The Annals of Statistics
2021, Vol. 49, No. 1, 346–369
https://doi.org/10.1214/20-AOS1959

ASYMPTOTICS FOR SPHERICAL FUNCTIONAL AUTOREGRESSIONS

BY ALESSIA CAPONERA[1] AND DOMENICO MARINUCCI[2]

[1]*Department of Statistical Sciences, Sapienza University of Rome, alessia.caponera@uniroma1.it*
[2]*Department of Mathematics, University of Rome Tor Vergata, marinucc@mat.uniroma2.it*

In this paper, we investigate a class of spherical functional autoregressive processes, and we discuss the estimation of the corresponding autoregressive kernels. In particular, we first establish a consistency result (in mean-square and sup norm), then a quantitative central limit theorem (in Wasserstein distance), and finally a weak convergence result, under more restrictive regularity conditions. Our results are validated by a small numerical investigation.

1. Introduction. In recent years, a lot of interest has been drawn by the statistical analysis of spherical isotropic random fields. These investigations have been motivated by a wide array of applications arising in many different areas, including, in particular, cosmology, astrophysics, geophysics, climate and atmospheric sciences, and many others; see, for example, [2, 3, 9–13, 21, 28, 29]. Most papers in cosmology and astrophysics have focussed so far on spherical random fields with no temporal dependence; the next generation of cosmological experiments is however going to make the time dependence much more relevant. On the other hand, applications in climate, atmospheric sciences, geophysics, and several other areas have always been naturally modelled in terms of a double-dependence in the spatial and temporal domains. In many works of these fields, the attention has been focussed on the definition of wide classes of space-time covariance functions, and then on the derivation of likelihood functions; the literature on these themes is vast and we make no attempt to a complete list of references; see, for instance, [10, 13, 18, 29] and the references therein.

Our purpose in this paper is to investigate a class of space-time processes, which can be viewed as functional autoregressions taking values in $L^2(\mathbb{S}^2)$; we refer to [5] for a general textbook analysis of functional autoregressions taking values in Hilbert spaces, and [1, 16, 25, 26] for a very partial list of some important recent references.

Dealing with functional spherical autoregressions ensures some very convenient simplifications; in particular, we exploit the analytic properties of the standard orthonormal basis of $L^2(\mathbb{S}^2)$ and some natural isotropy requirements to obtain neat expressions for the autoregressive operators, which are then estimated by a form of frequency-domain least squares. For our estimators, we are able to establish rates of consistency (in L^2 and L^∞ norms) and a quantitative version of the central limit theorem, in Wasserstein distance. In particular, we derive explicit bounds for the rate of convergence to the limiting Gaussian distribution by means of the rich machinery of Stein–Malliavin methods (see [24]); to the best of our knowledge, this is the first quantitative central limit theorem established in the framework of functional-valued stationary processes. Under stronger regularity conditions, we are able to establish a weak convergence result for the kernel estimators; our results are then illustrated by simulations.

The plan of our work is then as follows: in Section 2, we present background results on the harmonic analysis of spherical random fields and on Stein–Malliavin methods. In Section 3, we present our basic model; we show how, under isotropy, the model enjoys a number of symmetry properties which greatly simplify our approach. Our main results are then collected in

Received July 2019; revised December 2019.
MSC2020 subject classifications. Primary 62M15; secondary 60G15, 60F05, 62M40.
Key words and phrases. Spherical functional autoregressions, spherical harmonics, quantitative central limit theorem, Wasserstein distance, weak convergence.

Section 4, where we investigate rates of convergence and the quantitative central limit theorem; we consider also weak convergence in $C_p([-1,1])$, under stronger regularity conditions for the autoregressive kernels. Large parts of the proofs and many auxiliary lemmas, some of possible independent interest, are collected in Sections 5 and in the Appendix (Supplementary Material [8]). Finally, Section 6 provides numerical estimates on the behaviour of our procedures.

2. Background and notation.

2.1. *Spectral representation of isotropic random fields on the sphere.* Let $\{T(x), x \in \mathbb{S}^2\}$ denote a finite variance, isotropic random field on the unit sphere $\mathbb{S}^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$, by which we mean as usual that $T(g\cdot) \stackrel{d}{=} T(\cdot), \forall g \in \mathrm{SO}(3)$ the standard 3-dimensional group of rotations; here, the identity in distribution must be understood in the sense of stochastic processes. For notational simplicity, and without loss of generality, we will assume in the sequel that $\mathbb{E}[T(x)] = 0$. It is well known that the following representation holds, in the mean-square sense:

$$T(x) = \sum_{\ell=0}^{\infty} T_\ell(x), \quad T_\ell(x) = \sum_{m=-\ell}^{\ell} a_{\ell,m} Y_{\ell,m}(x), \tag{1}$$

where $\{Y_{\ell,m}(\cdot), \ell \geq 0, m = -\ell, \dots, \ell\}$ is the standard basis of spherical harmonics, which satisfy (for $\varphi \in [0, 2\pi)$, $\vartheta \in 0, \pi]$)

$$\Delta_{\mathbb{S}^2} Y_{\ell,m} = -\ell(\ell+1) Y_{\ell,m}, \quad \Delta_{\mathbb{S}^2} := \frac{1}{\sin\vartheta} \frac{\partial}{\partial\vartheta}\left(\sin\vartheta \frac{\partial}{\partial\vartheta}\right) + \frac{1}{\sin^2\vartheta} \frac{\partial}{\partial\varphi^2};$$

also $\{a_{\ell,m}, \ell \geq 0, m = -\ell, \dots, \ell\}$ is a triangular array of zero-mean, real-valued random coefficients whose covariance structure is given by

$$\mathbb{E}[a_{\ell,m} a_{\ell',m'}] = C_\ell \delta_\ell^{\ell'} \delta_m^{m'};$$

here, δ_a^b is the Kronecker delta function, and the sequence $\{C_\ell, \ell \geq 0\}$ represents the angular power spectrum of the field. Throughout this paper, we consider the real-valued basis of spherical harmonics and, therefore, the random coefficients are real-valued random variables for all (ℓ, m) (we refer for instance to [22] for a more detailed discussion on spectral representations on the sphere). Note that the random coefficients $\{a_{\ell,m}\}$ can be obtained by a direct inversion formula from the map $T(\cdot)$, indeed we have

$$a_{\ell,m} := \int_{\mathbb{S}^2} T(x) Y_{\ell,m}(x)\, dx.$$

Here, we recall also the following *addition formula* for spherical harmonics (see [22], equation (3.42)) which entails that, for any $x, y \in \mathbb{S}^2$,

$$\sum_{m=-\ell}^{\ell} Y_{\ell,m}(x) Y_{\ell,m}(y) = \frac{2\ell+1}{4\pi} P_\ell(\langle x, y\rangle), \tag{2}$$

where $\langle x, y\rangle$ denotes the standard inner product in $\mathbb{R}^3$, and $P_\ell(\cdot)$ represents the ℓth Legendre polynomial, defined as usual by

$$P_\ell(t) = \frac{1}{2^\ell \ell!} \frac{d^\ell}{dt^\ell} (t^2 - 1)^\ell, \quad t \in [-1, 1], \ell \geq 0.$$

It is easy to show that $P_\ell(1) = 1$; moreover, the following *duplication property* is satisfied, that is,

$$\int_{\mathbb{S}^2} \frac{2\ell+1}{4\pi} P_\ell(\langle x, y\rangle) \frac{2\ell+1}{4\pi} P_\ell(\langle y, z\rangle)\, dy = \frac{2\ell+1}{4\pi} P_\ell(\langle x, z\rangle).$$

Under isotropy, from (1) and (2) the covariance function $\Gamma(x,y)=\mathbb{E}[T(x)T(y)]$ satisfies

$$\Gamma(x,y)=\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}C_\ell Y_{\ell,m}(x)Y_{\ell,m}(y)$$
$$=\sum_{\ell=0}^{\infty}C_\ell\frac{2\ell+1}{4\pi}P_\ell\bigl(\langle x,y\rangle\bigr)\quad\text{for all } x,y\in\mathbb{S}^2.$$

In the sequel, given any two positive sequences $\{a_k, k\in\mathbb{N}\}$, $\{b_k, k\in\mathbb{N}\}$, we shall write $a_k\sim b_k$ if $\exists c_1,c_2>0$ such that $c_1b_k\le a_k\le c_2b_k$, $\forall k\in\mathbb{N}$. In addition, we will denote with *const* a positive real constant, which may change from line to line; also, we use $\|\cdot\|_{L^2(\mathbb{S}^2)}$ for the usual L^2 norm on the sphere, $\Lambda_{\min}(A)$ and $\Lambda_{\max}(A)$ for the minimum and maximum eigenvalues of the matrix A, respectively, $\|A\|_{\mathrm{op}}$ for the operator norm of A, that is, $\|A\|_{\mathrm{op}}=\sqrt{\lambda_{\max}(A'A)}$, and $\mathrm{Tr}(A)$ for the trace of A.

2.2. *Hermite polynomials and Stein–Malliavin results.* Let us recall the family of Hermite polynomials $\{H_q(\cdot), q\ge 0\}$, defined by

$$H_q(x)=(-1)^q e^{x^2/2}\frac{d^q}{dx^q}e^{-x^2/2},\quad x\in\mathbb{R};$$

for instance, the first few are given by $H_1(x)=x$, $H_2(x)=x^2-1$, $H_3(x)=x^3-3x$ and $H_4(x)=x^4-6x^2+3$. Any finite variance transform of a standard Gaussian random variable X has a representation in terms of Hermite polynomials (see [24], Example 2.2.6, p. 27), that is, for G such that $\mathbb{E}[G^2(X)]<\infty$,

$$G(X)=\sum_{q=0}^{\infty}J_q(G)\frac{H_q(X)}{q!},\quad J_q(G):=\mathbb{E}\bigl[G(X)H_q(X)\bigr];$$

more generally, for any $L^2(\Omega)$-closed linear Gaussian space $\mathcal{X}$, we can write the Stroock–Varadhan decomposition

$$\mathcal{X}=\bigoplus_{q=0}^{\infty}\mathcal{H}_q,$$

where $\mathcal{H}_q$ is the qth order Wiener chaos, that is, the space spanned by qth order Hermite polynomials; see [24], Chapter 2, for more discussions and details.

We shall exploit extensively a very powerful technique, recently discovered by [23], to establish quantitative central limit theorems for sequences of random variables belonging to Wiener chaos. To explain what we mean by a quantitative central limit theorem, we recall first the notion of *Wasserstein distance*, that is, for any two d-dimensional random variables X,Y,

$$d_W(X,Y)=\sup_{h(\cdot):\|h\|_{\mathrm{Lip}}\le 1}\bigl|\mathbb{E}\bigl[h(X)\bigr]-\mathbb{E}\bigl[h(Y)\bigr]\bigr|\quad\text{where } \|h\|_{\mathrm{Lip}}=\sup_{\substack{x\neq y\\ x,y\in\mathbb{R}^d}}\frac{|h(x)-h(y)|}{\|x-y\|},$$

with $\|\cdot\|$ the usual Euclidean norm on $\mathbb{R}^d$, where we assume that $\mathbb{E}|h(X)|<\infty$, $\mathbb{E}|h(Y)|<\infty$ for every $h(\cdot)$. See [24] for a discussion of the main properties of $d_W(\cdot,\cdot)$ and for other examples of probability metrics; here, we recall simply that

$$d_W(X,Y)\le\mathbb{E}\|X-Y\|. \tag{3}$$

It is shown in [24], Theorems 5.2.6 and 5.2.7, p. 99, that for sequences of zero-mean scalar random variables $\{Z_k, k \in \mathbb{N}\}$ belonging to $\mathcal{H}_q$ $(q \geq 2)$ and such that $\mathbb{E}[Z_k^2] = \sigma^2 > 0$, one has the remarkable inequality

$$d_W(Z, Z_k) \leq \frac{1}{\sigma}\sqrt{\frac{2q-2}{3\pi q}\left(\mathbb{E}[Z_k^4] - 3\sigma^4\right)}, \tag{4}$$

where $Z \overset{d}{=} \mathcal{N}(0, \sigma^2)$ (in our proof below we will actually exploit a multivariate extension of this inequality, also given in [24], Theorems 6.2.2 and 6.2.3, p. 121). The inequality in (4) can be proved by means of the so-called Stein–Malliavin approach, which establishes a deep and surprising connection between Malliavin calculus and Stein's equation as a tool for the investigation of limiting distributions. In particular, in view of (4) for sequences that belong to Wiener chaoses the investigation of the asymptotic behaviour of the fourth-moment is enough to investigate not only the validity of a central limit theorem, but also the rate of convergence to the Gaussian limiting distribution.

3. Spherical random fields with temporal dependence. We are now ready to introduce our model of interest. As usual, by space-time spherical random fields we mean a collection of random variables $\{T(x,t), (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ such that, for every $t \in \mathbb{Z}$, the mapping $(x, \omega) \mapsto T(x,t,\omega)$ is $\mathscr{B}(\mathbb{S}^2) \otimes \mathscr{F}$-measurable, for some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The following definition is standard.

DEFINITION 1. $\{T(x,t), (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is 2-weakly isotropic stationary if $\mathbb{E}[T(x,t)]$ is constant $\forall (x,t) \in \mathbb{S}^2 \times \mathbb{Z}$ and the covariance function Γ is a spatially isotropic and temporally stationary function on $(\mathbb{S}^2 \times \mathbb{Z})^2$, that is, there exists $\Gamma_0 : [-1,1] \times \mathbb{Z} \to \mathbb{R}$ such that

$$\Gamma(x,t,y,s) = \Gamma_0(\langle x, y\rangle, t-s) \quad \forall (x,t), (y,s) \in \mathbb{S}^2 \times \mathbb{Z}.$$

In particular, we will focus on Gaussian random fields, where of course weak isotropy and stationarity entails strong isotropy and stationarity, that is, the law of $T(g\cdot, \cdot + \tau)$ is the same as the law of $T(\cdot,\cdot)$, in the sense of processes, for all $g \in SO(3)$ and $\tau \in \mathbb{Z}$. Note that, for (zero-mean) finite variance, isotropic random fields, $T_t(\cdot) \equiv T(\cdot, t)$ is a random function of $L^2(\mathbb{S}^2)$, $t \in \mathbb{Z}$. Thus, for any fixed $t \in \mathbb{Z}$, the following spectral representation holds:

$$T(x,t) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell,m}(t) Y_{\ell,m}(x), \tag{5}$$

where $\{Y_{\ell,m}(\cdot), \ell \geq 0, m = -\ell, \dots, \ell\}$ are spherical harmonics, and $\{a_{\ell,m}(t), \ell \geq 0, m = -\ell, \dots, \ell\}$ (zero-mean) random coefficients which satisfy

$$\mathbb{E}\left[a_{\ell,m}(t) a_{\ell',m'}(s)\right] = C_\ell(t-s)\delta_\ell^{\ell'}\delta_m^{m'}, \quad t, s \in \mathbb{Z}.$$

Note that $\{C_\ell(0), \ell \geq 0\}$ corresponds to the angular power spectrum of the spherical field at a given time point, for which we will simply write $\{C_\ell\}$. As for the isotropic case, for fixed $t, s \in \mathbb{Z}$, the covariance function $\Gamma(x,t,y,s)$ is easily shown to have a spectral decomposition in terms of Legendre polynomials (Schoenberg's theorem, see also [3]), that is, for every $(x,t), (y,s) \in \mathbb{S}^2 \times \mathbb{Z}$,

$$\Gamma(x,t,y,s) = \sum_{\ell=0}^{\infty} C_\ell(t-s) \frac{2\ell+1}{4\pi} P_\ell(\langle x, y\rangle).$$

REMARK 2. By exploiting results from [25], it would also be possible to rewrite (5) by means of the Cramér–Karhunen–Loève representation

$$T(x,t) = \int_{-\pi}^{\pi} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \exp(-i\lambda t) Y_{\ell,m}(x)\, dW_{\ell,m}(\lambda) \quad \text{in } L^2(\Omega),$$

where $\{W_{\ell,m}(\cdot)\}$ is a family of independent complex-valued Gaussian random measures, with mean zero and covariance structure

$$\mathbb{E}\big[W_{\ell,m}(A)\overline{W_{\ell,m}}(B)\big] = \int_{A\cap B} f_\ell(\lambda)\, d\lambda \quad \text{for all } A, B \subset [-\pi,\pi],$$

where $f_\ell(\cdot)$ denotes the spectral density of the process $\{a_{\ell,m}(t), t \in \mathbb{Z}\}$, which is introduced below and satisfies

$$\mathbb{E}\big[a_{\ell,m}(t)a_{\ell,m}(t+\tau)\big] = \int_{-\pi}^{\pi} \exp(i\lambda\tau) f_\ell(\lambda)\, d\lambda.$$

This approach is not pursued here; see also [7] for more discussion and details.

3.1. *Spherical autoregressions.* In this section, we introduce a particular class of space-time spherical random fields, that is, what we call *spherical functional autoregressions*. As usual in the context of autoregressive processes, we start with the definition of a *spherical white noise*.

DEFINITION 3 (Spherical white noise). The space-time spherical random field $\{Z(x,t), (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is a spherical white noise if:

(i) for every fixed $t \in \mathbb{Z}$, $\{Z(x,t), x \in \mathbb{S}^2\}$ has mean zero and covariance function

$$\Gamma_Z(x,y) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_{\ell;Z} P_\ell\big(\langle x,y\rangle\big), \qquad \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_{\ell;Z} < \infty,$$

$\{C_{\ell;Z}\}$ denoting as usual the angular power spectrum of $Z(\cdot,t)$;

(ii) for every $t \neq s$, the random fields $\{Z(x,t), x \in \mathbb{S}^2\}$ and $\{Z(x,s), x \in \mathbb{S}^2\}$ are independent.

REMARK 4. Note that we are writing the spherical white noise as a collection of random variables defined on every pair $(x,t) \in \mathbb{S}^2 \times \mathbb{Z}$. Alternatively, one could introduce $\{Z(\cdot,t)\}$ as a sequence of random elements in a Hilbert space (in our case, corresponding to $L^2(\mathbb{S}^2)$), see [5], p. 72). The two approaches are equivalent here, because throughout this paper we will always be dealing with jointly-measurable mean-square continuous random fields.

DEFINITION 5. A spherical isotropic kernel operator is an application $\Phi : L^2(\mathbb{S}^2) \to L^2(\mathbb{S}^2)$ which satisfies

$$(\Phi f)(x) = \int_{\mathbb{S}^2} k\big(\langle x,y\rangle\big) f(y)\, dy, \quad x \in \mathbb{S}^2,$$

for some continuous $k : [-1,1] \to \mathbb{R}$.

The following representation holds, in the L^2 sense, for the kernel associated with Φ:

$$k\big(\langle x,y\rangle\big) = \sum_{\ell=0}^{\infty} \phi_\ell \frac{2\ell+1}{4\pi} P_\ell\big(\langle x,y\rangle\big). \tag{6}$$

The coefficients $\{\phi_\ell, \ell \geq 0\}$ corresponds to the eigenvalues of the operator Φ and the associated eigenfunctions are the family of spherical harmonics $\{Y_{\ell,m}\}$, yielding

$$\Phi Y_{\ell,m} = \phi_\ell Y_{\ell,m}.$$

Thus, it holds $\sum_\ell (2\ell + 1)\phi_\ell^2 < \infty$, and hence this operator is Hilbert–Schmidt (see, e.g., [17]). In this paper, we shall also consider trace class operators, namely such that $\sum_\ell (2\ell + 1)|\phi_\ell| < \infty$, for which the representation (6) holds pointwise for every $x, y \in \mathbb{S}^2$.

DEFINITION 6. A space-time spherical random field $\{T(x,t), (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is called the spherical autoregressive process of order p (written SPHAR(p)) if there exist p isotropic kernel operators $\{\Phi_1, \dots, \Phi_p\}$ and a spherical white noise $\{Z(x,t), (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ such that

$$T_t(x) - (\Phi_1 T_{t-1})(x) - \cdots - (\Phi_p T_{t-p})(x) - Z_t(x) = 0 \tag{7}$$

for all $(x,t) \in \mathbb{S}^2 \times \mathbb{Z}$, the equality holding both in the $L^2(\Omega)$ and in the $L^2(\mathbb{S}^2 \times \Omega)$ sense.

REMARK 7. It should be noted that the solution process is defined pointwise, that is, for each (x,t) there exists a random variable defined on $(\Omega, \mathscr{F}, \mathbb{P})$ such that the identity (7) holds.

REMARK 8. The definition of spherical functional autoregressions could be given in a more general form than we did here; for instance, it is clearly possible to define the SPHAR(p) process with a sequence of anisotropic "innovation" fields $\{Z(\cdot,t), t \in \mathbb{Z}\}$. However, in the absence of isotropy the spectral representation theorem would no longer hold, and the same notion of random spectral coefficients $\{a_{\ell,m}(t)\}$ may become ill-defined. Similarly, it would also be possible to relax the isotropy assumption on the auto-regressive kernels, for example, considering continuous symmetric functions on $\mathbb{S}^2 \times \mathbb{S}^2$ or more general compact self-adjoint operators (see [5, 17]). In this case, however, we would not be in the position to exploit the harmonic expansion (6). For these reasons, in this paper we just restrict ourselves to the isotropic framework.

Let us define the eigenvalues $\{\phi_{\ell;j}, \ell \geq 0, j = 1, \dots, p\}$ which satisfy

$$\Phi_j Y_{\ell,m} = \phi_{\ell;j} Y_{\ell,m} \quad \text{and} \quad k_j(\langle x, y\rangle) = \sum_{\ell=0}^{\infty} \phi_{\ell;j} \frac{2\ell+1}{4\pi} P_\ell(\langle x, y\rangle).$$

Hence, for 2-weakly isotropic stationary solutions, it holds

$$(\Phi_j T_{t-j})(x) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \phi_{\ell;j} a_{\ell,m}(t-j) Y_{\ell,m}(x),$$

that is, $(\Phi_j T_{t-j})(\cdot)$ admits a spectral representation in terms of spherical harmonics with coefficients $\{\phi_{\ell;j} a_{\ell m}(t-j), \ell \geq 0, m = -\ell, \dots, \ell\}$. Likewise, we obtain

$$a_{\ell,m}(t) = \phi_{\ell;1} a_{\ell,m}(t-1) + \cdots + \phi_{\ell;p} a_{\ell,m}(t-p) + a_{\ell,m;Z}(t); \tag{8}$$

to ensure identifiability, we assume that there exists at least an ℓ such that $\phi_{\ell;p} \neq 0$, so that $\Pr\{(\Phi_p T_t)(\cdot) \neq 0\} > 0, t \in \mathbb{Z}$; see again [5]. Now, define as usual the associated polynomials $\phi_\ell : \mathbb{C} \to \mathbb{C}, \ell \geq 0$:

$$\phi_\ell(z) = 1 - \phi_{\ell;1} z - \cdots - \phi_{\ell;p} z^p. \tag{9}$$

CONDITION 9. The sequence of polynomials (9) is such that

$$|z| \le 1 \quad \Rightarrow \quad \phi_\ell(z) \neq 0.$$

More explicitly, there are no roots in the unit disk, for all $\ell \ge 0$.

REMARK 10. Condition 9, together with the summability of $\{\phi^2_{\ell;j}\}$, ensures that the smallest root taken among all nondegenerate polynomials is bounded away from one. Indeed, if $\xi_{\ell;1}, \dots, \xi_{\ell;d_\ell}$ are the roots of the d_ℓ-degree polynomial (9), $1 \le d_\ell \le p$, then

$$|\xi_{\ell;j}| \ge \xi_* > 1,$$

uniformly over ℓ. Equivalently, there exists $\delta > 0$ such that

$$|z| < 1 + \delta \quad \Rightarrow \quad \phi_\ell(z) \neq 0 \quad \text{for all } \ell \ge 0.$$

As a consequence, equation (7) admits a unique 2-weakly isotropic stationary solution; the proof can be given along the same lines as in [5], and it is omitted for brevity's sake; see [7] for more discussion and details.

EXAMPLE 11 (SPHAR(1)). The family of random variables $\{T(x,t), (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is a spherical autoregressive process of order one if for all pairs $(x,t) \in \mathbb{S}^2 \times \mathbb{Z}$ it satisfies

$$T_t(x) = (\Phi T_{t-1})(x) + Z_t(x).$$

In this case, Condition 9 simply becomes $|\phi_\ell| < 1$, for all $\ell \ge 0$, which is equivalent to ask

$$\|\Phi\|_{\mathrm{op}} := \max_{\ell \ge 0} |\phi_\ell| < 1;$$

see also [5], Section 3.4.

REMARK 12. The autocovariance function of a 2-weakly isotropic stationary SPHAR(1) process is easily seen to be given by (writing $\tau = t - s$),

$$\Gamma(x,t,y,s) = \Gamma_0(\langle x,y\rangle, \tau) = \sum_{\ell=0}^{\infty} C_\ell(\tau) \frac{2\ell+1}{4\pi} P_\ell(\langle x,y\rangle) = \sum_{\ell=0}^{\infty} \frac{\phi_\ell^{|\tau|} C_{\ell;Z}}{1-\phi_\ell^2} \frac{2\ell+1}{4\pi} P_\ell(\langle x,y\rangle).$$

It is easy hence to envisage a number of parametric models for sphere-time covariances; for instance, a simple proposal is

$$\begin{aligned} \phi_\ell &= G \times \{|\ell - \ell^*| + 1\}^{-\alpha_\phi}, \quad \ell^* \ge 0, \alpha_\phi > 2, 0 < G < 1, \\ C_{\ell;Z} &= G_Z (1+\ell)^{-\alpha_Z}, \quad \alpha_Z > 2. \end{aligned} \tag{10}$$

Here, the parameters α_Z and α_ϕ control, respectively, the smoothness of the innovation process and the regularity of the autoregressive kernel (see [20]); the positive integer ℓ^* can be seen as a sort of "characteristic scale", where the power of the kernel is concentrated. More generally, we can take $\phi_\ell = G(\ell; \alpha_1, \dots, \alpha_q)$, where $\alpha_1, \dots, \alpha_q$ are fixed parameters and G is any function such that

$$\sup_\ell |G(\ell; \alpha_1, \dots, \alpha_q)| < 1 \quad \text{and} \quad \sum_\ell (2\ell+1) |G(\ell; \alpha_1, \dots, \alpha_q)| < \infty,$$

uniformly over all values of $(\alpha_1, \dots, \alpha_q)$.

CONDITION 13 (Gaussianity and identifiability). The spherical white noise $\{Z(x,t), (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is Gaussian and such that $C_{\ell;Z} > 0$ for all $\ell = 0, 1, 2, \dots$.

REMARK 14. The previous condition contains an identifiability assumption; indeed, it is simple to verify from our arguments below that for $C_{\ell;Z} = 0$ the component of the kernel corresponding to the ℓth multipole is not observable, that is, the AR(p) process has the same distribution whatever the values of $\{\phi_{\ell;j}, j = 1, \dots, p\}$. It is possible, however, to estimate the "sufficient" version of the kernel, that is, its projection on the relevant subspace, such that $C_{\ell,Z} > 0$. The extension is straightforward and we avoid it just for brevity and notational simplicity. Of course, as a consequence we have that

$$\int_{\mathbb{S}^2\times\mathbb{S}^2} \Gamma_Z(x, y) f(x) f(y)\, dx\, dy > 0 \quad \forall f(\cdot) \in L^2(\mathbb{S}^2),\ f(\cdot) \neq 0.$$

4. Main results. Throughout this paper, we shall assume to be able to observe the projections of the fields on the orthonormal basis $\{Y_{\ell m}\}$, that is, we assume to observe

$$a_{\ell,m}(t) := \int_{\mathbb{S}^2} T(x, t) Y_{\ell,m}(x)\, dx, \quad t = 1, \dots, n.$$

The estimator we shall focus on is a form of least squares regression on an increasing subset of the orthonormal system $\{Y_{\ell,m}\}$; more precisely, we shall define $k(\cdot) := (k_1(\cdot), \dots, k_p(\cdot))'$ for the vector of nuclear kernels, a growing sequence of integers L_N, $L_N \to \infty$ as $N \to \infty$; and a vector of estimators

$$\widehat{k}_N(\cdot) := (\widehat{k}_{1;N}(\cdot), \dots, \widehat{k}_{p;N}(\cdot))' = \arg\min_{k(\cdot)\in\mathcal{P}_N^p} \sum_{t=1}^{N} \left\| T_{t+p} - \sum_{j=1}^{p} \Phi_j T_{t+p-j} \right\|_{L^2(\mathbb{S}^2)}^2, \tag{11}$$

where $N := n - p$, $N > p$, and $\mathcal{P}_N^p$ is the Cartesian product of p copies of

$$\mathcal{P}_N = \operatorname{span}\left\{ \frac{2\ell + 1}{4\pi} P_\ell(\cdot), \ell \le L_N \right\}.$$

As common in the autoregressive context, we drop the first p observations when computing our estimators, in order to avoid initialization issues. We shall write $\mathcal{L}_N(\cdot)$ for the function $\mathcal{L}_N : [-1, 1] \to \mathbb{R}$,

$$\mathcal{L}_N(z) = \sum_{\ell=0}^{L_N} \frac{2\ell + 1}{16\pi^2} P_\ell^2(z), \quad z \in [-1, 1]. \tag{12}$$

Note that

$$\mathcal{L}_N(1) = \mathcal{L}_N(-1) = \sum_{\ell=0}^{L_N} \frac{2\ell + 1}{16\pi^2} = \frac{(L_N + 1)^2}{16\pi^2};$$

on the other hand, for $z \in (-1, 1)$ we have the identity (see [15, 31])

$$\sum_{\ell=0}^{L_N} \frac{2\ell + 1}{16\pi^2} P_\ell^2(z) = \frac{L_N + 1}{16\pi^2} \left[P'_{L_N+1}(z) P_{L_N}(z) - P'_{L_N}(z) P_{L_N+1}(z) \right];$$

it is then possible to show that (see Lemma 4 in the Supplementary Material [8])

$$\mathcal{L}_N(z) \simeq \frac{2L_N}{\pi\sqrt{1 - z^2}} \quad \text{as } L_N \to \infty, \tag{13}$$

where $\simeq$ indicates that the ratio of left- and right-hand sides converges to unity.

For our results to follow, we need slightly stronger assumptions on the "high frequency" behaviour of the kernels $k_j(\cdot)$. More precisely, we shall introduce the following.

CONDITION 15 (Smoothness). For all $j = 1, \dots, p$, there exists positive constants β_j, γ_j such that

$$|\phi_{\ell;j}| \leq \frac{\gamma_j}{\ell^{\beta_j}}, \quad \beta_j > 1, \ell > 0. \tag{14}$$

We let $\beta_* = \min_{j \in \{1,\dots,p\}} \beta_j$. We shall say that this condition is satisfied in the *strong* sense if $\beta_j > 2$, $j = 1, \dots, p$.

REMARK 16. It is readily seen that Condition 15 leads to Hilbert–Schmidt operators, since it implies $\sum_\ell (2\ell + 1)\phi^2_{\ell;j} < \infty$, $j = 1, \dots, p$; whereas the strong version Condition 15 is specific for nuclear operators, since it entails $\sum_\ell (2\ell + 1)|\phi_{\ell;j}| < \infty$, $j = 1, \dots, p$; see again [17].

REMARK 17. Condition 15 is interpretable in terms of the regularity of each kernel $k_j(\cdot)$. Indeed, in [20] it is shown that

$$\sum_{\ell=0}^{\infty} |\phi_{\ell;j}|^2 \frac{2\ell + 1}{4\pi}(1 + \ell^{2\eta}) < \infty$$

implies integrability of the first η derivatives of $k_j(\cdot)$, that is, $k_j(\cdot)$ belongs to the Sobolev space $W_{1,\eta}$.

Our first result refers to the asymptotic consistency of the kernel estimators that we just introduced.

THEOREM 18 (Consistency). *Consider $\widehat{k}_N(\cdot)$ in equation (11). Under Conditions 9, 13 and 15, for $L_N \sim N^d$, $0 < d < 1$, we have that*

$$\mathbb{E}\left[\int_{-1}^{1} \|\widehat{k}_N(z) - k(z)\|^2 \, dz\right] = \mathcal{O}(N^{d-1} + N^{2d(1-\beta_*)}). \tag{15}$$

Moreover, under Conditions 9, 13 and 15 (in the strong *sense), for $L_N \sim N^d$, $0 < d < \frac{1}{3}$,*

$$\mathbb{E}\Big[\sup_{z\in[-1,1]} \|\widehat{k}_N(z) - k(z)\|\Big] = \mathcal{O}(N^{(3d-1)/2} + N^{d(2-\beta_*)}).$$

REMARK 19 (Optimal choice of d). The optimal choice of d, in terms of the best convergence rates, is given by $d^* = \frac{1}{2\beta_*-1}$, leading to the exponents $\frac{2-2\beta_*}{2\beta_*-1}$ and $\frac{2-\beta_*}{2\beta_*-1}$, respectively. Heuristically, the result can be explained as follows: larger values of β_* entail higher regularity/smoothness properties of the kernels to be estimated; as usual in nonparametric estimation, more regular functions can be estimated with better convergence rates, as the bias term is controlled more efficiently. Indeed, for $d = d^*$ and $\beta_* \to \infty$, the mean squared error approximates the parametric rate $1/N$, as expected.

REMARK 20 (Plug-in estimates). For applications to empirical data, the optimal rate can be implemented by means of plug-in techniques, that is, estimating (under additional regularity conditions) the value of the parameter β_* by means of first step-estimators of the coefficients $\{\phi_{\ell,j}\}$. Let us sketch the main ideas for this approach, omitting some details for brevity. Consider for simplicity the SPHAR(1) case, and let us make Condition 15 stronger by assuming that

$$|\phi_\ell| = \frac{\gamma}{\ell^\beta} + o\left(\frac{1}{\ell^\beta}\right) \quad \text{some } \gamma > 0, \beta > 1, \forall \ell > 0.$$

Consider the estimator

$$\widehat{\phi}_{\ell,N} := \frac{\sum_t a_{\ell,m}(t-1)a_{\ell,m}(t)}{\sum_t a_{\ell,m}^2(t-1)}, \quad \ell = 0, 1, 2, \dots,$$

from which we can now build the pseudo log-regression model

$$\log \widehat{\phi}_{\ell,N}^2 = \log \frac{\widehat{\phi}_{\ell,N}^2}{\gamma^2 \ell^{-2\beta}} + \log(\gamma^2 \ell^{-2\beta}) = \log(\gamma^2) - 2\beta \log \ell + v_\ell,$$

$$v_\ell := \log \frac{\widehat{\phi}_{\ell,N}^2}{\gamma^2 \ell^{-2\beta}}, \quad \ell = 0, 1, 2, \dots,$$

where the "regression residuals" $\{v_\ell\}$ are independent over ℓ, with asymptotically mean zero and bounded variance as $N \to \infty$. It is then possible to study the asymptotic consistency of the OLS-like estimator (see also [30] for the related log-periodogram estimator)

$$\widehat{\beta}_N := -\frac{\sum_\ell \{\log \ell \times \log \widehat{\phi}_{\ell,N}^2\}}{2\sum_\ell \{\log \ell\}^2}.$$

The optimal rates can then be consistently estimated by means of the plug-in estimates $\widehat{d}_N^* = \frac{1}{2\widehat{\beta}_N - 1}$.

A more rigorous and complete investigation on these issues is currently in preparation and is not reported here for brevity's sake.

Our second result refers to a quantitative central limit theorem for the kernel estimators. Consider $\widehat{k}_N(\cdot)$ in equation (11) and, for any $m \in \mathbb{N}$, any $z_1, \dots, z_m \in (-1, 1)$, $z_1 \neq \dots \neq z_m$, define the $mp \times 1$ vectors

$$K_N = K_N(z_1, z_2, \dots, z_m) := \begin{pmatrix} \sqrt{\frac{N}{\mathcal{L}_N(z_1)}}(\widehat{k}_N(z_1) - k(z_1)) \\ \vdots \\ \sqrt{\frac{N}{\mathcal{L}_N(z_m)}}(\widehat{k}_N(z_m) - k(z_m)) \end{pmatrix}, \qquad Z \stackrel{d}{=} \mathcal{N}_{mp}(0_{mp}, I_{mp}).$$

THEOREM 21. *Under Conditions 9, 13 and 15 (in the* strong *sense), for $L_N \sim N^d$, $d > \frac{1}{2\beta_* - 2}$, we have that*

$$d_W(Z, K_N) = \mathcal{O}(N^{-1/2} + N^{1/2 + d(1-\beta_*)} + N^{-d} \log N).$$

An immediate corollary is the following.

COROLLARY 22. *Under the same conditions and notation as in Theorem 21, for any fixed $z \in [-1, 1]$, we have that*

$$\sqrt{\frac{N}{\mathcal{L}_N(z)}}(\widehat{k}_N(z) - k(z)) \to \mathcal{N}_p(0_p, I_p), \quad N \to \infty.$$

REMARK 23. As usual, the values of d that guarantee asymptotic normality do not minimize the mean squared error; in fact, we have that $d^* = \frac{1}{2\beta_* - 1} < \frac{1}{2\beta_* - 2}$, which is the minimal value of d for Theorem 21 to hold. Indeed, asymptotic Gaussianity requires undersmoothing, that is, a value of d which makes the asymptotic bias negligible, rather than of the same order as the variance. Once again the rate can be taken to approach $N^{-1/2}$ for $\beta_* \to \infty$.

For our third and final result, we need to strengthen the conditions on the regularity of the autoregressive kernels.

CONDITION 24. The kernel $k_j(\cdot)$ admits a finite expansion in the Legendre basis, that is, there exist an (arbitrary large but finite) integer $L > 0$ such that

$$\int_{-1}^{1} k_j(z) P_\ell(z)\, dz = 0 \quad \text{for all } j = 1, \dots, p \text{ and } \ell > L.$$

Condition 24 clearly implies that there exist finite integers $L_1, \dots, L_p \le L$ such that

$$k_j(z) = \sum_{\ell=0}^{L_j} \frac{2\ell+1}{4\pi} \phi_{\ell;j} P_\ell(z), \quad z \in [-1,1], j = 1, \dots, p;$$

we also need to introduce, for $\ell = 0, 1, 2, \dots$, the $p \times p$ autocovariance matrix

$$\Gamma_\ell := \begin{pmatrix} C_\ell & C_\ell(1) & \cdots & C_\ell(p-1) \\ C_\ell(1) & C_\ell & \cdots & C_\ell(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ C_\ell(p-1) & C_\ell(p-2) & \cdots & C_\ell \end{pmatrix},$$

and we shall write $W_p(\cdot)$ for the zero-mean, p-dimensional Gaussian process with covariance function

$$\Gamma_k(z, z') = \sum_{\ell=0}^{L} C_{\ell;Z} \Gamma_\ell^{-1} \frac{2\ell+1}{16\pi^2} P_\ell(z) P_\ell(z').$$

We are now able to state our last theorem.

THEOREM 25. *Under Conditions 9, 13 and 24, we have that*

$$\sqrt{N}(\widehat{k}_N(\cdot) - k(\cdot)) \Longrightarrow W_p(\cdot), \quad N \to \infty,$$

where $\Longrightarrow$ *denotes weak convergence in* $C_p([-1,1])$ (*the space of continuous functions from* $[-1,1]$ *to* $\mathbb{R}^p$, *with the standard uniform metric*).

REMARK 26. At first sight, it may look surprising that the weak convergence for the estimators in Theorem 25 occurs at a faster rate $\sqrt{N}$ than the convergence in finite-dimensional distributions of Theorem 21. This comparison, however, is misleading; indeed, in Theorem 21 we are not assuming the expansion of the kernels to be finite and, therefore, we need to include a growing number of multi-poles L_N, to ensure that bias terms are asymptotically negligible. On the other hand, note that weak convergence cannot hold under the conditions of Theorem 21, as the limiting finite dimensional distributions correspond to Gaussian independent random variables for any choice of fixed points $(z_1, \dots, z_m)$: no Gaussian process with measurable trajectories can have these finite-dimensional distributions. The limiting distribution is characterized by the nuisance parameters $\{C_\ell, C_\ell(1), \dots, C_\ell(p-1), C_{\ell;Z}\}$; for brevity's sake, estimation of these parameters is deferred to future work.

5. Proofs of the main results. We now present the main arguments of our proofs, which are based on a number of technical results collected in the Appendix (Supplementary Material [8]). For $\ell = 0, 1, 2, \dots$, it is convenient to introduce the $N(2\ell+1)$-dimensional vectors

$$Y_{\ell;N} := (a_{\ell,-\ell}(p+1), \dots, a_{\ell,\ell}(p+1), \dots, a_{\ell,\ell}(n))',$$

$$\boldsymbol{\varepsilon}_{\ell;N} := (a_{\ell,-\ell;Z}(p+1), \dots, a_{\ell,\ell;Z}(p+1), \dots, a_{\ell,\ell;Z}(n))';$$

moreover, let us consider the $N(2\ell+1)\times p$ matrix

$$X_{\ell;N} := \{Y_{\ell;N-1} : Y_{\ell;N-2} : \cdots : Y_{\ell;N-p}\},$$

where

$$Y_{\ell;N-j} := \big(a_{\ell,-\ell}(p+1-j), \dots, a_{\ell,\ell}(p+1-j), \dots, a_{\ell,\ell}(n-j)\big)', \quad j=1,\dots,p.$$

We start from the proof of the consistency results.

PROOF OF THEOREM 18. It is easy to see that we have

$$\begin{aligned}
\widehat{k}_N(\cdot) &= \underset{k(\cdot)\in\mathcal{P}_N^p}{\arg\min} \sum_{t=p+1}^{n} \Big\| T_t - \sum_{j=1}^{p} \Phi_j T_{t-j} \Big\|^2_{L^2(\mathbb{S}^2)} \\
&= \sum_{\ell=0}^{L_N} \widehat{\boldsymbol{\phi}}_{\ell;N} \frac{2\ell+1}{4\pi} P_\ell(\cdot),
\end{aligned}$$

where

$$\begin{aligned}
\widehat{\boldsymbol{\phi}}_{\ell;N} &:= \big(\widehat{\phi}_{\ell;N}(1), \dots, \widehat{\phi}_{\ell;N}(p)\big)' \\
&= \underset{\boldsymbol{\phi}_\ell\in\mathbb{R}^p}{\arg\min} \sum_{t=p+1}^{n} \sum_{m=-\ell}^{\ell} \Big(a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \Big)^2.
\end{aligned}$$

Now, let $r_N(z)$ be the difference between the kernel and its truncated version

$$k_N(z) = \sum_{\ell=0}^{L_N} \boldsymbol{\phi}_\ell \frac{2\ell+1}{4\pi} P_\ell(z),$$

that is,

$$r_N(z) = k(z) - k_N(z) = \sum_{\ell=L_N+1}^{\infty} \boldsymbol{\phi}_\ell \frac{2\ell+1}{4\pi} P_\ell(z).$$

where the equality holds in the L^2 sense. Then

$$\mathbb{E}\Big[\int_{-1}^{1} \|\widehat{k}_N(z) - k(z)\|^2 \, dz\Big] = \mathbb{E}\Big[\int_{-1}^{1} \|\widehat{k}_N(z) - k_N(z)\|^2 \, dz\Big] + \int_{-1}^{1} \|r_N(z)\|^2 \, dz, \tag{16}$$

since $\mathbb{E}[\int_{-1}^{1} \langle \widehat{k}_N(z) - k_N(z), r_N(z)\rangle \, dz] = 0$, from orthogonality of Legendre polynomials.

Now notice that

$$\begin{aligned}
&\int_{-1}^{1} \|\widehat{k}_N(z) - k_N(z)\|^2 \, dz \\
&\quad = \sum_{\ell=0}^{L_N} \sum_{\ell'=0}^{L_N} \langle \widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell, \widehat{\boldsymbol{\phi}}_{\ell';N} - \boldsymbol{\phi}_{\ell'} \rangle \frac{2\ell+1}{4\pi} \frac{2\ell'+1}{4\pi} \int_{-1}^{1} P_\ell(z) P_{\ell'}(z) \, dz \\
&\quad = \sum_{\ell=0}^{L_N} \sum_{\ell'=0}^{L_N} \langle \widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell, \widehat{\boldsymbol{\phi}}_{\ell';N} - \boldsymbol{\phi}_{\ell'} \rangle \frac{2\ell+1}{4\pi} \frac{2\ell'+1}{4\pi} \frac{2}{2\ell+1} \delta_\ell^{\ell'} \\
&\quad = \sum_{\ell=0}^{L_N} \|\widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell\|^2 \frac{2\ell+1}{8\pi^2}.
\end{aligned}$$

Then, from Lemma 2 in the Supplementary Material [8],

$$\mathbb{E}\left[\int_{-1}^{1}\|\widehat{k}_N(z)-k_N(z)\|^2\,dz\right]=\sum_{\ell=0}^{L_N}\mathbb{E}\|\widehat{\boldsymbol{\phi}}_{\ell;N}-\boldsymbol{\phi}_\ell\|^2\frac{2\ell+1}{8\pi^2}\leq \mathrm{const}\frac{L_N+1}{N}.$$

On the other hand,

$$\begin{aligned}\int_{-1}^{1}\|r_N(z)\|^2\,dz&=\sum_{\ell=L_N+1}^{\infty}\sum_{\ell'=L_N+1}^{\infty}\langle\boldsymbol{\phi}_\ell,\boldsymbol{\phi}_{\ell'}\rangle\frac{2\ell+1}{4\pi}\frac{2\ell'+1}{4\pi}\int_{-1}^{1}P_\ell(z)P_{\ell'}(z)\,dz\\&=\sum_{\ell=L_N+1}^{\infty}\sum_{\ell'=L_N+1}^{\infty}\langle\boldsymbol{\phi}_\ell,\boldsymbol{\phi}_{\ell'}\rangle\frac{2\ell+1}{4\pi}\frac{2\ell'+1}{4\pi}\frac{2}{2\ell+1}\delta_\ell^{\ell'}\\&=\sum_{\ell=L_N+1}^{\infty}\|\boldsymbol{\phi}_\ell\|^2\frac{2\ell+1}{8\pi^2}.\end{aligned}$$

Therefore, under Condition 15 and for $L_N\sim N^d$, $0<d<1$, we have

$$\int_{-1}^{1}\|r_N(z)\|^2\,dz=\mathcal{O}(N^{2d(1-\beta_*)})$$

and

$$\mathbb{E}\left[\int_{-1}^{1}\|\widehat{k}_N(z)-k(z)\|^2\,dz\right]=\mathcal{O}(N^{d-1}+N^{2d(1-\beta_*)}),$$

where $\beta_*=\min_{j\in\{1,\dots,p\}}\beta_j$, as claimed.

Under the strong version of Condition 15, each kernel $k_j(\cdot)$ is defined for all $z\in[-1,1]$ as the pointwise limit of its expansion in terms of Legendre polynomials and

$$\mathbb{E}\Big[\sup_{z\in[-1,1]}\|\widehat{k}_N(z)-k(z)\|\Big]\leq\mathbb{E}\Big[\sup_{z\in[-1,1]}\|\widehat{k}_N(z)-k_N(z)\|\Big]+\sup_{z\in[-1,1]}\|r_N(z)\|,$$

by the triangle inequality. Hence, for the first component we have

$$\begin{aligned}\mathbb{E}\left[\sup_{z\in[-1,1]}\left\|\sum_{\ell=0}^{L_N}(\widehat{\boldsymbol{\phi}}_{\ell;N}-\boldsymbol{\phi}_\ell)\frac{2\ell+1}{4\pi}P_\ell(z)\right\|\right]&\leq\sum_{\ell=0}^{L_N}\mathbb{E}\|\widehat{\boldsymbol{\phi}}_{\ell;N}-\boldsymbol{\phi}_\ell\|\frac{2\ell+1}{4\pi}\\&\leq\mathrm{const}\sum_{\ell=0}^{L_N}\frac{\sqrt{2\ell+1}}{\sqrt{N}}\\&\leq\mathrm{const}\frac{(L_N+1)^{3/2}}{\sqrt{N}},\end{aligned}$$

again in view of Lemma 2 in the Appendix (Supplementary Material [8]) and the Cauchy–Schwarz inequality. On the other hand,

$$\sup_{z\in[-1,1]}\|r_N(z)\|\leq\sum_{\ell=L_N+1}^{\infty}\|\boldsymbol{\phi}_\ell\|\frac{2\ell+1}{4\pi}.$$

Therefore, again under the strong version of Condition 15 and for $L_N\sim N^d$, $0<d<\frac{1}{3}$, we have

$$\sup_{z\in[-1,1]}\|r_N(z)\|=\mathcal{O}(N^{d(2-\beta_*)})$$

and thus

$$\mathbb{E}\Big[\sup_{z\in[-1,1]}\big\|\widehat{k}_N(z)-k(z)\big\|\Big]=\mathcal{O}\big(N^{(3d-1)/2}+N^{d(2-\beta_*)}\big),$$

as claimed. □

We are now in the position to establish the quantitative central limit theorem.

PROOF OF THEOREM 21. Let us recall that the minimizing estimator takes the form

$$\begin{aligned}\widehat{k}_N(\cdot)&=\operatorname*{arg\,min}_{k(\cdot)\in\mathcal{P}_N^p}\sum_{t=p+1}^{n}\Big\|T_t-\sum_{j=1}^{p}\Phi_j T_{t-j}\Big\|^2_{L^2(\mathbb{S}^2)}\\&=\sum_{\ell=0}^{L_N}\widehat{\boldsymbol{\phi}}_{\ell;N}\frac{2\ell+1}{4\pi}P_\ell(\cdot),\end{aligned}$$

where

$$\begin{aligned}\widehat{\boldsymbol{\phi}}_{\ell;N}&=\operatorname*{arg\,min}_{\boldsymbol{\phi}_\ell\in\mathbb{R}^p}\sum_{t=p+1}^{n}\sum_{m=-\ell}^{\ell}\Big(a_{\ell,m}(t)-\sum_{j=1}^{p}\phi_{\ell;j}a_{\ell,m}(t-j)\Big)^2\\&=\big(X'_{\ell;N}X_{\ell;N}\big)^{-1}X'_{\ell;N}Y_{\ell;N}=\boldsymbol{\phi}_\ell+\big(X'_{\ell;N}X_{\ell;N}\big)^{-1}X'_{\ell;N}\boldsymbol{\varepsilon}_{\ell;N}.\end{aligned}$$

We shall introduce some more notation:

$$A_{\ell;N}:=\frac{1}{C_\ell N(2\ell+1)}X'_{\ell;N}X_{\ell;N},\qquad \Sigma_\ell:=\mathbb{E}[A_{\ell;N}]=\frac{\Gamma_\ell}{C_\ell},$$

and

$$B_{\ell;N}:=\frac{1}{C_\ell\sqrt{N(2\ell+1)}}X'_{\ell;N}\boldsymbol{\varepsilon}_{\ell;N}.$$

Therefore,

$$\sqrt{N(2\ell+1)}(\widehat{\boldsymbol{\phi}}_{\ell;N}-\boldsymbol{\phi}_\ell)=A^{-1}_{\ell;N}B_{\ell;N}.$$

Heuristically, the proof of the quantitative central limit theorem can be described as follows: in order to be able to exploit Stein–Malliavin techniques, we need to deal with variables belonging to some qth order chaos; now the ratio above does not fulfill this requirement, because $A^{-1}_{\ell;N}$ is a random quantity which does not belong to any $\mathcal{H}_q$. On the other hand, componentwise we have $B_{\ell;N}\in\mathcal{H}_2$, for each ℓ. We shall then show that it is possible to replace $A^{-1}_{\ell;N}$ by its (deterministic) probability limit Σ_ℓ^{-1}, without affecting asymptotic results; because our kernel estimators will be written as linear combinations of $\widehat{\boldsymbol{\phi}}_{\ell;N}$, the proof can be completed by a careful investigation of multivariate fourth-order cumulants.

Let us now make the previous argument rigorous. Let K_N and U_N be two mp-dimensional random vectors, defined as

$$K_N:=\begin{pmatrix}\sqrt{\dfrac{N}{\mathcal{L}_N(z_1)}}\big(\widehat{k}_N(z_1)-k(z_1)\big)\\ \vdots\\ \sqrt{\dfrac{N}{\mathcal{L}_N(z_m)}}\big(\widehat{k}_N(z_m)-k(z_m)\big)\end{pmatrix}$$

and

$$U_N = \begin{pmatrix} U_N(z_1) \\ \vdots \\ U_N(z_m) \end{pmatrix} := \begin{pmatrix} \frac{1}{\sqrt{\mathcal{L}_N(z_1)}} \sum_{\ell=0}^{L_N} \Sigma_\ell^{-1} B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z_1) \\ \vdots \\ \frac{1}{\sqrt{\mathcal{L}_N(z_m)}} \sum_{\ell=0}^{L_N} \Sigma_\ell^{-1} B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z_m) \end{pmatrix}.$$

In particular, $\mathbb{E}[U_N] = 0_{mp}$ and $\mathbb{E}[U_N U_N'] = V_N$, where V_N is a block matrix whose generic ijth block, $i, j \in \{1, \dots, m\}$, is given by

$$\begin{aligned} V_N(i,j) &= \mathbb{E}\big[U_N(z_i) U_N'(z_j)\big] \\ &= \frac{1}{\sqrt{\mathcal{L}_N(z_i)}} \frac{1}{\sqrt{\mathcal{L}_N(z_j)}} \sum_{\ell=0}^{L_N} \frac{C_{\ell;Z}}{C_\ell} \Sigma_\ell^{-1} \frac{2\ell+1}{16\pi^2} P_\ell(z_i) P_\ell(z_j). \end{aligned}$$

Now, consider $Z \overset{d}{=} \mathcal{N}_{mp}(0_{mp}, I_{mp})$ and $Z_N \overset{d}{=} \mathcal{N}_{mp}(0_{mp}, V_N)$. Applying the triangle inequality twice, it follows that

$$\begin{aligned} d_W(Z, K_N) &\le d_W(Z, U_N) + d_W(U_N, K_N) \\ &\le d_W(Z, Z_N) + d_W(Z_N, U_N) + d_W(U_N, K_N). \end{aligned}$$

From [24], equation (6.4.2), p. 126, we have

$$d_W(Z, Z_N) \le \sqrt{mp} \min\big\{\|V_N^{-1}\|_{\mathrm{op}} \|V_N\|_{\mathrm{op}}^{1/2}, 1\big\} \|V_N - I_{mp}\|_{\mathrm{HS}},$$

where $\|A\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}(A'A)}$, and we observe that

$$\|V_N - I_{mp}\|_{\mathrm{HS}} \le mp \|V_N - I_{mp}\|_\infty = \mathcal{O}\big(N^{-d} \log N\big), \tag{17}$$

from Lemmas 3 and 4 in the Supplementary Material [8]. Indeed, for every $i \in \{1, \dots, m\}$,

$$\begin{aligned} \|V_N(i,i) - I_p\|_{\mathrm{HS}} &\le \frac{\mathrm{const}}{L_N+1} \sum_{\ell=0}^{L_N} \left\| \frac{C_{\ell;Z}}{C_\ell} \Sigma_\ell^{-1} - I_p \right\|_\infty (2\ell+1) \\ &\le \frac{\mathrm{const}}{L_N+1}; \end{aligned}$$

the logarithmic term comes from equation (8) in the Supplementary Lemma 4 [8]. Equation (17) entails that $V_N \to I_{mp}$, thus we have $\|V_N^{-1}\|_{\mathrm{op}} \|V_N\|_{\mathrm{op}}^{1/2} \to 1$, as $N \to \infty$, and

$$d_W(Z, Z_N) = \mathcal{O}\big(N^{-d} \log N\big). \tag{18}$$

Let us recall again from [24], p. 122 (second point of Theorem 6.2.2) that

$$d_W(Z_N, U_N) \le \sqrt{mp} \|V_N^{-1}\|_{\mathrm{op}} \|V_N\|_{\mathrm{op}}^{1/2} m(U_N),$$

where

$$m(U_N) = 2mp \sum_{i=1}^m \sum_{j=1}^p \sqrt{\mathrm{Cum}_4\left[\frac{1}{\sqrt{\mathcal{L}_N(z_i)}} \sum_{\ell=0}^{L_N} \tilde{b}_{\ell;N}(j) \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z_i)\right]},$$

$\tilde{b}_{\ell;N}(j)$ being the jth element of $\Sigma_\ell^{-1} B_{\ell;N}$. Moreover, for the jth element of $\Sigma_\ell^{-1} B_{\ell;N}$ we have

$$\mathrm{Cum}_4\big[\tilde{b}_{\ell;N}(j)\big] = \frac{6}{N(2\ell+1)} \left(\frac{C_{\ell;Z}}{C_\ell} s_\ell(j,j)\right)^2;$$

see equation (4) in Lemma 1. In addition,

$$\begin{aligned}
&\mathrm{Cum}_4\left[\frac{1}{\sqrt{\mathcal{L}_N(z_i)}}\sum_{\ell=0}^{L_N}\tilde{b}_{\ell;N}(j)\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z_i)\right] \\
&\quad = \frac{1}{\mathcal{L}_N^2(z_i)}\sum_{\ell=0}^{L_N}\mathrm{Cum}_4\big[\tilde{b}_{\ell;N}(j)\big]\frac{(2\ell+1)^2}{(4\pi)^4}P_\ell^4(z_i),
\end{aligned}$$

in view of the independence across different multi-poles ℓ. Therefore,

$$\begin{aligned}
&\mathrm{Cum}_4\left[\frac{1}{\sqrt{\mathcal{L}_N(z_i)}}\sum_{\ell=0}^{L_N}\tilde{b}_{\ell;N}(j)\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z_i)\right] \\
&\quad = \frac{6}{N\mathcal{L}_N^2(z_i)}\sum_{\ell=0}^{L_N}\left(\frac{C_{\ell;Z}}{C_\ell}s_\ell(j,j)\right)^2\frac{2\ell+1}{(4\pi)^4}P_\ell^4(z_i) \\
&\quad \le \frac{6}{N\mathcal{L}_N^2(z_i)}\sum_{\ell=0}^{L_N}\left[\frac{C_{\ell;Z}}{C_\ell}\mathrm{Tr}\big(\Sigma_\ell^{-1}\big)\right]^2\frac{2\ell+1}{(4\pi)^4}P_\ell^4(z_i) \\
&\quad \le \frac{\mathrm{const}}{N(L_N+1)^2}\sum_{\ell=0}^{L_N}(2\ell+1)P_\ell^4(z_i).
\end{aligned}$$

Thus, we have

$$m(U_N) \le \mathrm{const}\frac{m^2p^2}{L_N+1}\sqrt{\frac{\log N}{N}}$$

and

$$d_W(Z_N,U_N) = \mathcal{O}\big(N^{-(d+1/2)}(\log N)^{1/2}\big). \tag{19}$$

Now, consider the decomposition

$$\begin{aligned}
\sqrt{\frac{N}{\mathcal{L}_N(z)}}\big(\widehat{k}_N(z)-k(z)\big) &= \frac{1}{\sqrt{\mathcal{L}_N(z)}}\sum_{\ell=0}^{L_N}\sqrt{N(2\ell+1)}(\widehat{\boldsymbol{\phi}}_{\ell;N}-\boldsymbol{\phi}_\ell)\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z) \\
&\quad - \sqrt{\frac{N}{\mathcal{L}_N(z)}}\sum_{\ell=L_N+1}^{\infty}\boldsymbol{\phi}_\ell\frac{2\ell+1}{4\pi}P_\ell(z) \\
&= \frac{1}{\sqrt{\mathcal{L}_N(z)}}\sum_{\ell=0}^{L_N}\Sigma_\ell^{-1}B_{\ell;N}\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z) \\
&\quad + \frac{1}{\sqrt{\mathcal{L}_N(z)}}\sum_{\ell=0}^{L_N}\big[A_{\ell;N}^{-1}-\Sigma_\ell^{-1}\big]B_{\ell;N}\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z) \\
&\quad - \sqrt{\frac{N}{\mathcal{L}_N(z)}}\sum_{\ell=L_N+1}^{\infty}\boldsymbol{\phi}_\ell\frac{2\ell+1}{4\pi}P_\ell(z).
\end{aligned}$$

Without loss of generality, we shall focus on the case $m = 1$; the more general argument is basically identical, with a slightly more cumbersome notation. For $z \in (-1, 1)$,

$$\begin{aligned}
&\left\| \frac{1}{\sqrt{\mathcal{L}_N(z)}} \sum_{\ell=0}^{L_N} [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) \right\| \\
&\quad \le \frac{\text{const}}{\sqrt{L_N+1}} \sum_{\ell=0}^{L_N} \| [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \| \sqrt{2\ell+1} |P_\ell(z)|,
\end{aligned}$$

and then

$$\begin{aligned}
&\mathbb{E}\left[\left\| \frac{1}{\sqrt{\mathcal{L}_N(z)}} \sum_{\ell=0}^{L_N} [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) \right\| \right] \\
&\quad \le \frac{\text{const}}{\sqrt{L_N+1}} \sum_{\ell=0}^{L_N} \mathbb{E} \| [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \| \sqrt{2\ell+1} |P_\ell(z)| \\
&\quad \le \frac{\text{const}}{\sqrt{L_N+1}} \sum_{\ell=0}^{L_N} \frac{1}{\sqrt{N(2\ell+1)}} \sqrt{2\ell+1} |P_\ell(z)| \\
&\quad = \mathcal{O}\left(\frac{1}{\sqrt{N}} \right),
\end{aligned} \tag{20}$$

where for the second inequality we have exploited the Appendix Lemma 2, while for the last step the Hilb's asymptotics (11) in the Appendix (see, also, [31, 32]). Likewise,

$$\begin{aligned}
\left\| \sqrt{\frac{N}{\mathcal{L}_N(z)}} \sum_{\ell=L_N+1}^{\infty} \boldsymbol{\phi}_\ell \frac{2\ell+1}{4\pi} P_\ell(z) \right\| &\le \text{const} \sqrt{\frac{N}{L_N+1}} \sum_{\ell=L_N+1}^{\infty} \|\boldsymbol{\phi}_\ell\| (2\ell+1) |P_\ell(z)| \\
&\le \text{const} \sqrt{\frac{N}{L_N+1}} \sum_{\ell=L_N+1}^{\infty} \|\boldsymbol{\phi}_\ell\| \sqrt{2\ell+1} \\
&= \mathcal{O}\left(\frac{1}{N^{d(\beta_*-1)-1/2}} \right).
\end{aligned} \tag{21}$$

From equations (20) and (21),

$$d_W(U_N, K_N) = \mathcal{O}(N^{-1/2} + N^{1/2+d(1-\beta_*)}). \tag{22}$$

In the end, combining equations (18), (19) and (22), it holds that

$$d_W(Z, K_N) = \mathcal{O}(N^{-1/2} + N^{1/2+d(1-\beta_*)}).$$

Note that the constant in this bound may depend on the choice of m and $z_1, \dots, z_m$. □

We can now give the proof of the third (and final) result.

PROOF OF THEOREM 25. Under Condition 24, we have that, for $z \in [-1, 1]$,

$$\begin{aligned}
\sqrt{N}(\widehat{k}_N(z) - k(z)) &= \sum_{\ell=0}^{L} \sqrt{N(2\ell+1)} (\widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell) \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) \\
&= \sum_{\ell=0}^{L} A_{\ell;N}^{-1} B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z)
\end{aligned} \tag{23}$$

$$= \sum_{\ell=0}^{L} \Sigma_\ell^{-1} B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) + \sum_{\ell=0}^{L} [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z).$$

Then

$$\sup_{z\in[-1,1]} \left\| \sum_{\ell=0}^{L} [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) \right\| \le \sum_{\ell=0}^{L} \| [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \| \frac{\sqrt{2\ell+1}}{4\pi},$$

and hence

$$\mathbb{E}\left[\sup_{z\in[-1,1]} \left\| \sum_{\ell=0}^{L} [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) \right\| \right] \le \sum_{\ell=0}^{L} \mathbb{E} \| [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \| \frac{\sqrt{2\ell+1}}{4\pi} \to 0, \quad N \to \infty,$$

in view of the Supplementary Lemma 2 [8]. Then the second part of the sum in (23) goes to zero in probability. Since the sum (over ℓ) has independent components, we just need to prove that, for each $\ell = 0, 1, 2, \dots, L$, $\{B_{\ell;N} P_\ell(\cdot)\}$ forms a *tight* sequence. Using the tightness criterion given in [4], equation 13.14, p. 143, it is sufficient to show that, for $z_1 \le z \le z_2$,

$$\begin{aligned} &\mathbb{E}\| B_{\ell;N} P_\ell(z) - B_{\ell;N} P_\ell(z_1) \| \| B_{\ell;N} P_\ell(z_2) - B_{\ell;N} P_\ell(z) \| \\ &= |P_\ell(z) - P_\ell(z_1)| |P_\ell(z_2) - P_\ell(z)| \mathbb{E}\| B_{\ell;N} \|^2 \\ &\le p \frac{C_{\ell;Z}}{C_\ell} Q_\ell^2 |z - z_1| |z_2 - z| \\ &\le p \frac{C_{\ell;Z}}{C_\ell} Q_\ell^2 (z_2 - z_1)^2. \end{aligned}$$

Convergence of the finite-dimensional distributions is standard and we omit the details, which are close to those given in the proofs of the previous theorem. Thus the sequence converges weakly to a zero-mean multivariate Gaussian process with covariance function

$$\Gamma_{k_L}(z, z') = \sum_{\ell=0}^{L} C_{\ell;Z} \Gamma_\ell^{-1} \frac{2\ell+1}{16\pi^2} P_\ell(z) P_\ell(z'). \qquad \square$$

6. Some numerical evidence. In this section, we present some short numerical results to illustrate the models and methods that we discussed in this paper.

We stress first that random fields on the sphere cross time can be very conveniently generated by combining the general features of Python with the HEALPix software (see [14] and https://healpix.sourceforge.io). More precisely, HEALPix (which stands for *Hierarchical Equal Area and iso-Latitude Pixelation*) is a multipurpose computer software package for a high resolution numerical analysis of functions on the sphere, based on a clever tessellation scheme: the spherical surface is hierarchically partitioned into curvilinear quadrilaterals of equal area (at a given resolution), distributed on lines of constant latitude, as suggested in the name. In particular, we shall make use of `healpy`, which is a Python package based on the HEALPix C++ library. HEALPix was developed to efficiently process Cosmic Microwave Background data from cosmological experiments (like *Planck*, [27]), but it is now used in many other branches of astrophysics and applied sciences.

In short, HEALPix allows to create spherical maps according to the spectral representation (1), accepting in input either an array of random coefficients $\{a_{\ell,m}\}$, or the angular power spectrum $\{C_\ell\}$, by means of the routines `alm2map` and `synfast`: in the latter case, random $\{a_{\ell,m}\}$ are generated according to a Gaussian zero mean distribution with variance $\{C_\ell\}$. The routine is extremely efficient and allows to generate maps of resolution up to a few thousands multi-poles in a matter of seconds on a standard laptop computer.

In our case, however, we need random fields where the random harmonic coefficients have themselves a temporal dependence structure. For this reason, we implemented a simple routine in Python, to simulate Gaussian $\{a_{\ell,m}(t)\}$ processes, each following an AR(p) de-

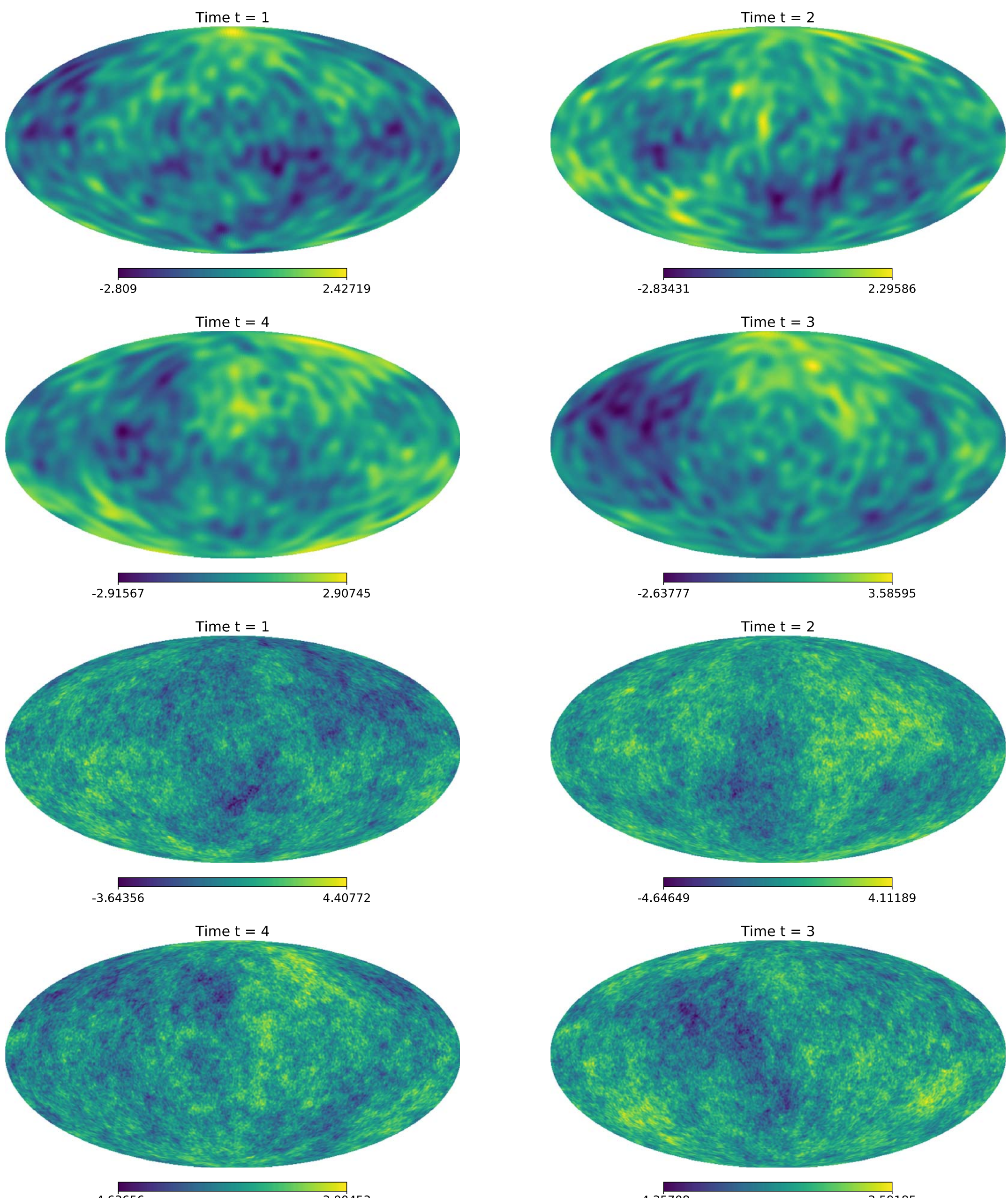

FIG. 1. *Two realizations of* $\{T(x,t)\}$ *at times* $t = 1, 2, 3, 4$ (*clockwise*). *Upper panel*: *maximum resolution* $L_{\max} = 30$. *Lower panel*: *maximum resolution* $L_{\max} = 200$.

pendence structure. These random harmonic coefficients are then uploaded into HEALPix, to generate maps such as those that are given in Figure 1. In particular, in these two cases we fixed $L_{\max} = \max(\ell) = 30, 200$, respectively. Then we generated $\{a_{\ell,m}(t)\}$ according to stationary AR(1) processes, with parameters $\phi_\ell \simeq \text{const} \times \ell^{-3}$; similarly, we took here $C_{\ell;Z} \simeq \text{const} \times \ell^{-2}$. In the figure, we report the realization for the first 4 periods, simply for illustrative purposes.

We are now in the position to use simulations to validate the previous results. In our first Tables 1–3, we report for $B = 1000$ Monte Carlo replications the values of the "variance" and "bias" terms, that is, the first and second summand in the mean squared error equation (16); the second term is actually deterministic, and it is reported to illustrate the approximation one obtains by cutting the expansion to a finite multi-pole value. In the third column, we report, the actual (squared) L^2 error. On the left-hand side, we fix the number of multi-poles to be exploited in the reconstruction of the kernel; on the right-hand side, we consider a sort of "oracle" estimator, where the number of multi-poles grows with the optimal rate $N^{\frac{1}{2\beta_*-1}}$. As before, we took $C_{\ell;Z} \simeq \text{const} \times \ell^{-2}$, $\phi_\ell \simeq \text{const} \times \ell^{-\beta}$ for $\beta = 2, 2.5, 3$; for $N = 100, 300, 700$ the left-hand side uses $L_N \sim N^{0.6}$, while the right-hand side takes $L_N \sim N^{\frac{1}{2\beta_*-1}}$, as explained above.

We note how the estimators perform very efficiently, and show the errors scale approximately as N^α, where $\alpha \approx \frac{2-2\beta_*}{2\beta_*-1}$, as predicted by our computations; see Remark 19. In particular, Figure 2 shows the behaviour of the L^2 error, as a function of N. For $\beta_* = 2, 2.5, 3$, the empirical mean squared error is computed over a grid of N which ranges from 50 to 1000 in

TABLE 1

L^2 errors obtained with $\beta_* = 2$; $L_N \sim N^{0.6}$ (left) and $L_N \sim N^{\frac{1}{2\beta_*-1}}$ (right)

N	Variance	Bias	MSE	N	Variance	Bias	MSE
100	0.00082	0.00006	0.00088	100	0.00041	0.00023	0.00065
300	0.00057	0.00001	0.00059	300	0.00022	0.00010	0.00031
700	0.00041	0.00001	0.00042	700	0.00012	0.00005	0.00018

TABLE 2

L^2 errors obtained with $\beta_* = 2.5$; $L_N \sim N^{0.6}$ (left) and $L_N \sim N^{\frac{1}{2\beta_*-1}}$ (right)

N	Variance	Bias	MSE	N	Variance	Bias	MSE
100	0.00081	0.00007	0.00088	100	0.00063	0.00014	0.00077
300	0.00056	0.00001	0.00057	300	0.00029	0.00006	0.00035
700	0.00041	0.00000	0.00041	700	0.00016	0.00003	0.00019

TABLE 3

L^2 errors obtained with $\beta_* = 3$; $L_N \sim N^{0.6}$ (left) and $L_N \sim N^{\frac{1}{2\beta_*-1}}$ (right)

N	Variance	Bias	MSE	N	Variance	Bias	MSE
100	0.00082	0.00001	0.00084	100	0.00041	0.00021	0.00062
300	0.00058	0.00000	0.00058	300	0.00021	0.00004	0.00025
700	0.00041	0.00000	0.00041	700	0.00009	0.00004	0.00013

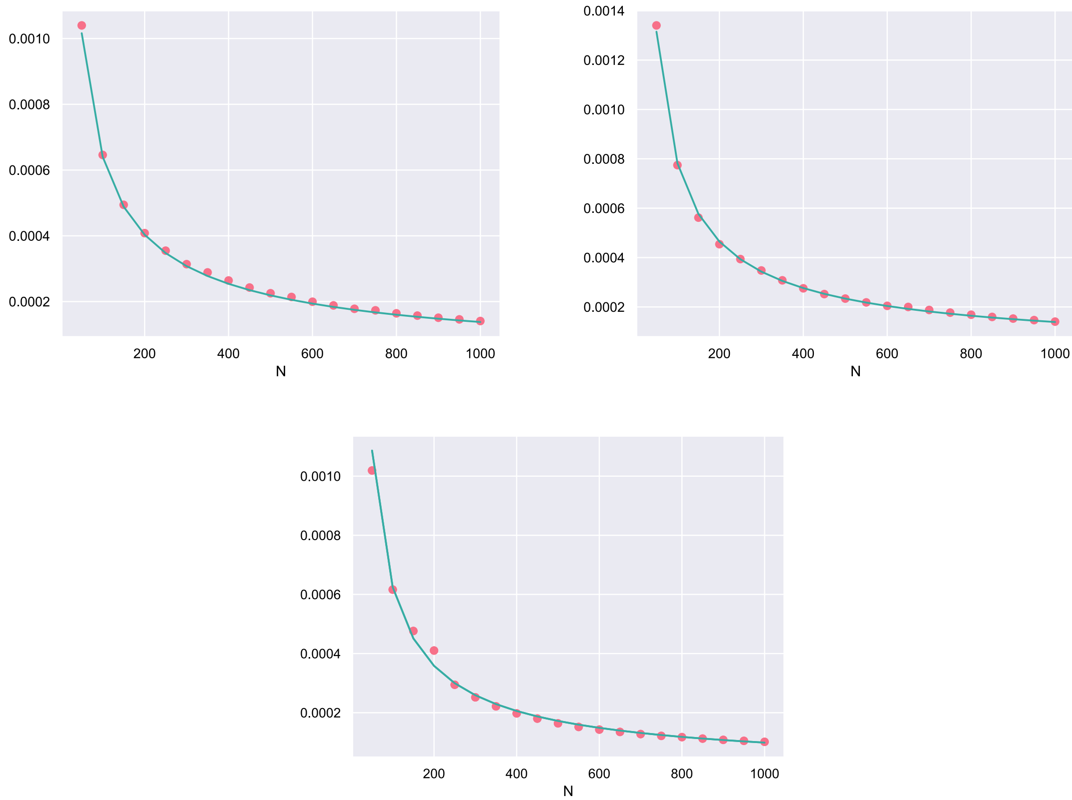

FIG. 2. *L^2 errors (dots) over a grid of N, for $\beta_* = 2, 2.5, 3$ (clockwise) and $L_N \sim N^{\frac{1}{2\beta_*-1}}$. The green lines represent the (calibrated) theoretical upper bounds in equation (24).*

steps of 50. The green lines represent respectively the curves

$$y = \exp(-4.28)x^{-0.667}, \qquad y = \exp(-3.7)x^{-0.75}, \qquad y = \exp(-3.7)x^{-0.80}. \tag{24}$$

As explained earlier, the exponents match our theoretical results, whereas the multiplicative constants have been chosen by a least squares fit.

We can now focus quickly on the main result of our paper, dealing with the quantitative central limit theorem, in Wasserstein distance; the latter is computed following the Python routine (`scipy.stats.wasserstein_distance`). We consider again a model where the autoregressive parameter and the angular power spectra are exactly the same as in the previous settings, in particular, taking $\beta = 3$ and $d = 0.5$, up to integer approximations; we fix $L_{\max} = 1000$ for the number of components under the null hypothesis. Under these circumstances, we evaluate (univariate) Wasserstein distances for the kernel estimators at $m = 9$ different locations, performing $B = 10{,}000$ Monte Carlo replications.

In our simulations, we took a number of time-domain observations ranging from $N = 100$ to $N = 1000$ in steps of 100; it should be noted that huge sample sizes are quite common when dealing with sphere-cross-time data; see, for example, the NCEP/NCAR reanalysis datasets [19] for atmospheric research. In Table 4, we report for brevity a subset of these results, while the full sample is considered in Figure 3.

Again, we note that simulations track closely the theoretical predictions. More precisely, by our theoretical upper bound, we expect the Wasserstein distance $d_W(\cdot,\cdot)$ to decay faster than $N^{-0.5}$ (up to logarithmic factors) in the setting of Table 4, in good agreement with simulations. To help visualize this behaviour, we report in Figure 3 the decay of numerically estimated Wasserstein distances for $K_N(z)$ (see Theorem 21) considered for three different

TABLE 4
Wasserstein distances obtained with $\beta_* = 3$ *and* $L_N \sim N^{0.5}$

$N \setminus z$	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
100	0.52	0.15	0.70	0.19	0.69	0.66	0.42	1.13	0.79
500	0.04	0.10	0.11	0.11	0.10	0.08	0.08	0.13	0.26
1000	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.03	0.07

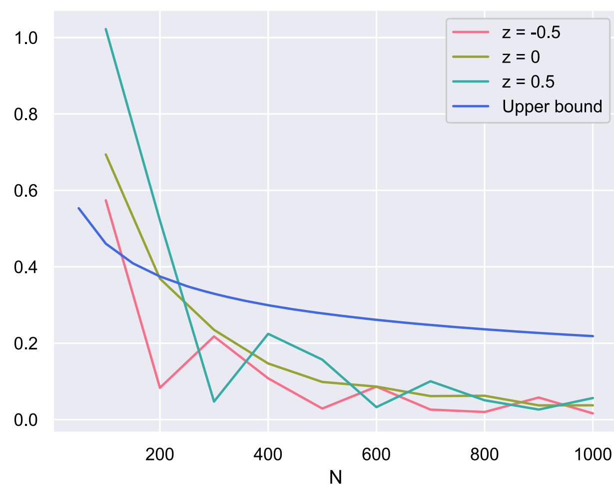

FIG. 3. *Wasserstein distances for* $z = -0.5, 0, 0.5$ *and theoretical upper bound* $\log N \times N^{-0.5}$.

values $z = -0.5, 0, 0.5$, for N in steps of 100 ranging from 100 to 1000; in blue, we reproduce also the expected upper bound, of order $\log N \times N^{-0.5}$. It is evident that the realized values are well controlled by the theoretical bound, with the exception of the smallest samples.

REMARK 27. Although the setting considered in this paper is mainly theoretical, we believe that the models and procedures introduced here have plenty of potential for important applications. A possible dataset, which is in our view amenable to SPHAR modeling, is the NCEP reanalysis catalogue (see [19]); it provides the near-surface air temperature of the planet Earth over a grid of 94×192 unique spatial locations with a time span of 50 years (starting in 1948), sampled every day; overall, then, there are publicly available $18{,}048 \times 18{,}250$ space-cross-time observations. Clearly, for temperature (and, more generally, climate) variables we cannot expect isotropy to hold exactly, due to the presence of features which depend on the location on the surface of the Earth; our idea, however, is that these anisotropic components can be estimated and removed in a preliminary stage of the analysis, just like trend and cyclical components are usually subtracted from time series data before standard ARMA models are implemented (see [6], Section 1.4). These topics are the object of current ongoing research; however, because those investigations require considerable extra work, together with ideas and techniques which are specific to a given application, they will be dealt in a future, more applied paper.

Acknowledgments. DM acknowledges the MIUR Excellence Department Project awarded to the Department of Mathematics, University of Rome Tor Vergata, CUP E83C180 00100006. We are also grateful to Pierpaolo Brutti for many insightful suggestions and conversations.

Some of the results in this paper have been derived using the HEALPix/healpy package.

SUPPLEMENTARY MATERIAL

Supplement to "Asymptotics for spherical functional autoregressions" (DOI: 10.1214/20-AOS1959SUPP; .pdf). This appendix collects a numbers of lemmas which are instrumental for the proofs of the main results in our paper; see [8].

REFERENCES

[1] AUE, A. and VAN DELFT, A. (2020). Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.* **40** 2505–2547. MR4152111 https://doi.org/10.1214/19-AOS1895 To appear. Available at arXiv:1701.01741.

[2] BALDI, P., KERKYACHARIAN, G., MARINUCCI, D. and PICARD, D. (2009). Asymptotics for spherical needlets. *Ann. Statist.* **37** 1150–1171. MR2509070 https://doi.org/10.1214/08-AOS601

[3] BERG, C. and PORCU, E. (2017). From Schoenberg coefficients to Schoenberg functions. *Constr. Approx.* **45** 217–241. MR3619442 https://doi.org/10.1007/s00365-016-9323-9

[4] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, New York. MR1700749 https://doi.org/10.1002/9780470316962

[5] BOSQ, D. (2000). *Linear Processes in Function Spaces*: *Theory and Applications*. *Lecture Notes in Statistics* **149**. Springer, New York. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9

[6] BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1093459 https://doi.org/10.1007/978-1-4419-0320-4

[7] CAPONERA, A. (2020). Statistical inference for spherical functional autoregressions Ph.D. thesis, Sapienza University of Rome.

[8] CAPONERA, A. and MARINUCCI, D. (2021). Supplement to "Asymptotics for spherical functional autoregressions." https://doi.org/10.1214/20-AOS1959SUPP

[9] CHENG, D., CAMMAROTA, V., FANTAYE, Y., MARINUCCI, D. and SCHWARTZMAN, A. (2020). Multiple testing of local maxima for detection of peaks on the (celestial) sphere. *Bernoulli* **26** 31–60. MR4036027 https://doi.org/10.3150/18-BEJ1068

[10] CLARKE DE LA CERDA, J., ALEGRÍA, A. and PORCU, E. (2018). Regularity properties and simulations of Gaussian random fields on the sphere cross time. *Electron. J. Stat.* **12** 399–426. MR3763911 https://doi.org/10.1214/18-EJS1393

[11] FAN, M., PAUL, D., LEE, T. C. M. and MATSUO, T. (2018). A multi-resolution model for non-Gaussian random fields on a sphere with application to ionospheric electrostatic potentials. *Ann. Appl. Stat.* **12** 459–489. MR3773401 https://doi.org/10.1214/17-AOAS1104

[12] FAN, M., PAUL, D., LEE, T. C. M. and MATSUO, T. (2018). Modeling tangential vector fields on a sphere. *J. Amer. Statist. Assoc.* **113** 1625–1636. MR3902234 https://doi.org/10.1080/01621459.2017.1356322

[13] GNEITING, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. MR3102554 https://doi.org/10.3150/12-BEJSP06

[14] GORSKI, K. M., HIVON, E., BANDAY, A. J., WANDELT, B. D., HANSEN, F. K., REINECKE, M. and BARTELMANN, M. (2005). HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *Astrophys. J.* **622** 759–771.

[15] GRADSHTEYN, I. S. and RYZHIK, I. M. (2015). *Table of Integrals*, *Series*, *and Products*, 8th ed. Elsevier/Academic Press, Amsterdam. MR3307944

[16] HÖRMANN, S., KOKOSZKA, P. and NISOL, G. (2018). Testing for periodicity in functional time series. *Ann. Statist.* **46** 2960–2984. MR3851761 https://doi.org/10.1214/17-AOS1645

[17] HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis*, *with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3379106 https://doi.org/10.1002/9781118762547

[18] JUN, M. (2014). Matérn-based nonstationary cross-covariance models for global processes. *J. Multivariate Anal.* **128** 134–146. MR3199833 https://doi.org/10.1016/j.jmva.2014.03.009

[19] KALNAY, E., KANAMITSU, M., KISTLER, R., COLLINS, W., DEAVEN, D., GANDIN, L., ZHU, Y. et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77** 437–472.

[20] LANG, A. and SCHWAB, C. (2015). Isotropic Gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations. *Ann. Appl. Probab.* **25** 3047–3094. MR3404631 https://doi.org/10.1214/14-AAP1067

[21] LEONENKO, N. N., TAQQU, M. S. and TERDIK, G. H. (2018). Estimation of the covariance function of Gaussian isotropic random fields on spheres, related Rosenblatt-type distributions and the cosmic variance problem. *Electron. J. Stat.* **12** 3114–3146. MR3857874 https://doi.org/10.1214/18-EJS1473

[22] MARINUCCI, D. and PECCATI, G. (2011). *Random Fields on the Sphere*: *Representation*, *Limit Theorems and Cosmological Applications*. *London Mathematical Society Lecture Note Series* **389**. Cambridge Univ. Press, Cambridge. MR2840154 https://doi.org/10.1017/CBO9780511751677

[23] NOURDIN, I. and PECCATI, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. MR2520122 https://doi.org/10.1007/s00440-008-0162-x

[24] NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus*: *From Stein's Method to Universality*. *Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. MR2962301 https://doi.org/10.1017/CBO9781139084659

[25] PANARETOS, V. M. and TAVAKOLI, S. (2013). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Process*. *Appl*. **123** 2779–2807. MR3054545 https://doi.org/10.1016/j.spa.2013.03.015

[26] PANARETOS, V. M. and TAVAKOLI, S. (2013). Fourier analysis of stationary time series in function space. *Ann*. *Statist*. **41** 568–603. MR3099114 https://doi.org/10.1214/13-AOS1086

[27] PLANCK COLLABORATION (2016). Planck 2015 results. I. Overview of products and scientific results. *Astron*. *Astrophys*. **594** Art. ID A1.

[28] PORCU, E., ALEGRIA, A. and FURRER, R. (2018). Modelling temporally evolving and spatially globally dependent data. *Int*. *Stat*. *Rev*. **86** 344–377. MR3852415 https://doi.org/10.1111/insr.12266

[29] PORCU, E., BEVILACQUA, M. and GENTON, M. G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J*. *Amer*. *Statist*. *Assoc*. **111** 888–898. MR3538713 https://doi.org/10.1080/01621459.2015.1072541

[30] ROBINSON, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Ann*. *Statist*. **23** 1048–1072. MR1345214 https://doi.org/10.1214/aos/1176324636

[31] SZEGŐ, G. (1975). *Orthogonal Polynomials*, 4th ed. *Colloquium Publications* **XXIII**. Amer. Math. Soc., Providence, RI. MR0372517

[32] WIGMAN, I. (2010). Fluctuations of the nodal length of random spherical harmonics. *Comm*. *Math*. *Phys*. **298** 787–831. MR2670928 https://doi.org/10.1007/s00220-010-1078-8