrightarrow{P} 0. \]
Therefore, \( \hat{T}_p \xrightarrow{d} \chi_{2p}^2 \).

\[ \text{Proof of Proposition 1.} \] We only consider the case of \( \hat{\sigma}_{S,m}^2 \), as the calculations for \( \hat{\sigma}_{K,m}^2 \) are similar. Before we start the calculations, we want to remark that in this proof we need to deal simultaneously with the estimation of the scores, the truncation of the summation, and the bias of the autocovariance estimators. This simultaneity requires the calculations to be somewhat convoluted. Observe that
\[ |\hat{\sigma}_{S,m}^2 - \sigma_{S,m}^2| \leq \sum_{\|k\|_{\infty} \leq L'} |\hat{r}_k^3 - r_k^3| + \sum_{\|k\|_{\infty} > L'} |\hat{r}_k| \leq \sum_{\|k\|_{\infty} \leq L'} |\hat{r}_k - r_k|^3 + 3 \sum_{\|k\|_{\infty} \leq L'} |\hat{r}_k - r_k|^2 |\hat{r}_k| + 3 \sum_{\|k\|_{\infty} \leq L'} |\hat{r}_k - r_k| |\hat{r}_k|^2 + \sum_{\|k\|_{\infty} > L'} |r_k|^3. \]
The summability of $\gamma_k$ implies that the last sum vanishes if $L' \to \infty$. We now take the first sum as an example to show how the convergence of each sum can be shown. The calculations are similar to the previous proofs in this section, and we will omit the terms $\hat{m}_1^m$ and $\tilde{m}_1^m$ for notational simplicity, as they are asymptotically negligible. Define

$$\tilde{c}_k := \frac{1}{N} \sum_{s \in M_{h,n}} \tilde{Y}_{s+k} \tilde{Y}_s$$

and

$$\tilde{D}_s := \tilde{Y}_s - \bar{Y}_s.$$

Then

$$\sum_{\|k\|_\infty \leq L'} |\hat{Y}_k - \gamma_k|^3 \leq \frac{a}{N^3} \sum_{\|k\|_\infty \leq L'} \left| \sum_s \hat{D}_{s+k} \hat{D}_s \right|^3 + \frac{a}{N^3} \sum_{\|k\|_\infty \leq L'} \left| \sum_s \hat{D}_{s+k} \gamma_s \right|^3$$

$$\quad + \frac{a}{N^3} \sum_{\|k\|_\infty \leq L'} \left| \sum_s Y_{s+k} \tilde{D}_s \right|^3 + a \sum_{\|k\|_\infty \leq L'} |\tilde{c}_k - \gamma_k|^3,$$

where $a$ is a fixed finite constant. Looking at the first sum, we see that

$$\frac{1}{N^3} \sum_{\|k\|_\infty \leq L'} \left| \sum_s \hat{D}_{s+k} \hat{D}_s \right|^3 = \frac{1}{N^3} \sum_{\|k\|_\infty \leq L'} \sum_s \sum_s \sum_s \sum_s \left( \sum_X X_{s+k-i} \otimes X_{s-j} \right) \left( \hat{\phi}_j - \phi_j \right) \left( \hat{\phi}_i - \phi_i \right)^3$$

$$\leq \frac{1}{N^3} \sup_{k \in \mathbb{Z}^d} \|\hat{\phi}_k - \phi_k\|^6 \sum_{\|k\|_\infty \leq L'} \sum_s \sum_s \sum_s \sum_s \|X_{s+k-i} \otimes X_{s-j}\|^3$$

$$\leq \left( L^d \sup_{k \in \mathbb{Z}^d} \|\hat{\phi}_k - \phi_k\| \right) \frac{1}{N^3 L^{2d}} \sum_{\|k\|_\infty \leq L'} \sum_s \sum_s \sum_s \sum_s \|X_{s+k-i} \otimes X_{s-j}\|^3.$$

Here we used an inequality of the form $(\sum_{k=1}^n a_k)^m \leq n^{m-1} \sum_{k=1}^n a_k^m$. The first factor converges to zero in probability by assumption. As for the single summands, stationarity implies that

$$\mathbb{E} \left[ \left\| \sum_X X_{s+k} \otimes X_s \right\|_S^4 \right]$$

$$= \sum_{s_1} \sum_{s_2} \sum_{s_3} \sum_{s_4} \mathbb{E} \left[ \left( X_{s_1+k} X_{s_1+k} X_{s_2} X_{s_2} \right) \left( X_{s_3+k} X_{s_3+k} X_{s_4} X_{s_4} \right) \right]$$

$$= \mathcal{O}(N^2 + N^4 \text{Tr}(C_k)^4).$$

It directly follows that

$$\mathbb{E} \left[ \frac{1}{N^3 L^{2d}} \sum_{\|k\|_\infty \leq L'} \sum_s \sum_s \sum_s \sum_s \|X_{s+k-i} \otimes X_{s-j}\|^3 \right] = \mathcal{O}((L')^{dN^{-3/2}} + \min(L' / L, 1)^d),$$
and in particular we can see that the expression is bounded. Since we assume that \( L^d G(N) \to 0 \), it holds that \( N^{-3} \sum_{\|k\|_\infty \leq L'} \left| \sum_s \hat{D}_{s+k} \hat{D}_s \right|^3 \to 0 \). The two other sums in (A18) can be bounded in a similar way.

What remains is to show that \( \sum_{\|k\|_\infty \leq L'} |\tilde{c}_k - \gamma_k|^3 \to 0 \). In the following calculations, we will use the notation \( c_k := \frac{1}{N} \sum_{s \in M, h, n} Y_{s+k} Y_s \). Note that these “estimators” still exhibit a bias because of the division of the sum by \( N \) which is not the number of summands. Because by assumption \( L' / \min n_i \to 0 \), it follows that the accumulated estimation bias converges to zero and

\[
\mathbb{E} \left[ \sum_{\|k\|_\infty \leq L'} |c_k - \gamma_k|^3 \right] \xrightarrow{N \to \infty} 0.
\]

Because we assume that \((L')^{d/3} H_m(L) \to 0\), it follows via Theorem 2 and (A9) that

\[
\mathbb{E} \left[ \sum_{\|k\|_\infty \leq L'} |\tilde{c}_k - \bar{c}_k|^3 \right] = \mathcal{O}\left( (L')^d H_m(L)^3 \right) \xrightarrow{N \to \infty} 0.
\]

This shows that \( |\hat{\sigma}_S^2 - \sigma_S^2| \to 0 \). The calculations for \( \hat{\sigma}_K^2 \) are analogous. Note that since \((L')^{d/4} \leq (L')^{d/3}\), the requirements on \( L' \) that stem from \( \hat{\sigma}_K^2 \) are fulfilled a fortiori.

\[\[
\]

Received 13 June 2020
Accepted 28 April 2021