Journal of Statistical Theory and Practice (2025) 19:35
https://doi.org/10.1007/s42519-025-00448-0

ORIGINAL ARTICLE

Intrinsic Random Function Kriging on the Sphere

Nicholas W. Bussberg[1] · **Jacob Shields**[2] · **Chunfeng Huang**[3]

Accepted: 6 April 2025 / Published online: 20 May 2025

Abstract
Intrinsic random functions (IRFs) can be used to model spatial processes on the sphere. To perform kriging using IRFs, one needs the knowledge of the IRF order (i.e., the degree of non-homogeneity) and its associate generalized covariance function, which are challenging in Euclidean spaces and hinder its practice. On the sphere, Huang et al. [1] showed that IRFs behave differently from their counterparts in Euclidean spaces and can be characterized by their lower-frequency truncated processes. Based on this, we develop procedures to estimate both the IRF order and the associated parametric generalized covariance function. Then, a truly IRF-based universal kriging can be implemented in practice on the sphere. We demonstrate our methods through simulations, and the numerical results show that IRF kriging outperforms ordinary kriging. Additionally, we apply our procedure to a global temperature dataset to determine the IRF order.

Keywords Non-homogeneity · Intrinsic Stationarity · Universal Kriging · Spatial Statistics

1 Introduction

The intrinsic random functions (IRF), developed by Matheron [2], generalize the concept of stationarity beyond the traditional second-order stationarity, providing a theoretical foundation for kriging. For example, one commonly used form of kriging in practice is ordinary kriging (see Cressie, [3]), which assumes intrinsic stationarity - corresponding to an IRF of order 0 in Matheron's notation. The associated generalized

✉ Nicholas W. Bussberg
nbussberg@elon.edu

Jacob Shields
jacob@biomedit.com

Chunfeng Huang
huang48@iu.edu

1 Department of Mathematics and Statistics, Elon University, Elon, North Carolina, USA

2 BiomEdit, Fishers, Indiana, USA

3 Department of Statistics, Indiana University, Bloomington, Indiana, USA

covariance function of this intrinsic stationary process can be expressed in the form of variograms. With a further assumption of isotropy, such a variogram can be readily estimated using method of moments [4]. Moreover, the variograms must be conditionally negative definite, a property that can be ensured through parametric modeling [5] or nonparametric procedures [6].

Ordinary kriging has been widely used in spatial statistics as a strong demonstration of IRF theory. However, extending it to higher-order IRFs and achieving a truly IRF-based universal kriging has been challenging. The difficulty arises from at least two major obstacles. First, the order of the IRFs play an essential role in universal kriging [2] and must be estimated beforehand. Higher orders of IRFs indicate weaker assumptions in stationarity. In Matheron's [2] notation, an IRF of order -1, or in short IRF(-1), assumes the typical second-order stationarity, while an IRF of order 0 assumes intrinsic stationarity. To the best of our knowledge, an effective IRF order selection procedure has yet to be developed in general Euclidean spaces. Cressie [3] introduced a graphical tool to select the order for integrated processes, which is a special case of an IRF in time series. Second, even if one can estimate (or assume) the order of IRF, the universal kriging procedure also relies on its generalized covariance function [2], which models the spatial dependency structure. For example, in ordinary kriging, variograms can be used in the place of the generalized covariance functions. For higher order IRFs, generalized covariance functions are defined in an abstract way, which makes it difficult to develop a statistical estimation procedure from observed data. These are major challenges and hurdles that make an IRF-based universal kriging less practical in spatial statistics applications.

In this manuscript, we attempt to address these two IRF-based kriging challenges when modeling processes are on the sphere. The solutions rely on the IRF extension from Euclidean spaces to circles and spheres in Huang et al. [1, 7]. In this extension, the geodesic distance replaces the chord distance, spherical harmonics substitute for monomials in Euclidean spaces, and rotation invariance is used instead of shift invariance. Huang et al. [1] formally define the IRFs on the sphere and prove that a process is an IRF of order κ if and only if its κ-lower-frequency truncated process is homogeneous (Theorem 1, [1]). The intuition behind Theorem 1 in Huang et al. [1] is, while the generalized differential operators can annihilate lower degree monomials in Euclidean spaces, they will not have the same effect on the spherical harmonics. Instead, Huang et al. [1] discovered that truncation operators can be used to annihilate the lower spherical harmonics.

Remark 1 On the sphere, stationary processes are more conventionally termed homogeneous processes. In this manuscript, these two terms are considered interchangeable, but we will prefer to use the term homogeneous.

Given this development, notice that a truncated process can be roughly approximated by a residual process when a regression on the lower spherical harmonics is applied to the observed data (e.g., [3, 8]). If this residual process can be tested for homogeneity, then we may be able to identify the order of the IRF. A test procedure for homogeneity can be found in Jun and Genton [9] for a spatio-temporal process on the sphere. In practice, this test requires a time component, which may not be read-

ily available. In this manuscript, we develop the procedure to identify the degree of non-homogeneity for spatial processes on the sphere without temporal information.

We develop a novel approach based on the residual process to infer the IRF order regardless of what order is present. Note that on the sphere, the order indicates the degree of non-homogeneity. Therefore, this procedure could be used to check whether a process is homogeneous or not in practice.

With the order selected, IRF-based kriging can be applied if we can estimate its associated generalized covariance function. Following Huang et al. [1], when the correct IRF order is reached, the corresponding truncated process is homogeneous, and its covariance function can be readily estimated. Based on the residual process, we can obtain the method of moments estimates of this covariance function. To ensure its conditional positive definiteness, we use parametric modeling in this manuscript. Therefore, a truly IRF-based kriging can be implemented for observations on the sphere. This is presented in Section 2.

In Section 3, numerical studies are conducted to compare IRF-based kriging and ordinary kriging, and we show the superior performance of IRF-based kriging. Simulation studies also show that our proposed IRF order selection methods (both graphical and numerical) work well. Additionally, a data analysis is conducted (Section 4), where temperature data on the sphere is shown to be non-homogeneous.

This manuscript develops statistical methods to select an IRF order and implement IRF-based kriging in practice. It follows the theory developed in Huang et al. [1] closely. While we will re-introduce some key notion and theorems in Huang et al. [1], some formula and technical details are shortened in this manuscript. The program can be found in our Github repository (https://github.com/nbussberg/IRF-kriging).

2 IRF Order Selection and Kriging

In this section, we develop truly IRF-based kriging on the sphere. While more technical details are in Huang et al. [1], we briefly introduce IRFs on the sphere here. A real-valued continuous random process $Z(x), x \in S^2$, where S^2 is an unit sphere, is an IRF of order κ, if the process $Z(g\lambda)$ is rotation invariant with respect to any rotation g on the sphere, where a measure λ is an allowable measure of order κ such that it annihilates the spherical harmonics of orders less than κ. That is,

$$\int_{S^2} Y_l^m(x)\lambda(dx) = 0, \quad 0 \le l < \kappa, \quad |m| \le l. \tag{1}$$

Here, the real-valued spherical harmonic functions are

$$\begin{cases} Y_l^m(x) = \sqrt{\frac{2l+1}{2\pi}\frac{(l-m)!}{(l+m)!}}\, P_l^m(\cos\zeta)\cos(m\psi), & m = 1, \ldots, l, \\ Y_l^0(x) = \sqrt{\frac{2l+1}{4\pi}}\, P_l^0(\cos\zeta), \\ Y_l^{-m}(x) = \sqrt{\frac{2l+1}{2\pi}\frac{(l-m)!}{(l+m)!}}\, P_l^m(\cos\zeta)\sin(m\psi), & m = 1, \ldots, l, \end{cases}$$

$P_l^m(\cdot)$ are the associated Legendre polynomials, $P_l^0(\cdot) \equiv P_l(\cdot)$ are the Legendre polynomials, and $x = (\psi, \zeta)$ with longitude $\psi \in [0, 2\pi)$ and latitude $\zeta \in [0, \pi]$. The process $Z(g\lambda) = \int_{S^2} Z(gx)\lambda(dx)$ being rotation invariant with respect to any rotation g means that $\mathrm{E}(Z(\lambda)) = \mathrm{E}Z(g\lambda)$ and $\mathrm{Cov}(Z(\lambda_1), Z(\lambda_2)) = \mathrm{Cov}(Z(g\lambda_1), Z(g\lambda_2))$, where λ, λ_1, and λ_2 are allowable measures of order κ (Equation 1).

This definition is the direct extension of IRFs in Euclidean spaces [2, 10] to the sphere, where geodesic distance, spherical harmonics, and rotation invariance are used instead of chord distance, monomials and shift invariance, respectively. Huang et al. [1] proceed to develop a theorem that says a real-valued continuous random process $Z(x)$ is an IRFκ if and only if its lower-frequency truncated process $Z_\kappa(x)$ is homogeneous. In particular, note that $Z(x)$ has the following expansion that is convergent in quadratic mean

$$Z(x) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Z_{l,m} Y_l^m(x),$$

so its corresponding lower-frequency truncated process is

$$Z_\kappa(x) = \sum_{l=\kappa}^{\infty} \sum_{m=-l}^{l} Z_{l,m} Y_l^m(x).$$

It is an equivalence theorem and can be used as the formal definition of an IRFκ. This equivalence theorem will help us deal with the two challenges mentioned in Section 1.

Remark 2 Note that in this manuscript, the indices for κ are different from the original notation for IRFs in Matheron [2]. In this manuscript, the order is increased by 1 relative to Matheron's notation. For example, second-order homogeneity is notated as IRF0, and intrinsic homogeneity is IRF1. This increment shift is because we use the inequality $0 \le l < \kappa$ in Equation (1). It can easily be switched back to be consistent with Matheron's notation by using $0 \le l \le \kappa$ instead. However, we feel it is more intuitive to denote a homogeneous process by IRF0 instead of the negative index in IRF(-1).

2.1 IRF Order Selection

Given this equivalence theorem, we now proceed to propose a statistical method to select the order based on observed data. The key component relies on the fact that when the truncated process $Z_\kappa(\cdot)$ is homogeneous, then the further truncated process $Z_{\kappa+1}(\cdot)$ is also homogeneous with a closely related covariance function. To see this, assume that $\{(x_i, Z(x_i), i = 1, \ldots, n\}$ are observed on the sphere, where $x_i \in S^2$, and $Z(x)$ is an IRFκ. Since $Z_\kappa(x)$ is a homogeneous process, this implies that its coefficients are uncorrelated [11, 12]

$$\mathrm{Cov}(Z_{l,m}, Z_{l',m'}) = a_l I(l, l') I(m, m'), \quad l, l' \ge \kappa, |m| \le l, |m'| \le l', \quad a_l \ge 0,$$

where $I(l, l') = 1$ if $l = l'$ and 0 otherwise is an indicator function. Therefore, its associated covariance function $\phi_\kappa(d(x, y))$ is

$$\phi_\kappa(d(x, y)) = \text{Cov}(Z_\kappa(x), Z_\kappa(y)) = \sum_{l=\kappa}^{\infty} \frac{2l+1}{4\pi} a_l P_l(\cos(d(x, y))),$$

where $d(x, y)$ is the geodesic distance between $x, y \in S^2$. This function plays the essential role of the generalized covariance function and is termed the intrinsic covariance function (ICF) in this manuscript. A further truncated process $Z_{\kappa+1}(x)$ is clearly also homogeneous, and its covariance function is

$$\text{Cov}(Z_{\kappa+1}(x), Z_{\kappa+1}(y)) = \sum_{l=\kappa+1}^{\infty} \frac{2l+1}{4\pi} a_l P_l(\cos(d(x, y))),$$

which is denoted as $\phi_{\kappa+1}(d(x, y))$. This covariance function shares the same coefficients a_l as in $\phi_\kappa(d(x, y))$ for $l \geq \kappa + 1$. The difference is then

$$\phi_\kappa(d(x, y)) - \phi_{\kappa+1}(d(x, y)) = \frac{2\kappa+1}{4\pi} a_\kappa P_\kappa(\cos(d(x, y))). \tag{2}$$

Therefore, if we have the estimates $\hat{\phi}_\kappa(h_j), \hat{\phi}_{\kappa+1}(h_j)$, $j = 0, 1 \ldots, n_h$ for spatial lags $h_0 = 0 < h_1 < \cdots < h_{n_h}$, we can obtain the following summation

$$S(\kappa) = \sum_{j=0}^{n_h} \left(\hat{\phi}_\kappa(h_j) - \hat{\phi}_{\kappa+1}(h_j) - \frac{2\kappa+1}{4\pi} a_\kappa P_\kappa(\cos(h_j)) \right)^2. \tag{3}$$

Given Equation 2, one may expect this sum of squares to be small. To fully compute this summation, one needs to compute a_κ. We can utilize a property of Legendre polynomials that $P_\kappa(\cos(0)) \equiv 1$. Then, based on Equation 2 we have

$$\phi_\kappa(0) - \phi_{\kappa+1}(0) = \frac{2\kappa+1}{4\pi} a_\kappa.$$

Therefore, we can rewrite $S(\kappa)$ in Equation 3 and obtain

$$S(\kappa) = \sum_{j=1}^{n_h} \left(\hat{\phi}_\kappa(h_j) - \hat{\phi}_{\kappa+1}(h_j) - \left\{ \hat{\phi}_\kappa(0) - \hat{\phi}_{\kappa+1}(0) \right\} P_\kappa(\cos(h_j)) \right)^2, \tag{4}$$

where the term with $j = 0$ is dropped from Equation 3. Then, it is left to find reasonable estimates for $\hat{\phi}_\kappa(\cdot)$ and $\hat{\phi}_{\kappa+1}(\cdot)$ to use Equation 4 in an IRF order selection procedure.

Note that the original spatial process $Z(x)$ can be partitioned as follows

$$Z(x) = \sum_{l<\kappa} \sum_{m=-l}^{l} Z_{l,m} Y_l^m(x) + Z_\kappa(x).$$

We may view this as a regression equation of $Z(x)$ on the lower spherical harmonics $Y_l^m(\cdot)$, $l < \kappa$, $|m| \leq l$, and obtain the residual process $Z_{\kappa,res}(x)$ to approximate the truncated process $Z_\kappa(x)$. Such use of residual processes is quite common in time series (e.g., [8]) and spatial statistics [3]. When the correct order κ is used, the truncated process $Z_\kappa(x)$ is homogeneous with the covariance function $\phi_\kappa(d(x, y))$. With the residual process $Z_{\kappa,res}(x)$, we can write the following method of moments estimate

$$\hat{\phi}_\kappa(h_j) = \frac{1}{|N_{h_j}|} \sum_{(x,y)\in N_{h_j}} Z_{\kappa,res}(x) Z_{\kappa,res}(y), \tag{5}$$

where N_{h_j} is the set of all pairs (x, y) such that $d(x, y) = T_\delta(h_j)$ with $T_\delta(h_j)$ being a tolerance region centered at h_j and δ ([3], Section 2.4), and $|N_{h_j}|$ is the cardinality of this set. A further truncated process $Z_{\kappa+1}(x)$ is also homogeneous with the covariance function $\phi_{\kappa+1}(d(x, y))$. We can obtain the estimates $\hat{\phi}_{\kappa+1}(h_j)$ based on the residual process $Z_{\kappa+1,res}(\cdot)$ of $Z(\cdot)$ regressed on $\{Y_l^m(x), l < \kappa + 1, |m| \leq l\}$. Then, we can compute $S(\kappa)$ in Equation 4.

We may compute $S(k)$ for any k following the same procedure. However, when $k < \kappa$, the lower frequency truncated process $Z_k(x)$ is not homogeneous. That is, its corresponding covariance function $\text{Cov}(Z_k(x), Z_k(y))$ is no longer a function of the geodesic distance $d(x, y)$, but instead it is governed by a non-homogeneous covariance function [1]. Thus, the notation $\hat{\phi}_k(d(x, y))$ in Equation 5 is not appropriate, so we introduce the notation $G(k, h_j)$, where

$$G(k, h_j) = \frac{1}{|N_{h_j}|} \sum_{(x,y)\in N_{h_j}} Z_{k,res}(x) Z_{k,res}(y). \tag{6}$$

To phrase another way, $Z_{k,res}(x)Z_{k,res}(y)$ does not share the same expectation anymore even when the pairs $(x, y) \in N(h_j)$ with the similar geodesic distance. Therefore, the averaging in $G(k, h_j)$ does not yield any reasonable statistical meaning. The corresponding

$$S(k) = \sum_{j=1}^{n_h} \left(G(k, h_j) - G(k+1, h_j) - \{G(k, 0) - G(k+1, 0)\} P_k(\cos(h_j))\right)^2, \tag{7}$$

will be arbitrary. On the other hand, when $k \geq \kappa$, then $G(k, h_j) = \hat{\phi}_k(h_j)$, where $\phi_k(d(x, y))$ is the covariance function of the truncated process $Z_k(x)$, which is homogenous. Here, we conclude that the quantity $S(k)$ will be small and remain small for any $k \geq \kappa$ and will be arbitrary when $k < \kappa$.

Next, we develop the procedure to select the IRF order (i.e., the degree of non-homogeneity) based on the previous discussion on the quantity $S(k)$. We can start by computing $S(k)$ for $k = 0, 1, 2, \ldots$ up to a fixed number. This fixed number can be pre-chosen based on data assumptions and the available computation capacity. For data observed from an IRFκ process, the quantities $S(k)$ will be any arbitrary numbers when

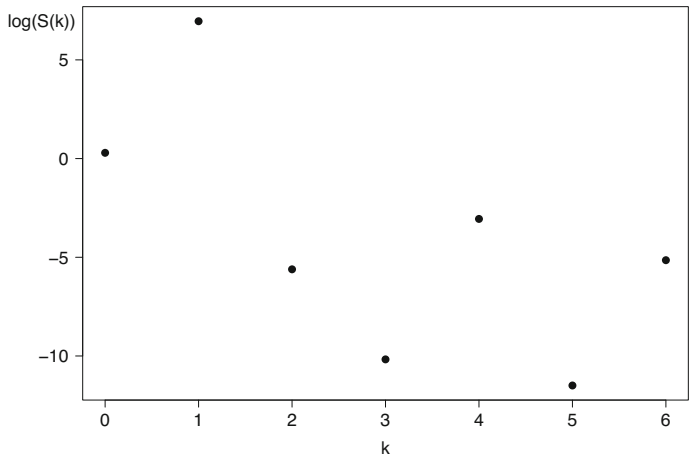

Fig. 1 Criterion plot of $\log(S(k))$ vs. varying values of k for a simulated IRF2 process

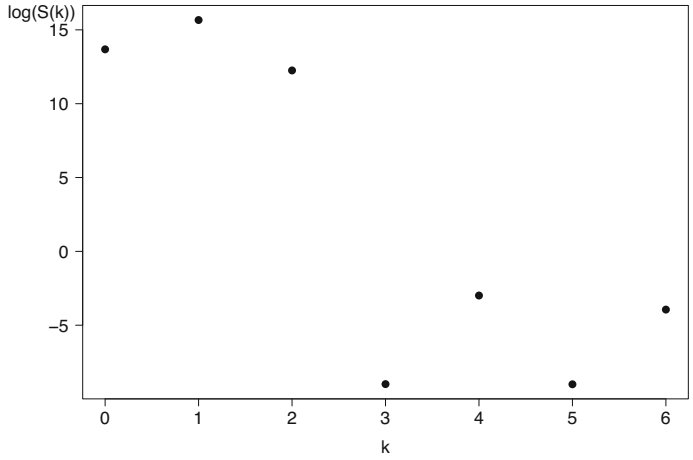

Fig. 2 Criterion plot of $\log(S(k))$ vs. varying values of k for a simulated IRF3 process

$k < \kappa$, but those quantities become relatively small and remain small when $k \geq \kappa$. To illustrate this, Figures 1 and 2 show the results from simulated IRFκ processes, where $\kappa = 2$ and 3, respectively. Details of the simulation will be discussed in the next section, but in Figure 1, where the true $\kappa = 2$, we can clearly see that $\log(S(k))$ is small and remains small when $k \geq 2$ and is relatively large in magnitude for $k = 0$ and 1. This is a clearly indication that the correct order is 2, or $\hat{\kappa} = 2$. The same pattern can be observed in Figure 2 with an IRF3 process: the quantities $\log(S(k))$ are large in magnitude when $k = 0$, 1, and 2 compared to $k \geq 3$, where the quantities $S(k)$ become comparably small and remain small. It is noteworthy that both figures show the $\log(S(k)$. The use of a log scale is useful in cases where $S(k)$ are very large, which is true for the IRF3 plot.

This graphical procedure is fundamental in revealing the non-homogeneity of the underlying process, and it is our primary recommendation for selecting the IRF order in practice. However, numerical procedures could be developed. Recall that $S(k)$ will be small and remains small when k reaches the correct order. This leads to two considerations in developing a numerical procedure. First, note that before k reaches the correct order, $S(k)$ is completely arbitrary. In our observations, it is usually relatively large, but it may be comparably small. This can be easily seen in Figure 1, where $S(0) = 1.3$, but $S(1) = 1051.6$. Such peaks are often observed in our experience. Therefore, the goal is to select the order after the peak in the $S(k)$ series. Second, when the correct order is reached, the series $S(k)$ will be small and remains small. This reflects the fact that an IRFκ is necessarily an IRF$(\kappa + 1)$ process, and so on. In principle, all these k may be appropriate. In statistical practice, one typically chooses the smallest such k for model simplicity, which means various penalty terms for higher orders are employed. Based on these two considerations, we reach the following numerical procedure:

$$\hat{\kappa} = \text{argmin}_{S(k)<\max\{S(k)\}} \{S(k) + \alpha k\}. \tag{8}$$

Unlike other selection procedures, such as AIC or BIC, the penalty coefficient α in Equation 8 is for purely numerical protection against choosing the largest order. As we have discussed, since $S(k)$ is usually much larger in magnitude when $k > \kappa$, most choices of α will work in practice. We use $\alpha = 0.5$ in this manuscript, but we will also include results for $\alpha = 0.2$ and 1.5.

It is clear from our discussion that it comes down to test whether the truncated process is homogeneous. The higher κ is, the more non-homogeneity is in the spatial process. Both the graphical and numerical procedures can be useful in practice to determine whether a truncated process is homogeneous on the sphere.

2.2 IRF-based Kriging Implementation

In previous section, we proposed procedures to choose the IRF order. Given such order $\hat{\kappa}$, we now turn to implementing the kriging formula developed in Huang et al. ([1], Equation 8). To do this, we need to estimate the generalized covariance functions of the IRF process. While the generalized covariance functions of the IRF in Euclidean spaces are defined in an abstract way, its counterpart on the sphere is much more straightforward. One can directly use the intrinsic covariance function (ICF) of the truncated process in the IRF-based kriging formula (see [1]). Conditional positive definiteness is needed for the generalized covariance function and thus the intrinsic covariance function. However, on the sphere, the ICF is positive definite, and the conditional positive definiteness is automatically satisfied. The extra complexity of the conditional positive definiteness can further be avoided on the sphere. To ensure the positive definiteness, one may use a parametric model approach, where a parametric model is developed to be positive definite, and one can estimate the parameter directly. Nonparametric approaches are a possibility, and these remain future research directions.

Parametric models that are positive definite on the sphere can be found in Huang et al. [6], Gneiting [13], and Porcu [14]. These models to be used for ICFs by directly expanding them in spherical harmonics and only keeping the higher order components. To illustrate this, we use the following simple model in this manuscript,

$$\phi(h,r) = \frac{1-r^2}{4\pi}(1-2r\cos(h)+r^2)^{-3/2}, \quad 0 \le r < 1,$$

where h is the geodesic distance, and r is the parameter. This is clearly positive definite on the sphere with the expansion (generating function)

$$\phi(h,r) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} r^l P_l(\cos(h)).$$

This leads to a direct IRF of order κ:

$$\phi_\kappa(h,r) = \sum_{l=\kappa}^{\infty} \frac{2l+1}{4\pi} r^l P_l(\cos(h)).$$

This can then be written as a parametric model in r, where

$$\phi_\kappa(h,r) = \frac{1-r^2}{4\pi}(1-2r\cos(h)+r^2)^{-3/2} - \sum_{l=0}^{\kappa-1} \frac{2l+1}{4\pi} r^l P_l(\cos(h)). \tag{9}$$

Other ICF parametric models can be similarly developed.

To use this parametric model in the kriging formula, we need to estimate the parameter. For Model 9, we can use the residual process to estimate the parameter r. Various methods can be used, for example, MLE, generalized least squares, and weighted least squares [3, 14, 15]. We use weighted least squares in this manuscript:

$$\sum_{j=0}^{n_h} |N_{h_j}| \left(\frac{G(\hat{\kappa}, h_j)}{\phi_{\hat{\kappa}}(h_j, r)} - 1 \right)^2.$$

With the IRF order selected and the ICF estimated, we can now implement kriging on the sphere to make predictions on unsampled locations. This is truly an IRF-based kriging procedure. We are not aware of any such development in Euclidean spaces. In the next section, we conduct numerical studies and explain how to carry out these analyses in practice.

3 Numerical Studies

In this section, we first discuss how to simulate an IRFκ process on the sphere. Then, we conduct simulation studies to assess the numerical IRF order selection procedure and to compare IRF-based kriging and ordinary kriging.

To simulate a Gaussian IRFκ on the sphere, we need its covariance function. While an IRFκ is characterized by its intrinsic covariance function $\phi_\kappa(d(x, y))$, the covariance function of the original process is not homogeneous and has the following form

$$H_\kappa(x, y) = \phi_\kappa(d(x, y)) - \sum_{\nu=1}^{\kappa^2}\{\phi_\kappa(d(x, \tau_\nu))p_\nu(y) + \phi_\kappa(d(y, \tau_\nu))p_\nu(x)\} + \sum_{\nu=1}^{\kappa^2}\sum_{\mu=1}^{\kappa^2}\phi_\kappa(d(\tau_\nu, \tau_\mu))p_\nu(x)p_\nu(y) + \sum_{\nu=1}^{\kappa^2}p_\nu(x)p_\nu(y), \tag{10}$$

where $\{\tau_1, \ldots, \tau_{\kappa^2}\} \in S^2$ and $p_1(\cdot), \ldots, p_{\kappa^2}(\cdot) \in N$ such that $p_\nu(\tau_\mu) = I(\nu, \mu)$ for $1 \le \nu, \mu \le \kappa^2$. This is clearly a non-homogeneous covariance function, and it can be shown to be a reproducing kernel [1].

In our simulation studies, IRF processes with $\kappa = 2$ were simulated on 1,500 randomly selected locations on the sphere. The parametric ICF model 9 was used with $r = 0.75$. For computational and interpretability reasons, we computed $G(k, h_j)$ (Equation 6) using orders $k = 0$ up to 6. In Equation 10, the four locations τ_ν, $\nu = 1, \ldots, 4$ were set to be

$$\left\{\left(\frac{\pi}{9}, \frac{\pi}{3}\right), \left(\frac{\pi}{3}, \frac{5\pi}{6}\right), \left(\frac{2\pi}{3}, \frac{6\pi}{5}\right), \left(\frac{8\pi}{9}, \frac{5\pi}{3}\right)\right\}.$$

Note that these four locations can be chosen arbitrarily. Figure 3 displays one simulated IRF2 process, with the locations τ_ν shown as white triangles. Similarly, we simulate IRF3 processes, and show an example in Figure 4. Note that now $\kappa = 3$, nine locations τ_ν, $\nu = 1, \ldots, 9$ need to be specified. We use the following locations (also shown as white triangles in Figure 4): $(\pi/12, \pi/6)$, $(\pi/9, \pi/3)$, $(\pi/6, 2\pi/3)$, $(\pi/3, 5\pi/6)$, $(\pi/2, \pi)$, $(2\pi/3, 6\pi/5)$, $(5\pi/6, 3\pi/2)$, $(8\pi/9, 5\pi/3)$, and $(11\pi/12, 9\pi/5)$.

Using this simulation setup, 1,000 simulated datasets for both IRF2 and IRF3 were created. For each simulated dataset, the numerical procedure (Equation 8) was used to select κ. Choices of α for the procedure were 0.2, 0.5, and 1.5, though 0.5 is used in Figures 5 and 6. For the IRF2 simulation, the numerical procedure with $\alpha = 0.5$ correctly selected the order to be 2 84.7% of the time, with just 0.7% below 2 (Figure 5). Setting $\alpha = 0.2$ performed worst, correctly choosing the order to be 2 only 69.0% of the time; however, only 0.5% of the orders were below 2.

As for the IRF3, the similar histogram is shown in Figure 6, where the numerical procedure selected 3 with 99.9% of the time, with only 0.01% below 3. These percentages were the same regardless of the choice of α for the procedure ($\alpha = 0.2$, 0.5, and 1.5). Overall, these simulations show that our numerical procedure works quite well.

Among kriging practices, ordinary kriging is one of the more commonly used procedures. Ordinary kriging on the sphere assumes intrinsic homogeneity [6], which is exactly an IRF1. It is clearly a special case of the IRF-based kriging discussed in this manuscript. Here, we conduct numerical experiments to compare truly IRF-based kriging to ordinary kriging predictions.

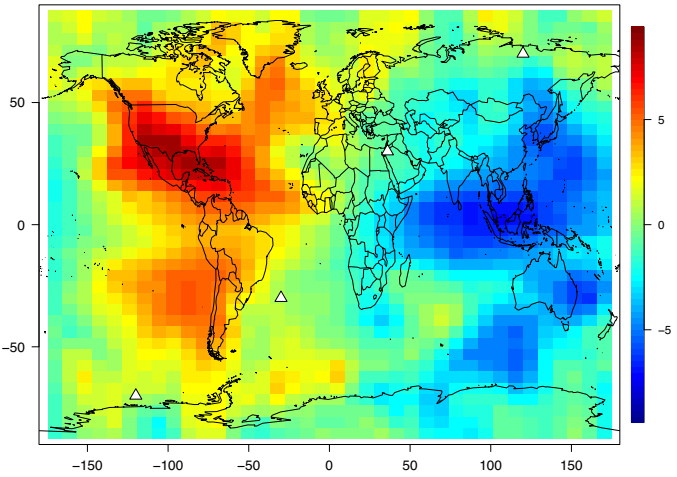

Fig. 3 Heat map of an IRF2 simulated processes. The arbitrary locations τ necessary for the simulation are plotted as white triangles

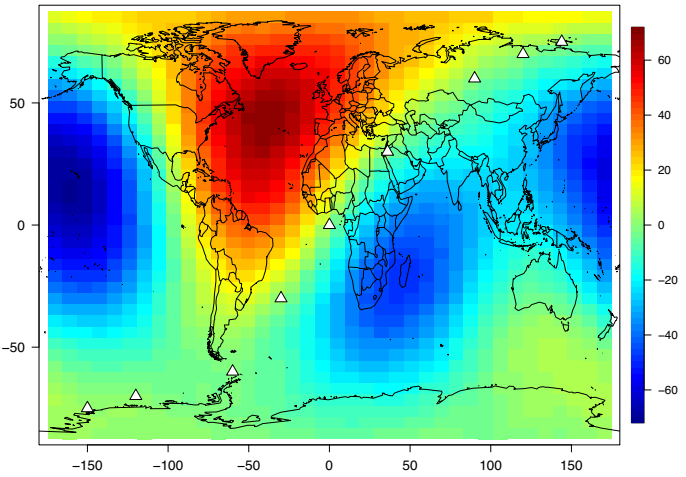

Fig. 4 Heat map of an IRF3 simulated processes. The arbitrary locations τ necessary for the simulation are plotted as white triangles

We simulated IRF2 and IRF3 processes using the simulation details discussed earlier. Then, the simulated datasets were split into training and testing subsets so that IRF-based kriging and ordinary kriging predictions could be compared against known values. In these simulations, 90% of the data were randomly selected to be the training dataset, and the other 10% were used as the testing dataset. For IRF-based kriging, the degree of non-homogeneity first needed to be estimated. Using the graphical procedure shown in Figure 7, it is clear that $\hat{\kappa} = 2$ for the IRF2 simulated data. Note that Figure 1 and Figure 7 are very similar, but Figure 7 was generated with 90% of the data whereas Figure 1 was created with the entire dataset. Then, the

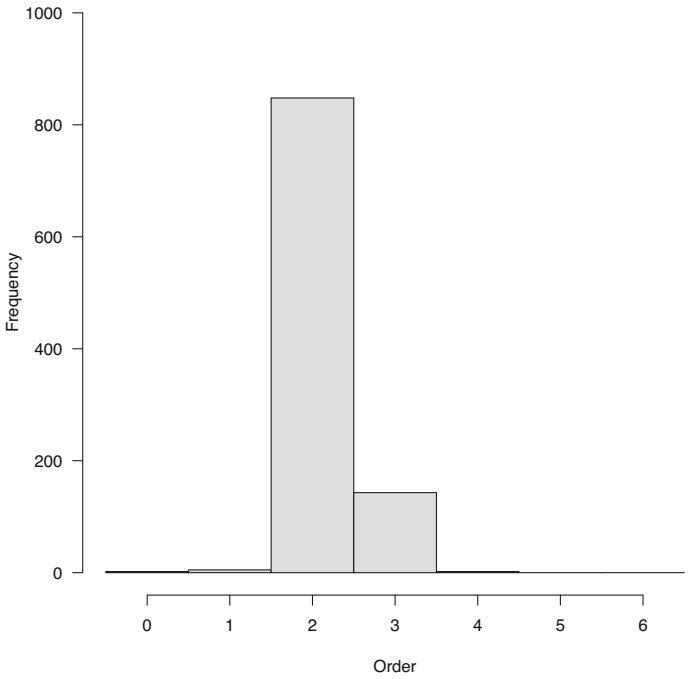

Fig. 5 Histogram showing $\hat{\kappa}$ for 1000 simulated IRF2 datasets with $\alpha = 0.5$. "Order" on the x-axis represents the value of $\hat{\kappa}$.

Table 1 Estimates for κ and r under ordinary kriging (OK) and IRF universal kriging (UK) using the training data subset. "True κ" represents the order from which the data were simulated. $\hat{r}$ was determined using weighted least squares. Root-mean squared error (RMSE) was calculated by comparing predictions to the testing subset values

True κ	Method	$\hat{\kappa}$	$\hat{r}$	RMSE
2	UK	2	0.714	0.04
2	OK	1	0.889	10.03
3	UK	3	0.717	0.04
3	OK	1	0.993	61.53

weighted least squares approach was used to estimate the parameter r in Equation 9 with $\hat{\kappa} = 2$. This produced $\hat{r} = 0.714$, which is quite close to the true parameter used in the simulation, $r = 0.75$. IRF-based kriging was then performed on the 10% of locations in the testing dataset. The actual values in the testing dataset were compared against the predictions by calculating the root mean square error (RMSE), which was only 0.03 (Table 1).

As for ordinary kriging, the assumption is an intrinsically homogeneous process, which implies $\hat{\kappa} = 1$. Weighted least squares was also used to estimate r in this case for the parametric model 9, where $\hat{\kappa} = 1$ by assumption. The estimate for r was $\hat{r} = 0.889$. This is clearly further from $r = 0.75$ than the IRF-based approach discussed earlier

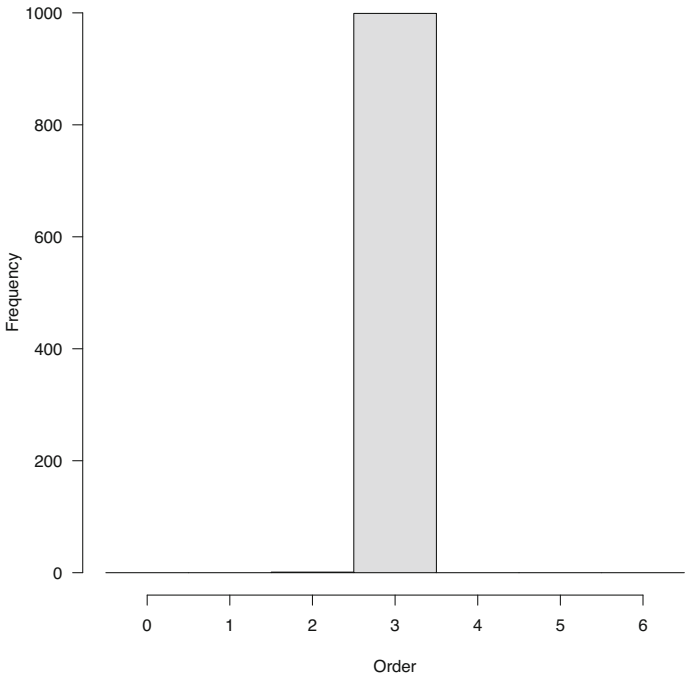

Fig. 6 Histogram showing $\hat{\kappa}$ for 1000 simulated IRF3 datasets with $\alpha = 0.5$. "Order" on the x-axis represents the value of $\hat{\kappa}$

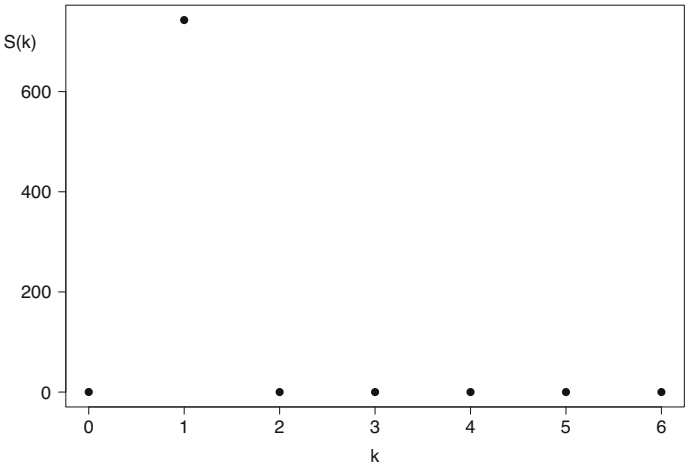

Fig. 7 Criterion plot of $S(k)$ for the simulated IRF2 process used in the kriging simulation study. The plot was generated with the training dataset, which consisted of a randomly selected 90% of the full dataset

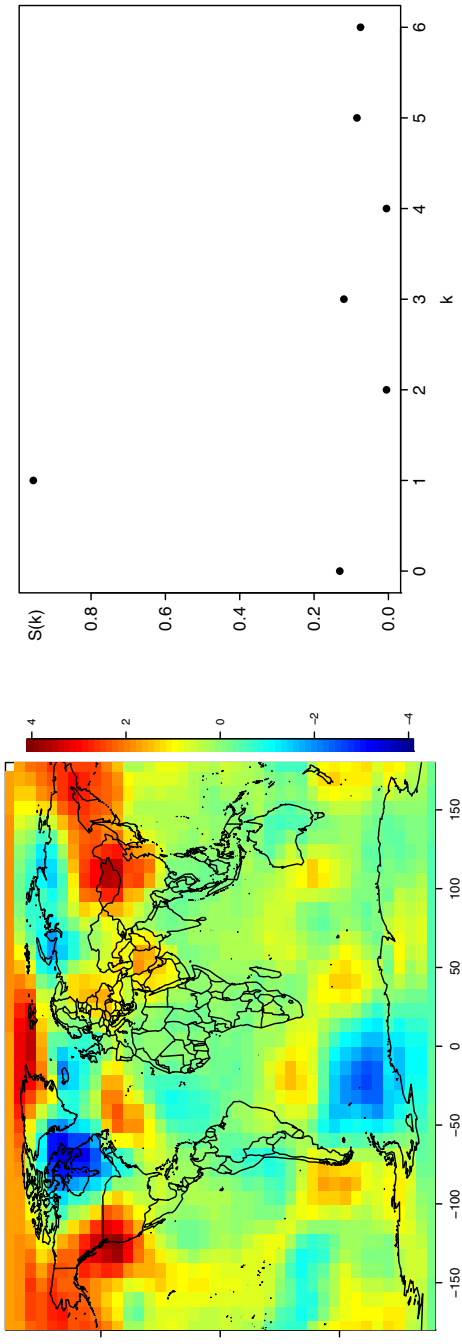

Fig. 8 Heat map (left) and criterion plot (right) for temperature anomaly data for January 2015. Data from National Space Science and Technology Center [16]

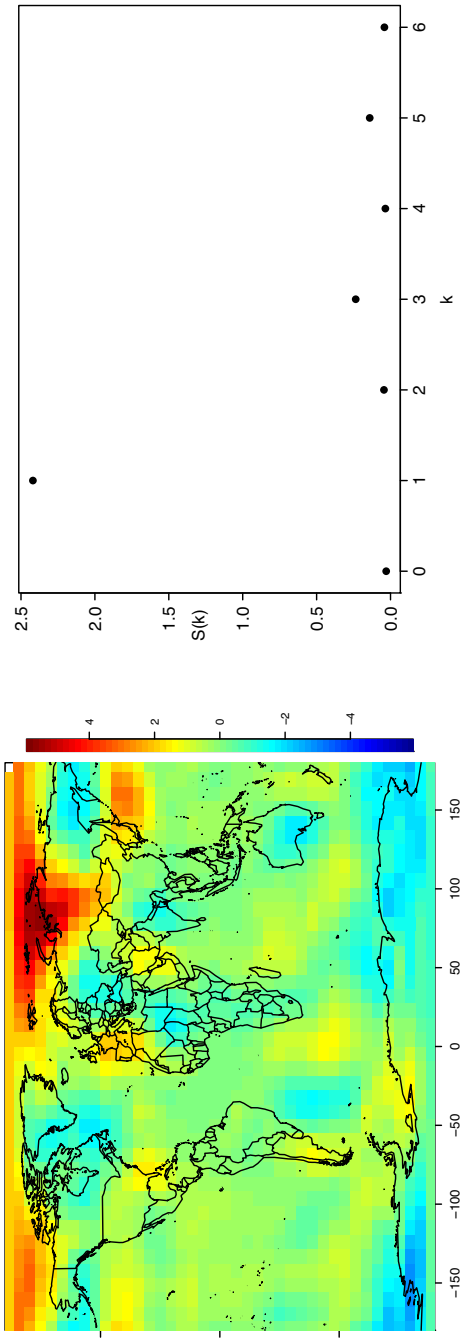

Fig. 9 Heat map (left) and criterion plot (right) for temperature anomaly data for April 2015. Data from National Space Science and Technology Center [16]

(Table 1). Ordinary kriging was then performed on the same 10% locations in testing dataset. We computed RMSE to be 10.08 (Table 1). Compared to RMSE of 0.03 in the IRF-based kriging approach, we can conclude that the ordinary kriging did not perform as well in these simulations. The same comparisons were done with IRF3 simulations, where the RMSEs were 0.06 and 61.53 for IRF-based kriging and ordinary kriging, respectively. These two examples clearly demonstrates the advantage of IRF-based kriging versus ordindary kriging. The intrinsic homogeneity assumption in ordinary kriging can be often too strict. Given the popularity of ordinary kriging in practice, this study provides evidence that a truly IRF-based kriging shall be preferred.

4 Application: Global Temperature Anomaly Data

Because it is commonly accepted that intrinsic homogeneity is too strong an assumption for many datasets, it is of interest to show that IRFκ data with $\kappa > 1$ exist in real-world applications. Further, the results in Section 3 are only relevant if there exist data for which the IRF-based kriging approach provides an advantage over other methods, e.g., ordinary kriging.

To investigate this issue, we used global temperature anomaly data from the National Space Science and Technology Center [16]. This data consists of satellite radiance measurements to determine global temperature anomalies in the lower troposphere for each month. The satellite readings are represented on a regular latitude-longitude grid, ranging from -88.75 to 88.75 degrees latitude and -178.75 to 178.75 degrees longitude and forming a 2.5×2.5 degree area.

Our goal in examining the temperature anomaly data was not to determine the prevalence of IRFκ with $\kappa > 1$, but rather to show examples of IRFκ data with $\kappa > 1$ process. With this in mind, both January and April of 2015 (Figures 8 and 9) were identified as IRF2 datasets.

Using the numerical procedure to find κ (Equation 8 with $\alpha = 0.5$) alongside graphical techniques (Figures 8 and 9), both January and April 2015 had $\hat{\kappa} = 2$. This brief exploration shows that there are indeed processes that are IRF2, but more importantly that the assumption of intrinsic homogeneity is not valid on at least some data. Combined with the results of Section 3, this implies that the proposed IRF-based procedure will likely lead to more accurate predictions in some real-world data applications. At a minimum, the procedures offer a technique to estimate κ and confirm what variations of kriging could or should be used. Finally, it is worth noting that these IRF order selection techniques are independent of parametric modeling. We can determine the degree of non-homogeneity without needing to model the rest of the data's structure.

Author Contributions Drs. Shields and Huang led the development of the graphical estimation procedure. Drs. Bussberg and Huang led the development of the universal kriging procedure and associated simulation studies. Dr. Bussberg led the manuscript development and temperature data analysis.

Funding Open access funding provided by the Carolinas Consortium.

Declarations

Conflicts of interests The authors do not have any conflicts of interests. All code is available on our Github repository: https://github.com/nbussberg/IRF-kriging.

References

1. Huang C, Zhang H, Robeson SM, Shields J (2019) Intrinsic random functions on the sphere. Statistics & Probability Letters 146:7–14
2. Matheron G (1973) The intrinsic random functions and their applications. Advances in Applied Probability 5(3):439–468
3. Cressie N (1993) Statistics for Spatial Data. John Wiley & Sons, New Jersey, USA
4. Matheron G (1963) Principles of geostatistics. Economic geology 58(8):1246–1266
5. Cressie N (1985) Fitting variogram models by weighted least squares. Journal of the International Association for Mathematical Geology 17(5):563–586
6. Huang C, Zhang H, Robeson SM (2011) On the validity of commonly used covariance and variogram functions on the sphere. Mathematical Geosciences 43(6):721–733
7. Huang C, Zhang H, Robeson SM (2016) Intrinsic random functions and universal kriging on the circle. Statistics & Probability Letters 108:33–39
8. Brockwell PJ, Davis RA (2016) Introduction to Time Series and Forecasting, 3rd edn. Springer, Switzerland
9. Jun M, Genton MG (2012) A test for stationarity of spatio-temporal random fields on planar and spherical domains. Statistica Sinica, 1737–1764
10. Chiles J-P, Delfiner P (2012) Geostatistics: Modeling Spatial Uncertainty vol. 497, 2nd edn. John Wiley & Sons, New Jersey, USA
11. Obukhov AM (1947) Statistically homogeneous fields on a sphere. Usp. Mat. Nauk 2(2):196–198
12. Yaglom AM (1961) Second-order homogeneous random fields. In: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, pp. 593–622
13. Gneiting T (2013) Strictly and non-strictly positive definite functions on spheres. Bernoulli 19(4):1327–1349
14. Porcu E, Bevilacqua M, Genton MG (2016) Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. Journal of the American Statistical Association 111(514):888–898
15. Stein ML (1999) Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York, USA
16. NSSTC (2017) Temperature anomaly data from 1978 to present. National Space Science and Technology Center (NSSTC)