Journal of Multivariate Analysis 179 (2020) 104640

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Testing normality of data on a multivariate grid

Lajos Horváth [a,1], Piotr Kokoszka [b,*,1], Shixuan Wang [c,1]

[a] Department of Mathematics, University of Utah, Salt Lake City, UT 84112–0090, United States
[b] Department of Statistics, Colorado State University, Fort Collins, CO 80523–1877, United States
[c] Department of Economics, University of Reading, Reading, RG6 6AA, United Kingdom

ARTICLE INFO

Article history:
Received 23 October 2019
Received in revised form 14 May 2020
Accepted 15 May 2020
Available online 28 May 2020

AMS 2010 subject classifications:
primary 62H11
secondary 62F15

Keywords:
Gaussian process
Lattice data
Significance test
Spatial statistics

ABSTRACT

We propose a significance test to determine if data on a regular d-dimensional grid can be assumed to be a realization of Gaussian process. By accounting for the spatial dependence of the observations, we derive statistics analogous to sample skewness and kurtosis. We show that the sum of squares of these two statistics converges to a chi-square distribution with two degrees of freedom. This leads to a readily applicable test. We examine two variants of the test, which are specified by two ways the spatial dependence is estimated. We provide a careful theoretical analysis, which justifies the validity of the test for a broad class of stationary random fields. A simulation study compares several implementations. While some implementations perform slightly better than others, all of them exhibit very good size control and high power, even in relatively small samples. An application to a comprehensive data set of sea surface temperatures further illustrates the usefulness of the test.

1. Introduction

Nearly all modern spatial statistics applications involve Gaussian processes. While for most large sample results it is not necessary to assume Gaussianity, it is often assumed to improve finite-sample inference and effectively apply Bayesian methods. The same goes for nearly all applications involving conditional and simultaneous autoregressive models in discrete space, see the monographs of Cressie [7], Stein [39], Schabenberger and Gotway [34], Cressie and Wikle [8] and Banerjee et al. [5]. A survey of Gaussian modeling in spatial statistics is given by Gelfand and Schliep [14], part III of Gelfand et al. [13] specifically focuses on methods for discrete spatial data which rely on the Gaussian assumption, and then those that do not. Recent research has focused on applying spatial statistics methods based on the assumption of Gaussianity to large data sets and advancing computational approaches, including parallel and distributed computing, see, e.g., Nychka et al. [26], Paciorek et al. [27], Katzfuss [20] and Guhaniyogi and Banerjee [15]. Methodology and theory for spatial Gaussian models continue to be developed, the references are very numerous. We note the recent work of Stroud et al. [41], which is concerned with missing values, and of Chang et al. [6] who study signal identification within the model involving a Gaussian field on a grid.

Despite the prevalence of the assumption of Gaussianity, there appears to exist no significance tests that could be used to assess if it is reasonable to assume that a given spatial data set can be treated as a realization of a Gaussian random field. This is a difficult problem because normality tests, and even exploratory tools like QQ-plots or histograms, require

* Corresponding author.
E-mail address: Piotr.Kokoszka@colostate.edu (P. Kokoszka).
[1] All authors contributed equally to this work.

https://doi.org/10.1016/j.jmva.2020.104640
0047-259X/

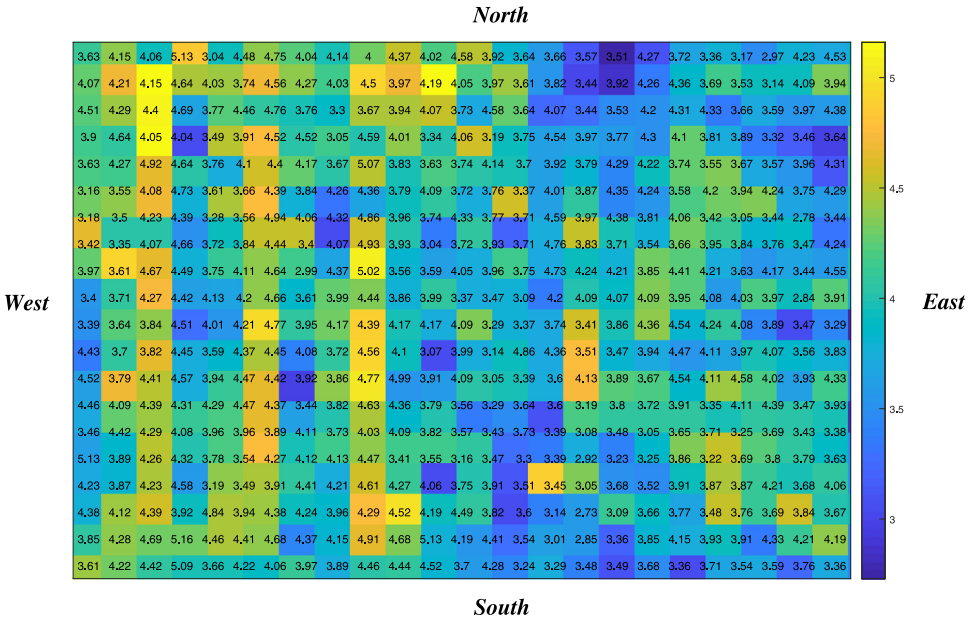

Fig. 1. Mercer and Hall wheat-yield data. The data were collected from wheat uniformity trials carried out at Rothamsted Experimental Station in 1910. The yield data are on 20 × 25 grid with each slot in a size of approximately 3.30 meter (east–west) × 2.5 meter (north–south), giving the total area of roughly one acre.

a random sample (iid observations) from a distribution whose Gaussianity is to be determined. For a general spatial data set, testing the joint normality of all finite-dimensional distributions is practically impossible. We will show that it is possible for data defined on a grid under the assumption of stationarity. When the original data do not appear stationary, it is a common practice to attempt to transform them to stationarity. For example, one can use the deformation approach pioneered by Sampson and Guttorp [33] and subsequently developed by Schmidt and O'Hogan [35], Anderes and Stein [1] and Fouedjio et al. [11], among others. A more common approach is to consider regression models, e.g., Chapter 6 of Schabenberger and Gotway [34], whose errors are stationary, and are often assumed jointly normal. These procedures should also be validated by suitable normality tests.

We illustrate an application of our methodology by considering a classical data set of wheat yields studied in some detail in Section 4.5 of Cressie [7], and many earlier papers cited there. The data are shown in Fig. 1. It is argued in Cressie [7] that no transformation of these data is needed to ensure stationarity. The question we want to answer is if these data can be considered to be a realization of a Gaussian process, i.e., if these values can be assumed to be a realization of a random field $X_{i,j}$, $i, j \in \mathbb{Z}$, whose all finite-dimensional distributions are multivariate normal. This question is difficult to answer because the pronounced spatial dependence of these data can "force" more large or small values in a finite region than univariate normality might suggest. More fundamentally, since these data are not a random sample, usual exploratory plots or tests cannot be relied on. Our significance test shows that these data can be assumed to be a realization of a Gaussian process. Depending on the implementation, the P-values are between 16% and 52%, details are shown in Section I of the online supplement. Our simulations show that most implementations have sufficient power to detect a departure from normality that matters, even for the relatively small sample size (20 × 25 grid) of the data in Fig. 1.

We hope that the test we propose will turn out to be a useful diagnostic tool, which may lend confidence in the application of various methodologies based on the normality assumption, or provide a caution on the validity of conclusions. An appealing feature of our test is that the test statistics can be computed fairly easily using existing R or MATLAB software, and the critical values are those of a chi-square distribution. The test has good empirical size and power, and can be justified asymptotically using recent advances in the asymptotic theory for random fields and new arguments related to the quantification of spatial dependence.

The assumption of normality has underlain much of the development of statistics, well beyond spatial statistics, and many tests have been proposed. Perhaps the best known is the Shapiro and Wilk [37] test, which has been extended and improved in many directions, Royston [30,31,32]. Tests based on the empirical distribution function have also been extensively used, Anderson and Darling [2], Stephens [40], Scholz and Stephens [36]. Great many other approaches have been proposed, Mardia [22,23] D'Agostino et al. [9], Henze and Zirkler [16], Doornik and Hansen [10], among many others. However, perhaps the most commonly used test is the Jarque and Bera [18,19] test. It checks if the first four moments of a distribution agree with the those of a normal distribution. This is a general direction we take. Our test will not detect very subtle departures from normality, which manifest themselves in discrepancy in moments beyond the first four, but it will detect most commonly encountered deviations from normality.

The paper is organized as follows. In Section 2 we develop the test. Its finite sample performance is evaluated in Section 3 by means of a simulation study and an application to a climate data set. There are many possible implementations of our general paradigm, which must be evaluated and compared. The proofs of the mathematical results of Section 2, needed to derive and justify the test, are presented in Section II of an online supplement, which also contains additional details of the test procedure and additional tables, which support our conclusions and recommendations.

2. Testing procedure and its large sample justification

We derive and formulate the testing procedure in Section 2.1, where we also specify the most important assumptions for its validity. A fundamental ingredient of our approach is the quantification and estimation of spatial dependence, this is treated in Section 2.2. Asymptotic theory underlying both Sections 2.1 and 2.2 is developed in Section 2.3.

2.1. Assumptions and test derivation

Let $\mathbb{Z}^d$ denote the set of d–dimensional vectors with integer coordinates. We assume that the observations X_i follow the model

$$X_i = \mu + e_i, \quad i \in \mathbb{Z}^d,$$

where $\{e_i\}$ is a strictly stationary, zero mean spatial process. The mean μ is unknown.

We want to test

$$H_0: \quad \text{the } X_i \text{ are jointly normal,}$$

against the alternative that H_0 does not hold. The test is based on observations X_i, $i \in \Gamma_n \subset \mathbb{Z}^d$. The domain Γ_n is indexed by positive integers n, which are not sample sizes, but sample indexes in increasing domain asymptotics. The sample size is denoted by n_Γ, the cardinality of the set Γ_n, $n_\Gamma = |\Gamma_n|$. If $d = 2$, and $\Gamma_n = \Gamma_{N,M} := \{(i, j), 1 \le i \le N, 1 \le j \le M\}$, then $n_\Gamma = NM$. Let $\partial\Gamma_n$ denote the boundary of Γ_n and $|\partial\Gamma_n|$ its cardinality. We assume that, as $n \to \infty$,

$$\frac{|\partial\Gamma_n|}{n_\Gamma} \to 0. \tag{1}$$

Condition (1) states that asymptotically there should be many more points in the interior of the domain than at its boundary. If $d = 2$, and $\Gamma_n = \Gamma_{N,M}$, defined above, then (1) holds if and only of $\min(N, M) \to \infty$.

We assume that under the null hypothesis $\{e_i\}$ is a Gaussian spatial linear process, i.e., it satisfies the following assumption.

Assumption 1. The e_i are spatial moving averages,

$$e_i = \sum_{s\in\mathbb{Z}^d} a_s \varepsilon_{i-s}, \quad i \in \mathbb{Z}^d, \tag{2}$$

with independent, standard normal innovations ε_i, and the coefficients a_s satisfying

$$\sum_{s\in\mathbb{Z}^d} |a_s| < \infty. \tag{3}$$

Assumption 1 implies that the field $\{X_i\}$ is strictly stationary and Gaussian, with spatial dependence quantified by conditions (2) and (3). Linearity in (2) is needed to ensure normality of the observations. The summability condition in (3) cannot be relaxed because the required CLT would not hold with standard rate, see Lahiri and Robinson [21]. Under Assumption 1, the random variables

$$z_i = \frac{X_i - \mu}{\sigma}, \quad \text{with} \quad \sigma^2 = \sum_{s\in\mathbb{Z}^d} a_s^2, \tag{4}$$

are standard normal (but, in general, not independent). The z_i must be approximated by random variables that can be computed from the sample. For this purpose, define

$$S_n^2 = \frac{1}{n_\Gamma} \sum_{i\in\Gamma_n} (X_i - \bar{X}_n)^2, \quad \bar{X}_n = \frac{1}{n_\Gamma} \sum_{i\in\Gamma_n} X_i.$$

Our tests statistics are based on the standardized observations

$$x_i = x_{i,n} = \frac{X_i - \bar{X}_n}{S_n}, \quad i \in \Gamma_n, \tag{5}$$

which are sample counterparts of the standard normal z_i defined above. Using the x_i, we define the sample skewness and kurtosis by

$$\mathcal{S}_n = \frac{1}{n_\Gamma^{1/2}} \sum_{i\in\Gamma_n} x_i^3 \quad \text{and} \quad \mathcal{K}_n = \frac{1}{n_\Gamma^{1/2}} \sum_{i\in\Gamma_n} (x_i^4 - 3). \tag{6}$$

As we will see in Section 2.3, the asymptotic variances of $\mathcal{S}_n$ and $\mathcal{K}_n$ are, respectively,

$$\phi_{\mathcal{S}}^2 = \sum_{i\in\mathbb{Z}^d} E\left[(z_0^3 - 3z_0)(z_i^3 - 3z_i)\right] \tag{7}$$

and

$$\phi_{\mathcal{K}}^2 = \sum_{i\in\mathbb{Z}^d} E\left[(z_0^4 - 6z_0^2 + 3)(z_i^4 - 6z_i^2 + 3)\right]. \tag{8}$$

In particular,

$$\phi_{\mathcal{K}}^2 \neq \sum_{i\in\mathbb{Z}^d} E\left[(z_0^4 - 3)(z_i^4 - 3)\right].$$

This motivates the introduction of modified sample skewness and kurtosis defined by

$$\mathcal{S}_n^\star = \frac{1}{n_\Gamma^{1/2}} \sum_{i\in\Gamma_n} (x_i^3 - 3x_i) \quad \text{and} \quad \mathcal{K}_n^\star = \frac{1}{n_\Gamma^{1/2}} \sum_{i\in\Gamma_n} (x_i^4 - 6x_i^2 + 3).$$

Observe that $\mathcal{S}_n^\star = \mathcal{S}_n$ because $\sum_{i\in\Gamma_n} x_i = 0$. The statistics $\mathcal{S}_n^\star$ and $\mathcal{K}_n^\star$ also have asymptotic variances, respectively, $\phi_{\mathcal{S}}^2$ and $\phi_{\mathcal{K}}^2$, and are better matched to them in finite samples because $\phi_{\mathcal{S}}^2$ and $\phi_{\mathcal{K}}^2$ are direct counterparts of spatial long-run variances of the sequences $\left\{x_i^3 - 3x_i\right\}$ and $\left\{x_i^4 - 6x_i^2 + 3\right\}$.

Denoting by $\hat{\phi}_{\mathcal{S}}$ and $\hat{\phi}_{\mathcal{K}}$ consistent estimators of $\phi_{\mathcal{S}}$ and $\phi_{\mathcal{K}}$, the test statistic is defined as

$$J_n^\star = \frac{\mathcal{S}_n^{\star 2}}{\hat{\phi}_{\mathcal{S}}^2} + \frac{\mathcal{K}_n^{\star 2}}{\hat{\phi}_{\mathcal{K}}^2}.$$

It is the sum of squares of normalized skewness and kurtosis. As will be stated in Section 2.3, $J_n^\star$ is asymptotically chi-square with two degrees of freedom. The test thus is:

Reject H_0 at significance level α if $J_n^\star > \chi_2^2(1-\alpha)$, where $\chi_2^2(1-\alpha)$ is the $(1-\alpha)$th quantile of the chi-square distribution with two degrees of freedom.

Suitable estimators $\hat{\phi}_{\mathcal{S}}^2$ and $\hat{\phi}_{\mathcal{K}}^2$ are derived in Section 2.2, see formulas (12) and (13).

The key to understanding the need for the modified kurtosis is the fact that

$$\phi_{\mathcal{K}}^2 \neq \sum_{i\in\mathbb{Z}^d} E\left[(z_0^4 - 3)(z_i^4 - 3)\right].$$

The formula given above must be used instead, which is the long-run variance of the unobservable field $\left\{z_i^4 - 6z_i^2 + 3\right\}$. We replace the z_i by the observable x_i, which approximate them with an asymptotically negligible effect. In particular, $\mathrm{Var}[\mathcal{K}_n^\star] = \phi_{\mathcal{K}}^2$, so $\mathcal{K}_n^{\star 2}$ divided by an estimator of the variance of $\mathcal{K}_n^\star$ is a Wald statistic, which is asymptotically χ_1^2. (The population kurtosis is zero under the null hypothesis.) The same argument applies the skewness. We show that these two components are asymptotically independent, so their sum is asymptotically χ_2^2.

2.2. Estimation of the spatial long run variances

It is useful to consider a more general setting. Suppose $\left\{y_i, i \in \mathbb{Z}^d\right\}$ is a zero mean strictly stationary scalar random field such that $Ey_0^2 < \infty$, whose covariances are $\gamma(j) = E[y_0 y_j]$, $j \in \mathbb{Z}^d$. The objective is to estimate the long-run, or asymptotic, variance defined by

$$\sigma^2 = \sum_{j\in\mathbb{Z}^d} \gamma(j) = \sum_{j\in\mathbb{Z}^d} E[y_0 y_j]. \tag{9}$$

We assume throughout that

$$\sum_{j\in\mathbb{Z}^d} |\gamma(j)| < \infty, \tag{10}$$

so that σ^2 can be defined. We observe $y_j \in \Gamma_n$, which is a rectangle whose all dimensions are increasing, as specified in the following assumption.

Assumption 2. The spatial domain Γ_n is given by

$$\Gamma_n = \{1, \dots, n_1\} \times \{1, \dots, n_2\} \times \cdots \times \{1, \dots, n_d\}$$

and $n^\star := \min_{1\le i\le d} n_i \to \infty$.

The sample covariances are defined by

$$\hat\gamma(j) = |\Gamma_n(j)|^{-1} \sum_{i\in\Gamma_n(j)} y_i y_{i+j}, \quad \text{where} \quad \Gamma_n(j) = \{i \in \Gamma_n : i + j \in \Gamma_n\}.$$

To provide explicit formulas, in the following we use the notation $j = (j_1, \dots, j_d)$. In this setting, σ^2 is estimated by the kernel estimator

$$\hat\sigma_n^2 = \sum_{\ell=1}^{d} \sum_{|j_\ell|\le n_\ell} \left\{ \prod_{\ell=1}^{d} K\left(\frac{j_\ell}{h_\ell}\right) \right\} \hat\gamma(j_1, \dots, j_d), \tag{11}$$

where K is a univariate kernel satisfying the following commonly used assumption.

Assumption 3. The kernel K is a continuous function on the interval $[-1, 1]$ satisfying $K(0) = 1$. The bandwidths h_ℓ satisfy $h^\star := \max_{1\le\ell\le d} h_\ell \to \infty$, as $n \to \infty$.

In our context, we use estimator (11) computed from $y_i = x_i^3 - 3x_i$ and $y_i = x_i^4 - 6x_i + 3$. These y_i do not form a strictly stationary random field. Due to the random normalization in (5), they form a structure which could be called a spatial triangular array. However, the z_i defined by (4) do form a strictly stationary random field, so it must be shown that replacing the x_i by the z_i introduces an asymptotically negligible effect into the estimation of $\phi_{\mathcal{S}}^2$ and $\phi_{\mathcal{K}}^2$. This will be established in the proof of Theorem 2. We first introduce the required notation. Set

$$y_i^{\mathcal{S}} = x_i^3 - 3x_i, \quad y_i^{\mathcal{K}} = x_i^4 - 6x_i + 3$$

and

$$\bar y_{\mathcal{S}} = \frac{1}{n_\Gamma} \sum_{i\in\Gamma_n} y_i^{\mathcal{S}}, \quad \bar y_{\mathcal{K}} = \frac{1}{n_\Gamma} \sum_{i\in\Gamma_n} y_i^{\mathcal{K}}.$$

Next, we define the sample covariances

$$\hat\gamma_{\mathcal{S}}(j) = |\Gamma_n(j)|^{-1} \sum_{i\in\Gamma_n(j)} \left(y_i^{\mathcal{S}} - \bar y_{\mathcal{S}}\right)\left(y_{i+j}^{\mathcal{S}} - \bar y_{\mathcal{S}}\right),$$
$$\hat\gamma_{\mathcal{K}}(j) = |\Gamma_n(j)|^{-1} \sum_{i\in\Gamma_n(j)} \left(y_i^{\mathcal{K}} - \bar y_{\mathcal{K}}\right)\left(y_{i+j}^{\mathcal{K}} - \bar y_{\mathcal{K}}\right).$$

Using notation

$$\sum_{j\in J(h)} w_h(j) g(j) = \sum_{\ell=1}^{d} \sum_{|j_\ell|\le n_\ell} \left\{ \prod_{\ell=1}^{d} K\left(\frac{j_\ell}{h_\ell}\right) \right\} g(j_1, \dots, j_d),$$

which applies to any function g on $\mathbb{Z}^d$, we define the *kernel* estimators

$$\hat\phi_{\mathcal{S},\text{kern}}^2 = \sum_{j\in J(h)} w_h(j)\hat\gamma_{\mathcal{S}}(j), \quad \hat\phi_{\mathcal{K},\text{kern}}^2 = \sum_{j\in J(h)} w_h(j)\hat\gamma_{\mathcal{K}}(j). \tag{12}$$

The idea behind the kernel estimators is as follows. Focus on $\hat\phi_{\mathcal{K},\text{kern}}^2$ and consult formula (8). We replace the model autocovariances $E\left[(z_0^4 - 6z_0^2 + 3)(z_j^4 - 6z_j^2 + 3)\right]$ by the sample autocovariances $\hat\gamma_{\mathcal{K}}(j)$. The latter are variable if the set $\Gamma_n(j)$ is small, i.e., if j is "spatially large". For this reason, we put smaller weights on them. This idea has been commonly used in time series analysis.

Another class of estimators can be derived as follows. Set $\rho_i = E[z_0 z_i]$. Tedious calculations, using the values of the moments of the standard normal distributions, show that

$$\phi_{\mathcal{S}}^2 = 6\sum_{i\in\mathbb{Z}^d} \rho_i^3 \quad \text{and} \quad \phi_{\mathcal{K}}^2 = 24\sum_{i\in\mathbb{Z}^d} \rho_i^4.$$

We estimate the ρ_i by the sample covariances of the x_i, i.e., by (recall that $\bar x = 0$)

$$\hat\gamma_x(j) = |\Gamma_n(j)|^{-1} \sum_{i\in\Gamma_n(j)} x_i x_{i+j}$$

and define the *power* estimators

$$\hat{\phi}^2_{\mathcal{S},\text{pow}} = 6\sum_{j\in J(h)} w_h(j)\hat{\gamma}^3_x(j), \quad \hat{\phi}^2_{\mathcal{K},\text{pow}} = 24\sum_{j\in J(h)} w_h(j)\hat{\gamma}^4_x(j), \tag{13}$$

i.e.,

$$\hat{\phi}^2_{\mathcal{S},\text{pow}} = 6\sum_{\ell=1}^{d}\sum_{|j_\ell|\le h_\ell}\left\{\prod_{\ell=1}^{d} K\left(\frac{j_\ell}{n_\ell}\right)\right\}\hat{\gamma}^3_x(j_1,\ldots,j_d),$$

$$\hat{\phi}^2_{\mathcal{K},\text{pow}} = 24\sum_{\ell=1}^{d}\sum_{|j_\ell|\le h_\ell}\left\{\prod_{\ell=1}^{d} K\left(\frac{j_\ell}{n_\ell}\right)\right\}\hat{\gamma}^4_x(j_1,\ldots,j_d).$$

The consistency of the above spatial long-run variance estimators is established in Section 2.3. More explicit formulas for the commonly encountered case of a 2D rectangular domain are given in Section III of the Supplement.

2.3. Asymptotic theory

This section contains asymptotic results, which justify the application of the test for a large class of stationary fields. All proofs are given in Section II of the supplement. The first result establishes the asymptotic distribution of the sample skewness $\mathcal{S}_n$ and kurtosis $\mathcal{K}_n$, and their modified versions $\mathcal{S}^\star_n$ and $\mathcal{K}^\star_n$. Very little must be assumed about the shape of the spatial domain Γ_n.

Theorem 1. *Suppose condition (1) and Assumption 1 hold. Then the series (7) and (8) defining, respectively, $\phi^2_{\mathcal{S}}$ and $\phi^2_{\mathcal{K}}$ are absolutely convergent, and the vectors $[\mathcal{S}_n, \mathcal{K}_n]^\top$ and $[\mathcal{S}^\star_n, \mathcal{K}^\star_n]^\top$ both converge to the bivariate normal distribution with mean zero and covariance matrix*

$$\begin{bmatrix} \phi^2_{\mathcal{S}} & 0 \\ 0 & \phi^2_{\mathcal{K}} \end{bmatrix}.$$

Based on Theorem 1, we consider the test statistics

$$\widehat{J}_n = \frac{\mathcal{S}^2_n}{\hat{\phi}^2_{\mathcal{S}}} + \frac{\mathcal{K}^2_n}{\hat{\phi}^2_{\mathcal{K}}} \quad \text{and} \quad J^\star_n = \frac{\mathcal{S}^{\star 2}_n}{\hat{\phi}^2_{\mathcal{S}}} + \frac{\mathcal{K}^{\star 2}_n}{\hat{\phi}^2_{\mathcal{K}}}.$$

The following corollary is an immediate consequence of Theorem 1.

Corollary 1. *Suppose condition (1) and Assumption 1 hold, and*

$$\hat{\phi}^2_{\mathcal{S}} \xrightarrow{P} \phi^2_{\mathcal{S}} \quad \text{and} \quad \hat{\phi}^2_{\mathcal{K}} \xrightarrow{P} \phi^2_{\mathcal{K}}. \tag{14}$$

Then $\widehat{J}_n \xrightarrow{D} \chi^2_2$ and $J^\star_n \xrightarrow{D} \chi^2_2$, where χ^2_2 is a chi-square random variable with two degrees of freedom.

We now turn to the consistency of the estimators given by (12) and (13). For these results more restrictive assumptions on the spatial domain are required. Recall that $n^\star := \min_{1\le i\le d} n_i$ and $h^\star = \max_{1\le \ell\le d} h_\ell$.

Theorem 2. *Suppose (1), Assumptions 1–3 hold, and $h^\star = o(n^{\star 1/2})$. Then relations (14) hold for the estimators $\hat{\phi}^2_{\mathcal{S},\text{kern}}$ and $\hat{\phi}^2_{\mathcal{K},\text{kern}}$ given by (12) and the estimators $\hat{\phi}^2_{\mathcal{S},\text{pow}}$ and $\hat{\phi}^2_{\mathcal{K},\text{pow}}$ given by (13).*

Estimation of the spatial long-run variance σ^2 given by (9) has been recently studied by Prause and Steland [29] who established consistency assuming φ-mixing with a suitable rate. If the errors ε_j are normal, even for $d = 1$, the moving average (2) is φ-mixing if only finitely many coefficients a_s are not zero, see Ibragimov and Linnik [17] and Sidorov [38]. For this reason, we use a different, more direct, approach to prove Theorem 2.

We now turn to the consistency of the test. We begin with an assumption which is essentially Assumption 1, but without assuming normality.

Assumption 4. The e_i are moving averages (2) with independent and identically distributed random variables ε_i, satisfying $E\varepsilon_\ell = 0$, $E\varepsilon^2_\ell = 1$, $E\varepsilon^8_\ell < \infty$, and the coefficients a_s satisfying (3).

Under Assumption 4, we can establish limits in probability of $n_\Gamma^{-1/2}\mathcal{S}^\star_n$ and $n_\Gamma^{-1/2}\mathcal{K}^\star_n$, as stated in Theorem 3. Notice that under H_0 these limits are zero.

Theorem 3. *If (1) and Assumption 4 hold, then*

$$n_\Gamma^{-1/2}\mathcal{S}^\star_n \xrightarrow{P} Ez_0^3 \quad \text{and} \quad n_\Gamma^{-1/2}\mathcal{K}^\star_n \xrightarrow{P} Ez_0^4 - 3,$$

where z_0 is defined by (4). The limit of $n_\Gamma^{-1/2}\mathcal{K}_n$ is the same as the limit of $n_\Gamma^{-1/2}\mathcal{K}^\star_n$.

Next we establish bounds on magnitudes of the estimators of the long-run variances.

Theorem 4. *Suppose (1) and Assumptions 3 and 4 hold, and* $h^\star = o(n^{\star 1/2})$. *Then*

$$\hat{\phi}^2_{\mathcal{S},\text{kern}} = O_P(h^\star), \quad \hat{\phi}^2_{\mathcal{K},\text{kern}} = O_P(h^\star)$$

and

$$\hat{\phi}^2_{\mathcal{S},\text{pow}} = O_P(1), \quad \hat{\phi}^2_{\mathcal{K},\text{pow}} = O_P(1).$$

Using Theorems 3 and 4, we can prove the consistency of the test.

Corollary 2. *If the conditions of Theorem 4 are satisfied and if* $Ez_0^3 \neq 0$ *and/or* $Ez_0^4 \neq 3$, *then* $\widehat{J}_n \xrightarrow{P} \infty$ *and* $J_n^\star \xrightarrow{P} \infty$.

3. Finite sample performance and application to temperature data

In Section 3.1, we explore the empirical size and power of several implementations of our test. In Section 3.2, we check if the spatial fields of sea surface temperature anomalies can be assumed to be Gaussian, and provide further insights into the behavior of the test.

3.1. A simulation study

In this section, we use Monte Carlo simulation to assess finite sample properties of the test derived in Section 2.1. We focus on the case of $d = 2$, most commonly encountered in applications. Explicit formulas in this case are given in Section III of the Supplement. We consider data generating processes (DGPs) defined by three different spatial models specified below, and by several grid sizes. We use 5000 independent replications, and record the count of rejections to calculate empirical size and power of the proposed test.

We generate realizations on a grid $\{1 \leq i, j \leq N\}$ of the following spatial models:

Spatial IID: $X_{i,j} = 2 + \sqrt{2}\xi_{i,j}$.
Spatial Moving-average (MA): $X_{i,j} = \xi_{i,j} + 0.5\xi_{i,j-1}$.
Spatial Autoregressive(AR): $X_{i,j} = 0.5X_{i-1,j-1} + \xi_{i,j}$.

Under H_0, $\xi_{i,j} \sim$ *i.i.d.* $\mathcal{N}(0, 1)$. We consider two error distributions under H_A: the $\xi_{i,j}$ are i.i.d. with either Student's ι-distribution with ν degrees of freedom or with the skew-normal distribution. We set ν to values ranging from 5 to 20. If $\nu \geq 30$, the univariate ι-distribution is visually almost indistinguishable from the standard normal distribution, and its quantiles are almost equal to the standard normal quantiles. Unlike the ι-distribution, the skew-normal distribution, treated in Azzalini [4], has nonzero skewness. Further details and power tables are presented in Section IV of the Supplement.

Both the *kernel* and *power* estimators, defined in Section 2.2 (and Section III of the Supplement), need the specification of the kernel and the smoothing bandwidth. Three kernel functions are compared.

The truncated kernel (TR): $K_{TR}(t) = I\{|t| \leqslant 1\}$.
The Bartlett kernel (BT): $K_{BT}(t) = (1 - |t|)\, I\{|t| \leqslant 1\}$.
The flat-top kernel (FT):

$$K_{FT}(t) = \begin{cases} 1, & 0 \leqslant t < 0.5 \\ 2 - |t|, & 0.5 \leqslant t < 1 \\ 0, & 1 \leqslant t. \end{cases}$$

The bandwidth h for these kernels is selected as

$$h_{TR} = \lfloor 4(N/100)^{1/5} \rfloor, \quad h_{BT} = \lfloor 4(N/100)^{2/9} \rfloor, \quad h_{FT} = \lfloor 4(N/100)^{1/5} \rfloor. \tag{15}$$

The choice of the smoothing bandwidth has been well studied. For the truncated and Bartlett kernels, Newey and West [25] compared the performance of different plug-in methods, while Andrews [3] proposed a data-driven bandwidth selection technique. Politis [28] developed an adaptive bandwidth choice for the flat-top kernel. It turns out that these choices work well for our purpose. We thus follow Newey and West [25] to select the bandwidth for the truncated and Bartlett kernels. Our simulations showed that choosing the bandwidth of the flat-top kernel the same as for the truncated kernel produces stable and satisfactory results.

Empirical size Table 1 reports the empirical sizes, the percentages of rejections under H_0. As can be seen, the empirical sizes are close to the theoretical levels, even for small grid size, such as $N = 100$. Comparing the results for the *kernel* estimator and the *power* estimator, it seems that there is no obvious pattern in the empirical sizes. The differences arising

Table 1
The empirical sizes of 5000 independent simulations with significant levels of 10%, 5%, and 1% for the spatial normality tests based on the *kernel* estimators $\hat{\phi}^2_{\mathcal{S},\text{kern}}$, $\hat{\phi}^2_{\mathcal{K},\text{kern}}$ and the *power* estimators $\hat{\phi}^2_{\mathcal{S},\text{pow}}$, $\hat{\phi}^2_{\mathcal{K},\text{pow}}$ for three DGPs of spatial IID ($X_{i,j} = 2+\sqrt{2}\xi_{i,j}$), spatial moving-average ($X_{i,j} = \xi_{i,j}+0.5\xi_{i,j-1}$), spatial autoregressive($X_{i,j} = 0.5X_{i-1,j-1}+\xi_{i,j}$), where $\xi_{i,j} \sim$ *i.i.d.* $\mathcal{N}(0,1)$.

Grid size	Kernel	*Kernel* estimator			*Power* estimator		
		10%	5%	1%	10%	5%	1%
Panel A: Spatial IID							
N = 100	Truncated	10.12%	4.92%	1.10%	10.52%	4.96%	1.16%
	Bartlett	9.52%	4.54%	1.08%	10.52%	4.96%	1.16%
	Flat-top	9.54%	4.84%	1.10%	10.52%	4.96%	1.16%
N = 500	Truncated	9.66%	5.08%	0.84%	10.48%	5.46%	1.08%
	Bartlett	9.66%	5.06%	0.84%	10.48%	5.46%	1.08%
	Flat-top	9.66%	5.08%	0.86%	10.48%	5.46%	1.08%
N = 1000	Truncated	9.68%	4.90%	0.96%	10.26%	5.06%	0.98%
	Bartlett	9.64%	4.86%	0.96%	10.26%	5.06%	0.98%
	Flat-top	9.70%	4.88%	0.98%	10.26%	5.06%	0.98%
Panel B: Spatial moving average							
N = 100	Truncated	10.72%	5.44%	1.30%	10.00%	4.68%	0.78%
	Bartlett	10.68%	5.70%	1.30%	10.42%	5.04%	0.86%
	Flat-top	10.36%	5.44%	1.26%	10.00%	4.68%	0.78%
N = 500	Truncated	10.16%	4.82%	1.12%	9.96%	4.76%	1.02%
	Bartlett	10.58%	5.10%	1.24%	10.46%	5.06%	1.12%
	Flat-top	10.14%	4.72%	1.12%	9.96%	4.76%	1.02%
N = 1000	Truncated	10.44%	5.38%	1.18%	10.00%	4.94%	1.02%
	Bartlett	10.84%	5.64%	1.24%	10.18%	5.12%	1.20%
	Flat-top	10.54%	5.42%	1.16%	10.00%	4.94%	1.02%
Panel C: Spatial autoregressive							
N = 100	Truncated	10.70%	5.82%	1.50%	9.34%	4.74%	0.96%
	Bartlett	12.32%	6.60%	1.64%	11.56%	5.74%	1.40%
	Flat-top	10.46%	5.42%	1.34%	9.36%	4.74%	0.96%
N = 500	Truncated	10.12%	5.06%	0.96%	9.94%	4.82%	1.00%
	Bartlett	11.58%	6.12%	1.14%	11.74%	5.82%	1.28%
	Flat-top	9.96%	5.02%	0.90%	9.98%	4.84%	1.02%
N = 1000	Truncated	10.00%	4.74%	1.00%	9.70%	4.66%	1.12%
	Bartlett	11.50%	5.62%	1.20%	10.90%	5.64%	1.30%
	Flat-top	10.00%	4.76%	1.00%	9.70%	4.66%	1.12%

from the application of different kernels are small and do not exhibit any clear pattern either. We conclude that our test controls size very well, not matter which one of the six considered implementations is used.

Empirical power Tables 2 and 3 present the empirical power of the test by 5% significance level critical values with the spatial long run variance estimated, respectively, by the *kernel* estimator and the *power* estimator. As expected, the power increases with the grid size N. Comparing the results for the three DGPs, we find that the test has higher power under the spatial IID than the two models with spatial dependence. This could be expected, as both the MA and AR models lead to some averaging of the $\xi_{i,j}$, bringing the observations $X_{i,j}$ a bit closer to normality. There is no apparent difference when using different kernels under the spatial IID, but the Bartlett kernel occasionally has marginally higher power under the spatial MA and AR models. An important observation is that different results are produced by using the two spatial long run variance estimators. When the *power* estimator is used, the power is monotonously decreasing as the degrees of freedom ν of the $\xi_{i,j}$ grow. However, this pattern does not occur when the *kernel* estimator is employed. To be specific for the *kernel* estimator, the expected power behavior is observed for $\nu > 8$, but not for $\nu \leqslant 8$. A reasonable explanation is that we use the 8th moment of the Student's t-distribution when estimating $\phi^2_{\mathcal{K}}$. However, the kth moment of a Student's t random variable is well-defined only for $k < \nu$. For the *power* estimator, we only use the 4th moment of observations in the spatial models generated by the Student's t random variable. Comparing Tables 2 and 3, we can conclude that the *power* estimator has better power properties than the *kernel* estimator. Additionally, the *power* estimator is more broadly applicable as it requires fewer moments of the data. We note that the kernel estimator requires the existence of first eight moments of the distribution, but we are still interested in the impact on power of the kernel estimator if some of the first eight moments do not exist. Thus, we also report the power of the kernel estimator for $\nu = 8, 5$ in Table 2. The empirical power when the skew-normal distribution is employed, has similar behavior, except that we do not see nonmonotonic

Table 2
The empirical power of 5000 independent simulations with significance level of 5% for the spatial normality test based on the *kernel* estimators $\hat{\phi}^2_{\mathcal{S},\text{kern}}$, $\hat{\phi}^2_{\mathcal{K},\text{kern}}$ for three DGPs of spatial IID ($X_{i,j} = 2 + \sqrt{2}\xi_{i,j}$), spatial moving-average ($X_{i,j} = \xi_{i,j} + 0.5\xi_{i,j-1}$), spatial autoregressive($X_{i,j} = 0.5X_{i-1,j-1} + \xi_{i,j}$), where $\xi_{i,j}$ are from i.i.d. Student's t-distribution with ν degrees of freedom.

Grid size	Kernel	$\nu = 20$	$\nu = 9$	$\nu = 8$	$\nu = 5$
Panel A: Spatial IID					
N = 25	Truncated	18.86%	43.42%	47.82%	54.08%
	Bartlett	9.18%	33.22%	38.92%	48.66%
	Flat-top	11.52%	36.70%	41.62%	50.00%
N = 50	Truncated	54.86%	90.14%	88.12%	76.64%
	Bartlett	53.52%	90.20%	88.10%	75.94%
	Flat-top	54.36%	90.22%	88.18%	76.14%
N = 100	Truncated	99.50%	98.28%	97.94%	87.10%
	Bartlett	99.50%	98.24%	97.94%	86.96%
	Flat-top	99.50%	98.26%	97.96%	87.06%
Panel B: Spatial moving average					
N = 25	Truncated	12.50%	28.80%	32.20%	45.78%
	Bartlett	6.68%	20.18%	23.22%	40.50%
	Flat-top	7.34%	21.54%	24.64%	40.58%
N = 50	Truncated	28.34%	82.56%	84.18%	73.96%
	Bartlett	27.62%	83.32%	84.78%	74.82%
	Flat-top	26.86%	81.94%	83.78%	73.58%
N = 100	Truncated	93.04%	98.14%	97.60%	86.30%
	Bartlett	93.48%	98.26%	97.76%	86.82%
	Flat-top	93.06%	98.10%	97.56%	86.24%
Panel C: Spatial autoregressive					
N = 25	Truncated	11.90%	22.80%	26.44%	41.64%
	Bartlett	7.44%	17.50%	20.58%	38.70%
	Flat-top	6.68%	16.38%	18.94%	35.84%
N = 50	Truncated	22.66%	75.30%	79.36%	74.06%
	Bartlett	24.10%	78.76%	81.70%	75.84%
	Flat-top	21.26%	75.16%	79.22%	73.48%
N = 100	Truncated	84.44%	97.70%	97.20%	86.32%
	Bartlett	86.60%	98.10%	97.50%	87.52%
	Flat-top	84.42%	97.76%	97.18%	86.16%

power for the *kernel* estimator; both estimators produce comparable results. The test is very powerful even for small departures of normality. Details are discussed in Section IV of the Supplement.

Broad conclusions Based on the simulations we performed, we recommend the implementation based on the *power* estimators (13) and any one of the three kernels listed in this section, with bandwidths given by (15).

We conclude this section by presenting in Table 4 the empirical size of the standard Jarque–Bera test. Under independence, this standard test has correct size, as does our test, but under spatial dependence it has overinflated size, while our test controls the size very well. The distortion increases as the nominal size decreases, and exceeds 100% of the nominal size at the 1 percent level.

3.2. Normality of Sea Surface Temperature anomalies

Sea Surface Temperatures (SSTs) are closely linked with El Niño/Southern Oscillation (ENSO) events, which are related to pattern changes in rainfall, wind speeds, ocean circulation, and general global weather patterns. The North Carolina Institute for Climate Studies (NCICS) provides monthly mean of daily Optimum Interpolation Sea Surface Temperature (OISST) analysis using Advanced Very High Resolution Radiometer (AVHRR) prepared for Observations for Model Intercomparisons Project by National Centers for Environmental Information (NCEI). The global SSTs are on a 1440×720 grid (in every 1/4 longitude degree and 1/4 latitude degree) observed daily for over 30 years, with missing pixels over land. The specific data we used was downloaded from the website https://esgf-node.llnl.gov/search/obs4mips. In the data set, there are 400 monthly observations in the period of September 1981 to December 2014. Fig. 2 shows a snapshot of the SST data in the month of September 1981.

Denote the SST observations by $Y_{i,j}(t)$, where t is a month, i is longitude, and j is latitude. These observations are available only for coordinates i, j which correspond to sea, not to land. For any sufficiently large region, and any month t, the observations $Y_{i,j}(t)$ cannot be considered as a realization of a *stationary* spatial field because of spatial trends in water temperature due to latitude, ocean currents and the shape of neighboring land. We must therefore transform these data

Table 3
The empirical power of 5000 independent simulations with significance level of 5% for the spatial normality test based on the *power* estimators $\hat{\phi}^2_{S,\text{pow}}, \hat{\phi}^2_{K,\text{pow}}$ for three DGPs of spatial IID ($X_{i,j} = 2 + \sqrt{2}\xi_{i,j}$), spatial moving-average ($X_{i,j} = \xi_{i,j} + 0.5\xi_{i,j-1}$), spatial autoregressive($X_{i,j} = 0.5X_{i-1,j-1} + \xi_{i,j}$), where $\xi_{i,j}$ are from i.i.d. Student's t-distribution with ν degrees of freedom.

Grid size	Kernel	$\nu = 20$	$\nu = 9$	$\nu = 8$	$\nu = 5$
Panel A: Spatial IID					
N = 25	Truncated	36.62%	87.62%	93.22%	99.84%
	Bartlett	36.64%	87.62%	93.22%	99.84%
	Flat-top	36.64%	87.62%	93.22%	99.84%
N = 50	Truncated	82.44%	100.00%	100.00%	100.00%
	Bartlett	82.44%	100.00%	100.00%	100.00%
	Flat-top	82.44%	100.00%	100.00%	100.00%
N = 100	Truncated	100.00%	100.00%	100.00%	100.00%
	Bartlett	100.00%	100.00%	100.00%	100.00%
	Flat-top	100.00%	100.00%	100.00%	100.00%
Panel B: Spatial moving average					
N = 25	Truncated	23.60%	67.98%	76.78%	98.02%
	Bartlett	24.50%	68.76%	77.54%	98.14%
	Flat-top	23.60%	68.00%	76.78%	98.02%
N = 50	Truncated	56.30%	99.36%	99.86%	100.00%
	Bartlett	57.02%	99.42%	99.86%	100.00%
	Flat-top	56.30%	99.36%	99.86%	100.00%
N = 100	Truncated	98.14%	100.00%	100.00%	100.00%
	Bartlett	98.26%	100.00%	100.00%	100.00%
	Flat-top	98.14%	100.00%	100.00%	100.00%
Panel C: Spatial autoregressive					
N = 25	Truncated	19.66%	59.84%	68.24%	96.00%
	Bartlett	22.60%	63.08%	71.02%	96.56%
	Flat-top	19.82%	60.16%	68.48%	96.04%
N = 50	Truncated	44.26%	97.74%	99.32%	100.00%
	Bartlett	47.40%	98.10%	99.42%	100.00%
	Flat-top	44.62%	97.78%	99.32%	100.00%
N = 100	Truncated	93.34%	100.00%	100.00%	100.00%
	Bartlett	94.10%	100.00%	100.00%	100.00%
	Flat-top	93.34%	100.00%	100.00%	100.00%

Table 4
The empirical sizes of Jarque–Bera test based on 5000 independent simulations with significant levels of 10%, 5%, and 1% for three DGPs of spatial IID ($X_{i,j} = 2 + \sqrt{2}\xi_{i,j}$), spatial moving-average ($X_{i,j} = \xi_{i,j} + 0.5\xi_{i,j-1}$), spatial autoregressive($X_{i,j} = 0.5X_{i-1,j-1} + \xi_{i,j}$), where $\xi_{i,j} \sim i.i.d.\ \mathcal{N}(0, 1)$.

Grid size	10%	5%	1%
Panel A: Spatial IID			
N = 100	9.62%	4.88%	1.02%
N = 500	10.48%	4.70%	0.98%
N = 1000	9.80%	5.18%	0.94%
Panel B: Spatial moving average			
N = 100	12.48%	6.66%	1.60%
N = 500	11.96%	6.74%	1.76%
N = 1000	12.84%	7.08%	1.68%
Panel C: Spatial autoregressive			
N = 100	15.20%	8.68%	2.36%
N = 500	14.88%	8.44%	2.30%
N = 1000	14.80%	8.10%	2.22%

to consider them as a realization of a stationary random field whose normality can be tested. A transformation that is of primary interest, see, e.g., NASA [24], is defined as follows. Compute the long term averages

$$A_{i,j}(T) = \frac{1}{T}\sum_{t=1}^{T} Y_{i,j}(t)$$

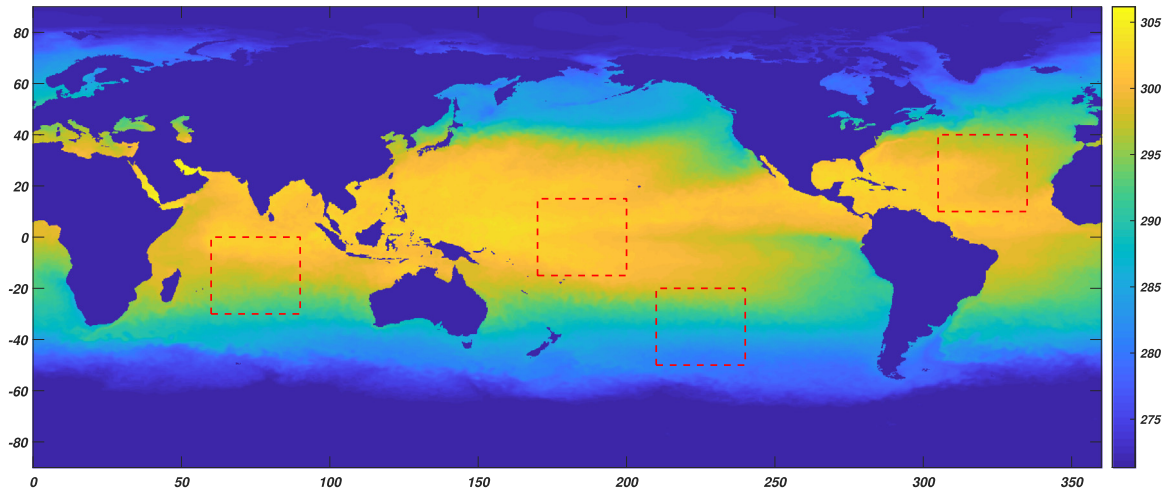

Fig. 2. SST data snapshot in September 1981 and the four selected regions. The global SST data is on a 1440 × 720 grid (in every 1/4 longitude degree and 1/4 latitude degree), with missing pixels over land. *Region 1* (longitude 60° to 90°, latitude −30° to 0°) lies in Indian Ocean and in the southern hemisphere. *Region 2* (longitude 170° to 200°, latitude −15° to 15°) is located in the Pacific Ocean and it is symmetric by the equator. *Region 3* (longitude 210° to 240°, latitude −50° to −20°) is in the Pacific Ocean but it is in the southern hemisphere, away from the equator. Region 4 (longitude 305° to 335°, latitude 10° to 40°) is in the Northern Atlantic. Each selected region is on a 120 × 120 grid, containing 30° of longitude and 30° of latitude.

Table 5
P-values of the spatial normality test for $D_{i,j}(t)$ in July. Bold numbers indicate rejections at the 5% significance level.

Date	*Kernel* estimator				*Power* estimator			
	Region 1	Region 2	Region 3	Region 4	Region 1	Region 2	Region 3	Region 4
Jul-1982	**3.3%**	**0.0%**	*36.7%*	**0.0%**	*9.0%*	**0.0%**	*34.5%*	**0.0%**
Jul-1983	*47.4%*	**0.0%**	**0.0%**	**0.0%**	*46.4%*	**1.8%**	**1.6%**	**0.0%**
Jul-1984	**0.0%**	**0.1%**	**0.0%**	**0.0%**	**0.0%**	**1.4%**	**2.5%**	**0.0%**
Jul-1985	**0.0%**	**0.0%**	**0.1%**	**0.0%**	**0.4%**	**0.0%**	**0.0%**	**0.4%**
Jul-1986	**4.7%**	**0.0%**	*84.8%*	**0.0%**	*5.4%*	**0.2%**	*85.8%*	**0.0%**
Jul-1987	*27.0%*	**1.4%**	*71.4%*	**0.0%**	*45.5%*	*11.8%*	*73.5%*	**0.0%**
Jul-1988	*11.9%*	**0.0%**	**1.5%**	**3.6%**	*8.0%*	**0.0%**	**3.2%**	*16.0%*
Jul-1989	**0.5%**	**0.0%**	**0.3%**	**0.0%**	**0.3%**	**0.0%**	*8.0%*	**0.0%**
Jul-1990	**0.1%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**1.5%**	**0.0%**
Jul-1991	**0.3%**	**0.0%**	**0.0%**	**0.0%**	*6.6%*	**0.0%**	**0.0%**	**0.0%**
Jul-1992	**0.0%**	*50.1%*	**0.0%**	**0.0%**	**0.0%**	*57.6%*	**0.1%**	**0.0%**
Jul-1993	*11.5%*	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.4%**	**0.0%**	**0.0%**
Jul-1994	**4.1%**	**0.0%**	**0.1%**	**0.0%**	*20.3%*	**0.0%**	**1.4%**	**0.0%**
Jul-1995	**0.0%**	*31.6%*	**1.9%**	**0.0%**	**0.1%**	*58.7%*	**1.8%**	**0.0%**
Jul-1996	*11.0%*	**0.0%**	**0.0%**	**0.0%**	*25.4%*	**0.0%**	**1.0%**	**0.1%**
Jul-1997	**0.8%**	**0.0%**	*74.2%*	**0.0%**	*10.2%*	**0.3%**	*52.5%*	**0.0%**
Jul-1998	**1.0%**	**0.0%**	**1.9%**	*15.6%*	*5.3%*	**0.0%**	**4.0%**	*33.7%*
Jul-1999	**0.1%**	**0.0%**	**0.0%**	*10.4%*	**3.3%**	**0.0%**	**0.0%**	*18.6%*
Jul-2000	**0.5%**	**0.0%**	**0.0%**	**0.0%**	**3.4%**	**0.0%**	**0.0%**	**0.0%**
Jul-2001	**0.0%**	**0.0%**	**0.0%**	**0.7%**	**0.0%**	**0.0%**	**0.1%**	*12.4%*
Jul-2002	*7.8%*	**0.0%**	**0.0%**	**0.0%**	*34.0%*	**0.0%**	**0.0%**	**0.0%**
Jul-2003	**0.0%**	**0.0%**	*57.8%*	**0.0%**	**0.0%**	**0.0%**	*49.6%*	**0.2%**
Jul-2004	**2.9%**	*18.5%*	*51.9%*	**0.0%**	**2.8%**	*26.5%*	*66.6%*	**0.1%**
Jul-2005	**1.5%**	*5.2%*	**0.0%**	**0.0%**	**0.0%**	**2.1%**	**0.1%**	**0.3%**
Jul-2006	**0.2%**	*23.8%*	*54.0%*	**0.0%**	**0.0%**	*40.3%*	*57.0%*	**0.0%**
Jul-2007	**0.0%**	**0.0%**	*8.9%*	**0.0%**	**0.0%**	**0.0%**	**0.9%**	**0.3%**
Jul-2008	*37.0%*	**0.0%**	**0.1%**	**0.0%**	*41.9%*	**0.0%**	**0.0%**	**0.0%**
Jul-2009	*34.3%*	**0.3%**	*48.9%*	**0.0%**	*45.6%*	*9.5%*	*56.8%*	**0.0%**
Jul-2010	**4.1%**	*9.7%*	**0.0%**	**0.0%**	*13.2%*	*21.5%*	**0.3%**	**0.0%**
Jul-2011	**0.0%**	**0.0%**	**0.6%**	**0.0%**	**0.3%**	**0.0%**	*13.8%*	**0.0%**
Jul-2012	**0.0%**	**0.0%**	**0.0%**	**0.1%**	**0.3%**	**0.0%**	**1.0%**	**2.5%**
Jul-2013	**0.0%**	**0.0%**	*12.1%*	**0.6%**	**0.1%**	**0.1%**	**4.0%**	**0.1%**
Jul-2014	**0.2%**	**0.0%**	*53.7%*	**0.8%**	**0.3%**	**0.0%**	*49.4%*	**2.0%**

where T is number of the same calendar months in the sample period. For example, if t corresponds to July, and we have $T = 33$ Julys in the sample period. The monthly anomalies are defined as

$$D_{i,j}(t) = Y_{i,j}(t) - A_{i,j}(T).$$

Table 6
P-values of the spatial normality test for $U_{i,j}(t)$ in July. Bold numbers indicate rejections at the 5% significance level.

Date	*Kernel* estimator				*Power* estimator			
	Region 1	Region 2	Region 3	Region 4	Region 1	Region 2	Region 3	Region 4
Jul-1982	*11.3%*	**0.0%**	**0.0%**	**0.0%**	*40.3%*	**0.0%**	**0.0%**	**0.0%**
Jul-1983	**1.3%**	**0.0%**	**0.1%**	**0.0%**	*13.1%*	**0.0%**	**0.7%**	**0.0%**
Jul-1984	**0.0%**	*6.8%*	**0.0%**	**0.0%**	**0.2%**	*29.3%*	**0.1%**	**0.0%**
Jul-1985	**0.0%**	**0.0%**	**1.2%**	**0.0%**	**0.1%**	**1.0%**	**3.2%**	**1.2%**
Jul-1986	**0.4%**	**0.0%**	**0.3%**	**0.0%**	**0.1%**	**0.0%**	*6.3%*	**0.0%**
Jul-1987	**3.9%**	**0.0%**	*77.7%*	**0.0%**	*5.9%*	**0.6%**	*83.5%*	**0.0%**
Jul-1988	*56.3%*	**0.0%**	*5.8%*	**0.0%**	*70.9%*	**0.1%**	*15.8%*	**0.0%**
Jul-1989	*19.4%*	**0.0%**	**0.0%**	**0.0%**	*40.4%*	**0.0%**	**0.1%**	**0.0%**
Jul-1990	*70.5%*	**0.1%**	**1.6%**	**0.1%**	*82.5%*	**0.3%**	*7.7%*	**0.4%**
Jul-1991	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.3%**	**1.9%**	**0.0%**	**0.0%**
Jul-1992	*8.0%*	**0.1%**	**0.0%**	**0.0%**	*12.1%*	**0.1%**	**0.0%**	**0.3%**
Jul-1993	*38.6%*	**0.0%**	**0.0%**	**0.2%**	*63.6%*	**0.0%**	**1.7%**	**2.5%**
Jul-1994	*17.9%*	**0.0%**	**0.0%**	**0.0%**	*43.1%*	**0.0%**	**1.5%**	**0.0%**
Jul-1995	**0.3%**	*43.0%*	*5.8%*	**0.0%**	*6.6%*	*47.3%*	*9.7%*	**0.0%**
Jul-1996	**0.0%**	**1.4%**	**1.4%**	**0.0%**	**2.3%**	*14.4%*	**3.4%**	**0.1%**
Jul-1997	**0.0%**	**0.0%**	*59.1%*	**0.0%**	**0.3%**	**0.0%**	*66.8%*	**0.0%**
Jul-1998	**0.2%**	**0.1%**	**0.1%**	**0.0%**	*6.9%*	**4.4%**	**1.9%**	**0.0%**
Jul-1999	**0.7%**	**0.0%**	**0.0%**	**3.2%**	*12.0%*	**0.0%**	**0.0%**	*7.6%*
Jul-2000	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**1.4%**	**0.0%**	**0.0%**	**0.0%**
Jul-2001	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.1%**	**0.0%**	**0.3%**
Jul-2002	**1.2%**	**0.0%**	**0.0%**	**0.0%**	*16.1%*	**1.1%**	**0.0%**	**0.0%**
Jul-2003	**0.1%**	**0.0%**	*53.9%*	**0.0%**	**0.3%**	**0.0%**	*48.8%*	**2.6%**
Jul-2004	*11.0%*	**0.0%**	**0.0%**	**0.0%**	*21.1%*	**0.0%**	**0.8%**	**0.1%**
Jul-2005	*96.4%*	*45.2%*	**0.0%**	**0.0%**	*97.4%*	*48.6%*	**0.5%**	**0.0%**
Jul-2006	**3.1%**	*46.0%*	**0.0%**	**0.1%**	**2.4%**	*60.6%*	**2.5%**	**0.8%**
Jul-2007	*10.4%*	**0.0%**	**0.0%**	**0.0%**	*10.1%*	**0.0%**	**0.0%**	**0.1%**
Jul-2008	**0.4%**	**0.0%**	*13.6%*	**4.7%**	*5.4%*	**0.0%**	**0.3%**	*6.7%*
Jul-2009	*35.2%*	**0.4%**	*6.1%*	**0.0%**	*59.5%*	*8.8%*	*18.0%*	**0.0%**
Jul-2010	**0.1%**	**0.0%**	*30.9%*	**0.0%**	**4.1%**	**0.0%**	*54.4%*	**0.0%**
Jul-2011	**3.9%**	**0.0%**	**0.0%**	**0.0%**	*19.8%*	**0.0%**	**0.0%**	**0.0%**
Jul-2012	**0.0%**	**0.0%**	**0.0%**	**0.0%**	**0.2%**	**0.0%**	**0.6%**	**1.2%**
Jul-2013	*13.3%*	**0.0%**	*34.2%*	*76.3%*	*25.5%*	**0.0%**	*44.9%*	*81.3%*
Jul-2014	*9.1%*	**0.0%**	**0.0%**	**0.0%**	*28.2%*	**0.0%**	**1.5%**	**1.2%**

They are deviations in a given year from what is typical for a given month at location (i, j). As quantified by French et al. [12], among others, surface temperatures exhibit complex spatial trends in their variability. These are more pronounced over continents (temperatures over coastal regions are less variable that those in the interior), but one can expect a similar, though smaller, effect over bodies of water. We therefore also consider standardized anomalies defined by

$$U_{i,j}(t) = \frac{Y_{i,j}(t) - A_{i,j}(T)}{SD_{i,j}(T)},$$

where

$$SD^2_{i,j}(T) = \frac{1}{T}\sum_{t=1}^{T}(Y_{i,j}(t) - A_{i,j}(T))^2.$$

As spatial domains, we selected four squared ocean regions with different characteristics. *Region 1* (longitude 60° to 90°, latitude -30° to 0°) lies in Indian Ocean and in the southern hemisphere. *Region 2* (longitude 170° to 200°, latitude −15° to 15°) is located in the Pacific Ocean and it is symmetric by the equator. *Region 3* (longitude 210° to 240°, latitude −50° to −20°) is also in the Pacific Ocean but it is in the southern hemisphere, away from the equator. The last region, Region 4 (longitude 305° to 335°, latitude 10° to 40°) is in the Northern Atlantic. The data over these regions are on a 120 × 120 grid, due to the fact that they all contain an area extending 30° of longitude and 30° of latitude. The four selected regions are highlighted in Fig. 2.

Conclusions from the application of the normality test We applied the implementations with both the *kernel* and the *power* estimator in order to see if the differences observed in Section 3.1 manifest themselves for the temperature data. It turns out that the *kernel* and the *power* estimators produce consistent results in the most cases, but not in all cases. We only reports results for the flat-top kernel as other kernels produce similar results.

The P-values for July in all years of the sample period for $D_{i,j}(t)$ and $U_{i,j}(t)$ are shown in Tables 5 and 6, respectively. The P-values for January, April, and October are provided in the Supplement. The most general observation is that normality of these spatial data cannot be assumed without further checks, so spatial statistics methods which rely on the assumption of Gaussianity must be used with caution. It might be best to use methods which do not assume Gaussianity. Comparing

the results for the two versions of monthly anomalies, $D_{i,j}(t)$ and $U_{i,j}(t)$, they generally lead to the same conclusion, but $U_{i,j}(t)$ tends to produce more acceptances of normality, indicated by the P-values greater than 5%. This effect is however not very large. By looking at the results in different four regions, we see that *Region 1*, which is in the Indian Ocean and in the southern hemisphere, is the one with the highest number of normality in the July monthly anomalies for all years. In particular, the test on the second version of monthly anomalies, $U_{i,j}(t)$, using the *power* estimator for the long run variance suggests the normality in 21 out of 33 years. On the opposite side, *Region 4*, which is located in the Atlantic Ocean and in the northern hemisphere, has the lowest number of acceptances of normality of the July monthly anomalies. Specifically, $U_{i,j}(t)$ with the *power* estimator only suggests the normality in 3 out of 33 years. These conclusions also hold for other months.

Acknowledgments

We are grateful to the Editor-in-Chief, Associate Editor, and reviewers for their valuable comments and suggestions, which have led to an improved version of this paper. We would also like to thank Professor Anil K. Bera for his helpful discussions at the 4th Workshop on "Goodness-of-fit, Change-point and Related Problems" (Trento, Italy, 2019). This research has been partially supported by the NSF, United States grant DMS-1914882.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2020.104640.

References

[1] E.B. Anderes, M.L. Stein, Estimating deformations of isotropic Gaussian random fields on the plane, Ann. Statist. 36 (2008) 719–741.
[2] T.W. Anderson, D.A. Darling, A test of goodness-of-fit, J. Amer. Statist. Assoc. 49 (1954) 765–769.
[3] D.W.K. Andrews, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59 (1991) 817–858.
[4] A. Azzalini, The Skew–Normal and Related Families, IMS, 2014.
[5] S. Banerjee, B. Carlin, A. Gelfand, Hierarchical Modeling and Analysis for Spatial Data, CRC Press, 2014.
[6] M.-C. Chang, S.-W. Cheng, C.-S. Cheng, Signal aliasing in Gaussian random fields for experiments with quantitative factors, Ann. Statist. 47 (2019) 909–935.
[7] N. Cressie, Statistics for Spatial Data, Wiley, 1993.
[8] N. Cressie, C. Wikle, Statistics for Spatio-Temporal Data, Wiley, 2011.
[9] R.B. D'Agostino, A. Belanger, R.B. D'Agostino Jr., A suggestion for using powerful and informative tests of normality, Amer. Statist. 44 (1990) 316–321.
[10] J.A. Doornik, H. Hansen, An omnibus test for univariate and multivariate normality, Oxford Bull. Econ. Stat. 70 (2008) 927–939.
[11] F. Fouedjio, N. Desassis, T. Romary, Estimation of space deformation model for non-stationary random functions, Spat. Stat. 13 (2015) 45–61.
[12] J. French, P. Kokoszka, S. Stoev, L. Hall, Quantifying the risk of heat waves using extreme value theory and spatio-temporal functional data, Comput. Statist. Data Anal. 131 (2019) 176–193.
[13] A.E. Gelfand, P.J. Diggle, M. Fuentes, P. Guttorp (Eds.), Handbook of Spatial Statistics, CRC Press, 2010.
[14] A. Gelfand, E. Schliep, Spatial statistics and Gaussian processes: A beautiful marriage, Spat. Stat. 18 (2016) 86–104.
[15] R. Guhaniyogi, S. Banerjee, Meta-kriging: Scalable Bayesian modeling and inference for massive spatial datasets, Technometrics 60 (2018) 430–444.
[16] N. Henze, B. Zirkler, A class of invariant consistent tests for multivariate normality, Comm. Statist. Theory Methods 19 (1990) 3595–3617.
[17] I.A. Ibragimov, Y.V. Linnik, Independent and Stationary Sequences of Random Variables, Wolters-Nordhoff, 1971.
[18] C.M. Jarque, A.K. Bera, Efficient tests for normality, homoskedasticity and serial independence of regression residuals, Econ. Lett. 6 (1980) 255–259.
[19] C.M. Jarque, A.K. Bera, A test of normality of observations and regression residual, Internat. Statist. Rev. 55 (1987) 163–172.
[20] M. Katzfuss, A multi-resolution approximation for massive spatial datasets, J. Amer. Statist. Assoc. 112 (2017) 201–214.
[21] S. Lahiri, P. Robinson, Central limit theorems for long range dependent spatial linear processes, Bernoulli 22 (2016) 345–375.
[22] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530.
[23] K.V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, Sankhya B 36 (1974) 115–128.
[24] NASA, Sea Surface Temperature Anomaly Global Maps, NASA Earth Observatory, 2019, http://earthobservatory.nasa.gov/global-maps/AMSRE_SSTAn_M, (Accessed April 2019).
[25] W. Newey, K. West, Automatic lag selection in covariance matrix estimation, Rev. Econom. Stud. 61 (1994) 631–653.
[26] D. Nychka, S. Bandyopadhyay, D. Hammerling, F. Lindgren, S. Sain, A multiresolution Gaussian process model for the analysis of large spatial datasets, J. Comput. Graph. Statist. 24 (2015) 579–599.
[27] C. Paciorek, B. Lipshitz, W. Zhuo, C. Kaufman, R. Thomas, Parallelizing Gaussian process calculations in R, J. Stat. Softw. 63 (2015) 1–23.
[28] D. Politis, Adaptive bandwidth choice, J. Nonparametr. Stat. 15 (2003) 517–533.
[29] A. Prause, A. Steland, Estimation of the asymptotic variance of univariate and multivariate random fields and statistical inference, Electron. J. Stat. 12 (2018) 890–940.
[30] J.P. Royston, An extension of Shapiro and Wilk's W test for normality to large samples, J. R. Stat. Soc. Ser. C Appl. Stat. 31 (2) (1982) 115–124.
[31] J.P. Royston, Some techniques for assessing multivarate normality based on the Shapiro–Wilk W, J. R. Stat. Soc. Ser. C Appl. Stat. 32 (2) (1983) 121–133.
[32] J.P. Royston, Approximating the Shapiro–Wilk W–test for non-normality, Stat. Comput. 2 (3) (1992) 117–119.
[33] P.D. Sampson, P. Guttorp, Nonparametric estimation of nonstationary spatial covariance structure, J. Amer. Statist. Assoc. 87 (1992) 108–119.
[34] O. Schabenberger, C.A. Gotway, Statistical Methods for Spatial Data Analysis, Chapman & Hall/CRC, 2005.
[35] A.M. Schmidt, A. O'Hogan, BayesIan inference for non-stationary spatial covariance structure via spatial deformations, J. R. Stat. Soc. Ser. B Stat. Methodol. 65 (2003) 743–758.
[36] F.W. Scholz, M.A. Stephens, K-sample Anderson–Darling tests, J. Amer. Statist. Assoc. 82 (1997) 918–924..

[37] S.S. Shapiro, M.B. Wilk, An analysis of variance test for normality (complete samples), Biometrika 52 (1965) 591–611.
[38] D.I. Sidorov, On mixing conditions for sequences of moving averages, Theory Probab. Appl. 54 (2010) 339–347.
[39] M.L. Stein, Interpolation of Spatial Data: Some Theory for Krigging, Springer, 1999.
[40] M.A. Stephens, EDF statistics for goodness of fit and some comparisons, J. Amer. Statist. Assoc. 69 (1974) 730–737.
[41] J. Stroud, M. Stein, S. Lysen, BayesIan and maximum likelihood estimation for Gaussian processes on an incomplete lattice, J. Comput. Graph. Statist. 26 (2017) 108–120.