مجلهی پژوهشهای آماری ایران
J. Statist. Res. Iran **17** (2020): 235–251
دورهی ۱۷، شمارهی ۱، بهار و تابستان ۱۳۹۹، صص ۲۳۵–۲۵۱
DOI: 10.52547/jsri.17.1.235
DOR: 20.1001.1.25385771.2020.17.1.12.3

A New Class of Spatial Covariance Functions Generated by Higher-order Kernels

Nafiseh Vafaei[†,*], Jorge Mateu[‡], Masoud Ganji[†] and Mohammad Ghorbani[⋆]

[†] University of Mohaghegh Ardabili
[‡] Jaume I University
[⋆] Luleå University of Technology

Received: 2022/20/11 Approved: 2023/26/02

Abstract. Covariance functions and variograms play a fundamental role in exploratory analysis and statistical modelling of spatial and spatio-temporal datasets. In this paper, we construct a new class of spatial covariance functions using the Fourier transform of some higher-order kernels. Moreover, we extend this class of spatial covariance functions to the spatio-temporal setting using the idea used in Ma (2003).

Keywords. Bochner's theorem, characteristic function, covariance model, higher-order Kernels, spatial data.

1 Introduction

Suppose that $\{Z(u); u \in D \subset \mathbb{R}^d\}$ is a spatial random process observed at n fixed locations $u_1, \ldots, u_n$. In most practical applications, it is assumed that $d = 2$ or 3. In spatial data analysis, the main goal is usually to optimally

[*] Corresponding author

predict unobserved parts of the process based on the observations. The most important ingredient to achieve this goal is an appropriate model that provides information about how the observations co-vary with respect to each other in space. In other words, for optimal prediction we need a suitable covariance or variogram model. The variogram is a suitable tool for describing the degree of spatial dependence of a spatial random field or stochastic process $Z(.)$. Thus, a fundamental concept for making inferences about the stochastic process under study is a covariance or variogram function, and a new class of models is a welcome contribution to the analysis of spatial and spatio-temporal data.

We now review a few important concepts of spatial statistics. Throughout the paper, we assume that the spatial process $Z(\cdot)$ satisfies the regularity condition, i.e., $\mathbb{V}ar[Z(u)] < \infty$, for all $u \in D$, which implies that the first two moments exist. By this assumption, at each location point u in D we can define the mean function as

$$\mu(u) = \mathbb{E}[Z(u)]$$

and for every two location points u_1 and u_2 in D, the covariance and the semivariogram functions are respectively defined as

$$C(u_1, u_2) = \mathbb{C}ov[Z(u_1), Z(u_2)]$$

and

$$\gamma(u_1, u_2) = \frac{1}{2}\mathbb{V}ar[Z(u_1) - Z(u_2)],$$

provided that they exist. The variogram is twice of the semivariogram. If the spatial process has a constant mean μ, the semivariogram corresponds to the expected value for the quadratic increase of the values between the locations u_1 and u_2 (Wackernagel (2003)), i.e.,

$$\gamma(u_1, u_2) = \frac{1}{2}\mathbb{E}[Z(u_1) - Z(u_2)]^2.$$

We also assume that the process $\{Z(u)\}$ is second-order stationary, meaning that the mean function is constant and the covariance function depends on the difference between two distinct points only, i.e., for some function C_0,

$$C(u_1, u_2) = C_0(u_1 - u_2).$$

The function C_0 is a valid covariance function if it is even and satisfies the positive definiteness condition. That is, for any $u_1, \ldots, u_m$ and reals $a_1, a_2, \ldots, a_m$, and any positive integer m, C_0 must satisfy

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j C_0(u_i - u_j) \geq 0.$$

For a continuous covariance function C_0 evaluated at spatial lag $h = u_i - u_j$ this is equivalent to the Bochner's theorem, which states that C_0 is positive definite if and only if it can be represented as

$$C_0(h) = \int e^{ih^{\mathsf{T}}\omega} \mathrm{d}F(\omega),$$

where F is a non-decreasing, right continuous, and bounded real-valued function called the spectral measure of $C_0(h)$ (Ripley (1981); Lindgren (2012); Finkenstadt et al. (2007)) and h^{T} is the transpose of h. If F is absolutely continuous with respect to the Lebesgue measure, then $\mathrm{d}F(\omega) = f(\omega)\mathrm{d}\omega$ and $f(\omega)$ is called the spectral density. Therefore for any non-decreasing, continuous, and bounded real-valued function F, the covariance function C_0 is its characteristic function.

There is another type of stationarity, called intrinsic stationarity, which is based on the variogram, and it is more general than second-order stationarity since there are processes for which the variogram is well defined but the covariance is not. The process $Z(.)$ is said to be intrinsically stationary, if its mean function is constant and the variogram depends only on the spatial distances $u_1 - u_2$ for every $u_1, u_2 \in \mathbb{R}^d$. The corresponding semivariogram for some function γ_0 is denoted by $\gamma_0(u)$ and said to be an intrinsically stationary semivariogram. If the process $Z(.)$ is second-order stationary, then it is also intrinsically stationary and

$$\gamma(h) = C(0) - C(h), \qquad u \in^d . \tag{1}$$

Hence, statistical methods for second-order stationary random fields can be represented using either the semivariogram or the covariance function. While statisticians are more familiar with variances and covariances, many geostatisticians prefer the semivariogram. Working with $\gamma(.)$ has distinct advantages over $C(.)$, especially when estimating these functions from observational data (see e.g., Cressie (1993); Wikle and Cressie (2019)).

Further, a second-order stationary random process $Z(\cdot)$ is said isotropic if its covariance function $C(h)$ (or the variogram function $\gamma(h)$) only depends on $\|h\|$, where $\|.\|$ indicates the Euclidean distance. Hereafter, we assume that the process $Z(\cdot)$ is isotropic. To assure positive definiteness, it is usually assumed that the covariance function C_0 belongs to a parametric family whose members are known to be positive definite. That is, one assumes that

$$C_0(h) = C^0(h;\theta), \tag{2}$$

where C^0 satisfies the positive definiteness condition for all $\theta \in \Theta \subset \mathbb{R}^d$. The vector parameter θ usually consists of one or more of the following parameters: the nugget effect, the sill, the partial sill, and the range. The nugget effect is given by $\gamma(h)$ when $h \to 0$, while the sill is $\gamma(h)$ when $h \to \infty$. The partial sill is the difference between the sill and the nugget effect. The range is the lag beyond which the dependence between the locations vanishes and the observations do not affect each other. There are several commonly used parametric models such as exponential, Gaussian, spherical, power exponential, Cauchy, and Matérn in the literature for the variogram and covariance functions in geostatistical modelling, see, e.g. Cressie (1993), Ripley (1981).Spatio-temporal covariances can be constructed directly as the product of purely valid spatial and purely valid temporal covariance functions. The drawback of these models is that they cannot model space-time interaction. Therefore, nonseparable stationary spatio-temporal covariance functions have been developed and studied in the literature to model space-time interactions, see, e.g. Cressie and Huang (1999), Jones and Zhang (1997), Brown et al. (2000), and Christakos (2000). To model the nonstationarity of a random field, Guttorp (1992)introduced a class of nonstationary spatial covariance functions and Smith (2001)presented a class of nonstationary spatio-temporal covariance functions by means of the convolution of a kernel function and a white noise process.

In this paper, we introduce new classes of spatial and spatio-temporal stationary covariance functions using the characteristic functions of absolutely continuous higher order kernels. A class of higher order kernels, which can be viewed as an extension of second order kernels, has some attractive properties such as smoothness, manageable convolution formulas, and Fourier transforms. See Schucany and Sommers (1977) for details. They have mainly been used for bias reduction in kernel density estimation (Marron, (1994); Hansen (2005); Tsuruta and Sagae (2017)).They have currently been employed: to capture filtrations of stochastic processes (Salvi et al. (2021)),to impute mixed-attribute datasets (Das et al. (2019));and to recognize CNN features with higher-order pooling (Cherian et al. (2017)).The rest of this paper is organized as follows. In Section 2, we give a brief overview of higher-order kernels. Section 3 introduces new covariance functions generated by higher-order kernels in space, and in Section 4, we develop the obtained models for the spatio-temporal case. Section 5 presents the analysis of Swiss rainfall data. The paper ends with some conclusions.

2 Higher-order Kernels

In this section we introduce some notation and define higher order kernels that we will use in the following sections. A function $K(x)$ is called a symmetric kernel if $K(x) = K(-x)$ and $\int_{\mathbb{R}} K(x)\mathrm{d}x = 1$. Further, it is called s-smooth if for $s \geq 1$ its $(s-1)$-th derivative, i.e. $K^{(s-1)}$, is absolutely continuous on $\mathbb{R}$. For any integer $j \geq 1$, we define the jth moment of the kernel K with $\mu_j(K) = \int_{\mathbb{R}} x^j K(x)\mathrm{d}x$.

Definition 1. The order of a kernel, r, is defined as the order of its first non-zero moment. A kernel K is a higher-order kernel if $r > 2$.

For example, if $\mu_1(K) = \mu_2(K) = \mu_3(K) = 0$ but $\mu_4(K) > 0$, then K is a fourth-order kernel and $r = 4$. The higher-order kernels are also called bias-reducing kernels (see Hansen (2005) and Marron (1994)).

Remark 1. The order of a symmetric kernel is always even.

Remark 2. Symmetric non-negative kernels are second order kernels.

A special function which we will use throughout the paper is the spherical Bessel function. For any integer $m \geq 0$, it is given by

$$j_m(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{\left(\frac{3}{2}\right)_{m+k} k!} \left(\frac{x}{2}\right)^{m+2k}. \tag{3}$$

Here $\left(\frac{3}{2}\right)_{m+k}$ obeys Pochhammer's symbol given by

$$(\alpha)_n = \alpha(\alpha+1)(\alpha+2)\ldots(\alpha+n-1) = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)},$$

where $\Gamma(\cdot)$ is the gamma function.

2.1 Müller's Kernel

We now consider a new class of s-smooth bounded-support kernels of order $2r$ introduced by Muller (1984).This class of s-smooth, $2r$th-order kernel on [-1,1] minimizes the mean integrated square error in kernel density estimation. Hansen (2005)presented an alternative representation of Müller's kernel functions. We use Hansen's representation of an s-smooth, $2r$th-order kernel for building a new class of spatial covariance functions. Following Hansen (2005),for any integer $r \geq 1$, the Müller's s-smooth, $2r$th-order kernel on [-1,1] is given by

$$M_{2r,s}(x) = B_{r,s}(x)M_s(x), \tag{4}$$

where

$$B_{r,s}(x) = \frac{\left(\frac{3}{2}\right)_{r-1}\left(\frac{3}{2}+s\right)_{r-1}}{(s+1)_{r-1}} \sum_{k=0}^{r-1} \frac{(-1)^k \left(\frac{1}{2}+s+r\right)_k x^{2k}}{k!(r-1-k)!\left(\frac{3}{2}\right)_k},$$

and for $s \geq 0$

$$M_s(x) = \frac{\left(\frac{1}{2}\right)_{s+1}}{s!}(1-x^2)^s.$$

Note that $M_s(x)$ is a special case of $M_{2r,s}(x)$ for the case $r = 1$. Granovsky and Muller (1991)showed that

$$\lim_{s\to\infty} \frac{1}{\sqrt{2s}} M_{2r,s}\left(\frac{x}{\sqrt{2s}}\right) = G_{2r}(x) = \frac{(-1)^r \phi^{(2r-1)}(x)}{2^{r-1}(r-1)!x}, \tag{5}$$

where $\phi^{(2r-1)}(x)$ is the $(2r-1)$th derivative of a standard normal density, and thus G_{2r} is the higher-order Gaussian kernel (Wand and Schucany (1990)). Wand and Schucany (1990)have shown that (5) can be represented as

$$G_{2r}(x) = \sum_{k=0}^{r-1} \frac{(-1)^k \phi^{(2k)}(x)}{2^k k!}. \tag{6}$$

This representation facilitates the calculation of its characteristic function (For more explanation see Wand and Schucany (1990).

Table 1. Isotropic stationary higher-order Gaussian covariance functions

r	1	2	3	4
$\tilde{G}_{2r}(h)$	$\tilde{\phi}(h)$	$\tilde{\phi}(h)(1+h^2/2)$	$\tilde{\phi}(h)(1+h^2/2+h^4/8)$	$\tilde{\phi}(h)(1+h^2/2+h^4/8+h^6/48)$

3 Spatial Covariance Functions Generated by Higher-order Kernels

We construct here a family of spatial covariance functions by using the characteristic function of the above kernels. Toward this end, for any function g, we denote its characteristic function by $\tilde{g}$.

3.1 Higher-order Gaussian Covariance Functions

For the higher-order Gaussian kernels given in (6), the characteristic function is given by $\tilde{G}_{2r}(h) = \exp(-h^2/2)\sum_{k=0}^{r-1} h^{2k}/(2^k k!)$, which introduce a new class of spatial covariance functions. Particularly, for $r = 1$, $\tilde{G}_2(h) = \exp(-h^2/2)$ is the classical Gaussian covariance model which we denote by $\tilde{\phi}(h)$, see, e.g. Wackernagel (2003).Isotropic stationary higher-order Gaussian covariance functions for some special values of r are listed in Table 1. The semivariograms of the functions given in Table 1 are demonstrated in Figure 1. They behave the same at the two ends but are slightly different along the support. The family of higher-order Gaussian covariance functions can be generalized by considering *Laguerre polynomials*, see, e.g. Fasshauer (2007).We leave such extensions for discussion in future work.

3.1.1 Müller's Covariance Functions

Another class of stationary spatial covariance function is obtained by using the characteristic function of Müller's s-smooth, $2r$th-order kernel. The characteristic function of the Müller's higher-order kernel is given by

$$C_1(h) = \tilde{M}_s(h) = \left(\frac{3}{2}\right)_s \left(\frac{2}{h}\right)^s j_s(h) \tag{7}$$

and

$$C_2(h) = \tilde{M}_{2r,s}(h) = \frac{2}{\sqrt{\pi}} \left(\frac{2}{h}\right)^s \sum_{m=0}^{r-1} \alpha_s(m) j_{s+2m}(h), \tag{8}$$

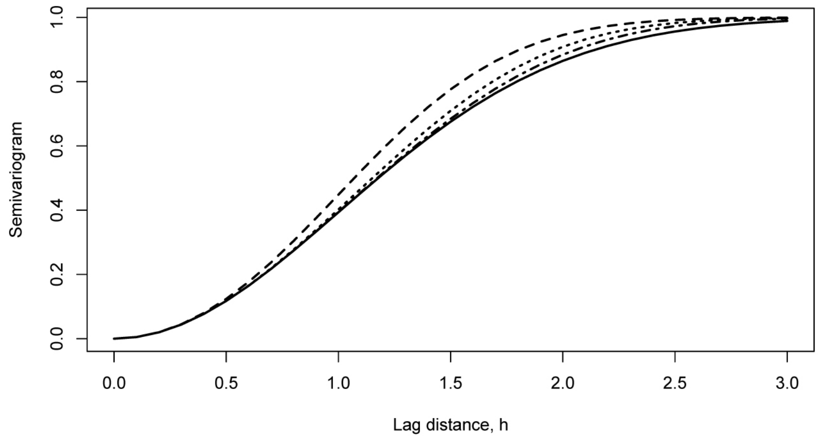

Figure 1. Isotropic stationary higher-order Gaussian semivariogram functions: Gaussian (solid line), Gaussian with order=4 (dashed line), Gaussian with order=6 (dotted line), and Gaussian with order=8 (dot-dashed line).

where

$$\alpha_s(m) = \frac{\Gamma\left(\frac{1}{2} + m + s\right)\left(\frac{1}{2} + 2m + s\right)}{m!}.$$

See more details in Hansen (2005).It is easy to show that (7) is another representation of the Bessel covariance function (Yaglom (1987), p. 139).Further, it can be shown that the Matérn covariance function (Matern (1960)) which is related to the characteristics function of the student's t distribution (Hurst (1995)) is a special case of (8). Thus (8) provides a general class of covariance functions for different values of r.

For computational purpose, we use the recurrence formula of the Bessel function, $j_m(h)$, given by

$$j_{m+1}(h) = \frac{2m+1}{h} j_m(h) + j_{m-1}(h)$$

with the initial condition

$$j_0(h) = \frac{\sin(h)}{h}$$

and

$$j_1(h) = \frac{\sin(h)}{h^2} - \frac{\cos(h)}{h}.$$

In the special case by setting $r = 1$ and $s = 0$ in (8), the hole effect (in some literature called the sine wave) model,

$$\tilde{M}_{2,0}(h) = \frac{\sin(h)}{h}, \tag{9}$$

is obtained. Note that the hole effect model is the characteristic function of a continuous uniform random variable on $[-1, 1]$. This model can be reparametrized as

$$C(\mathbf{h}, \theta) = \sigma_e^2 + \sigma^2 \frac{\eta}{\|h\|} \sin\left(\|h\|/\eta\right), \qquad \mathbf{h} \neq \mathbf{0},$$

$\theta = (\sigma_e^2, \sigma^2, \eta)$, where $\sigma_e^2 \geq 0$, $\sigma^2 \geq 0$ and $\eta \geq 0$ denote the nugget effect, the sill and the smoothing parameter, respectively. Considering (1), the corresponding semivariogram is given by

$$\gamma(\mathbf{h}, \theta) = \sigma^2 \left(1 - \frac{\eta}{\|h\|} \sin\left(\|h\|/\eta\right)\right), \qquad \mathbf{h} \neq \mathbf{0}.$$

This model is used for modelling the data where the empirical variogram shows strong cyclicity with decreasing amplitudes for increasing lag distances. In the literature, so far, this was the only example of a process where the covariance has a ciclicity behavior as a function of the distance h, and can be directly obtained from the characteristic function of a kernel function. Using this class of covariance models we are able to generate other processes where the covariance function has a sinusoidal behavior but not as strong as a sine wave. For example, consider the model $\tilde{M}_{2,1}(h)$ given by

$$\tilde{M}_{2,1}(h) = \frac{3}{h}\left[\frac{\sin(h)}{h^2} - \frac{cos(h)}{h}\right],$$

A parametric version of this model is given by

$$\tilde{M}_{2,1}(\mathbf{h}, \theta) = \sigma_e^2 + \sigma^2 \frac{3\eta}{\|h\|}\left[\sin\left(\frac{\|h\|}{\eta}\right)\frac{\eta^2}{\|h\|^2} - cos\left(\frac{\|h\|}{\eta}\right)\frac{\eta}{\|h\|}\right], \tag{10}$$

where $\sigma_e^2 \geq 0$, $\sigma^2 \geq 0$, and $\eta \geq 0$. We call this model the *sine-cosine wave*

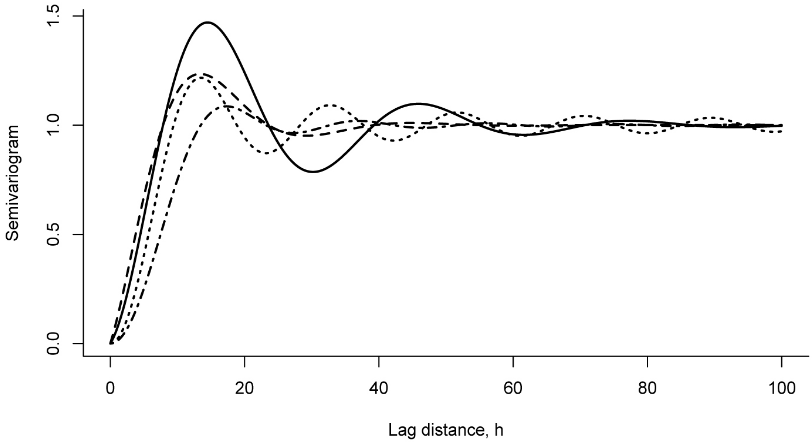

Figure 2. Examples of isotropic stationary semivariogram functions: Exponential-cosine composite semivariogram with $\nu = 60$ and $\eta = 5$ (solid line) and with $\nu = 30$ and $\eta = 5$ (dashed line). Sine-cosine wave with $\eta = 3$ (dot-dashed line) and sine wave with $\eta = 3$ (doted line). In all models $\sigma_e^2 = 0$ and $\sigma^2 = 1$.

model. As shown in Figure 2, this model can be used in a situation where the cyclicity weakens due to large fluctuations along the domain. Moreover, it can be used instead of or together with the cosine-based composite covariance functions obtained by the product of the cosine covariance function and other positive definite covariance functions. For example, the *cosine-exponential* composite model

$$C(\mathbf{h}, \theta) = \sigma_e^2 + \sigma^2 \exp(-3\|h\|/\nu) cos(\|h\|/\eta), \qquad \sigma^2 > 0,\ \nu > 0,\ \eta > 0, \tag{11}$$

results from the product of the covariance models *cosine* and *exponential*, see further details in Yaglom (1987), page 122.

4 Spatio-temporal Covariance Functions

Here we use the idea from Cressie and Huang (1999) and Ma (2003) to extend the generated spatial covariance function to the spatio-temporal setting. We consider a real-valued random process $Z(u;t)$ indexed in space by $u \in \mathbb{R}^d$ and in time by $t \in T$. As in the spatial case, the spatio-temporal dependence is usually characterized by the covariance function

$$C(u_1, u_2; t_1, t_2) = Cov[Z(u_1; t_1), Z(u_2; t_2)]. \tag{12}$$

The covariance function $C(u_1, u_2; t_1, t_2)$ is a well-defined space-time covariance function if $\mathbb{V}ar[Z(u;t)] < \infty$ (Cressie (1993).The spatio-temporal stochastic process $Z(u;t)$ is called second-ordery stationary if the mean function $\mu(u;t) = \mathbb{E}[Z(u;t)]$ is constant and the covariance function (12) is a function of the spatial distance $h = u_1 - u_2$ and temporal lag $t = t_1 - t_2$. Further, the process is called isotropic if $h = \|h\|$ and $t = t$. Assuming stationarity and isotropy, for a function C^0, we denote the covariance function by $C^0(h;t)$. Corollary 1.1 in ? states that for a constant vector $\beta \in \mathbb{R}^d$, if $C_S(h)$ is a stationary covariance function on $\mathbb{R}^d$, then

$$C(h;t) = C_S(h + \beta t), \qquad (h;t) \in \mathbb{R}^d \times \mathbb{R} \tag{13}$$

is a stationary covariance function on $\mathbb{R}^d \times \mathbb{R}$. The same idea with slightly different notation has been used in Cressic and Huang (1999).Thus, applying the above corollary to the covariance functions in (7) and (8) we obtain the following stationary space-time covariance functions,

$$C_1(h;t) = C_1(h + \beta t) = \left(\frac{3}{2}\right)_s \left(\frac{2}{h + \beta t}\right)^s j_s(h + \beta t).$$

and

$$C_2(h;t) = C_2(h + \beta t) = \frac{2}{\sqrt{\pi}} \left(\frac{2}{h + \beta t}\right)^s \sum_{m=0}^{r-1} \alpha_s(m) j_{s+2m}(h + \beta t),$$

Combining the above corollary with (9) and (11), we obtain the correspondence to the hole effect model and to the sine-cosine wave model in the space-time domain in the special case as follows,

$$\tilde{M}_{2,0}(h + \beta t) = \frac{\sin(h + \beta t)}{h + \beta t}, \tag{14}$$

and

$$\tilde{M}_{2,1}(h+\beta t) = \frac{3}{h+\beta t}\left[\frac{\sin(h+\beta t)}{h^2} - \frac{cos(h+\beta t)}{h+\beta t}\right]$$

are wave spatio-temporal covarince functions. Perspective and contour plots of (14) are shown in Figure 3.

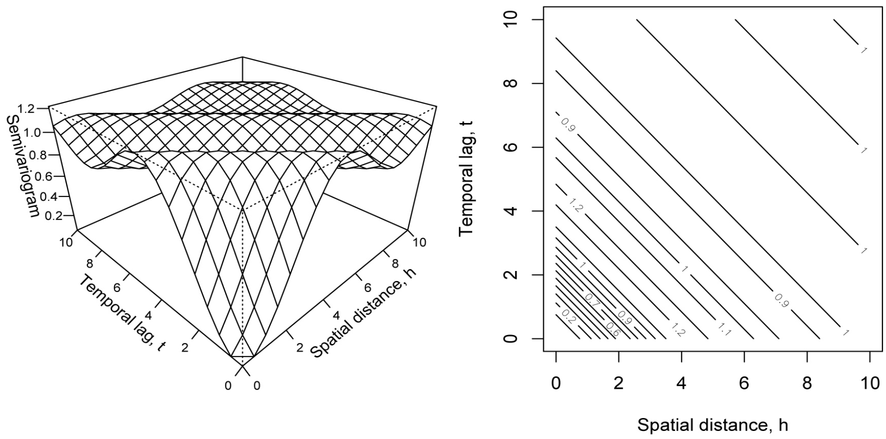

Figure 3. Isotropic stationary spatio-temporal covariance function given in (14) with $\beta = 1$.

5 Swiss Rainfall Data

In this section we illustrate how our sine-cosine wave model is applied to the Swiss rainfall data. The data are records of rainfall measured at 467 locations in Switzerland on May 8, 1986. The data analyzed in this section are from http://www.leg.ufpr.br/doku.php/pessoais:paulojus:mbgbook:datasets. This data collection was part of a workshop organized by AI-GEOSTATS to compare the different methods currently used to analyze spatial data, see Dubois (1998) for a detailed description of the data and the project.

Let us assume that data $\{Z(u_i) : i = 1, \ldots, n\}$ can be modeled with an stationary process. Under the stationarity assumption, a natural estimator based on the method of moment (Matheron (1962)),is

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|}\sum_{N(h)} (Z(u_i) - Z(u_j))^2, \quad h \in \mathbb{R}^d,$$

where $N(h) = \{(u_i, u_j) : u_i - u_j = h, i, j = 1, \ldots, n\}$ and $|N(h)|$ is the number of distinct pairs with distance h. This is called the empirical estimator of variogram. To estimate the parameters of a parametric variogram model $2\gamma(h, \theta)$, we minimize the weighted least square (WLS) method, which measure the weighted discrepancy between the empirical variogram and the parametric variogram. That is we minimize

$$Q(\theta) = \sum_{i=1}^{k} \left[\log(2\hat{\gamma}(h_i)) - \log(2\gamma(h_i, \theta))\right]^2 N(h_i)/2,$$

where, k denotes the number of lag distances at which the empirical and theoretical semivariograms are computed, and recall that $\hat{\gamma}(h)$ is the empirical semivariogram (see Cressie (1993),page 69 for more detail).

We obtained the empirical semivariogram, $\hat{\gamma}(h)$, using variog function inthe R package **geoR** (Ribeiro Jr and Diggle (2001)).To determine the WLS estimate of θ, the R package **nloptr** (Ypma (2017))was used. First, we used the DIRECT-L method to obtain a global optimum $\bar{\theta}$. Afterwards, to polish the optimum to a greater accuracy, we used $\bar{\theta}$ as a starting point for the local optimization 'bound-constrained by quadratic approximation' (BOBYQA) algorithm (Powell (2009))and obtained a final estimate $\hat{\theta}$. The WLS estimates of the sine-cosine wave model (10) parameters when $\sigma_e^2 = 2766$ (by the empirical variogram) are $\hat{\sigma} = 103.17$ and $\hat{\eta} = 14.13$.

Figure 4 shows the empirical semivariogram for the data together with simulated pointwise 0.05 significance envelopes obtained from 39 simulations of the sine-cosine wave model (such envelopes are obtained for each value of h by calculating the smallest and largest simulated values of $\hat{\gamma}(h)$; see Section 4.3.4 in Moller and Waagepetersen (2004)).We used the R package **geoR** with some slightly modifications in its functions geoRCovModels, cov.spatial and grf for the simulations. For the sine-cosine wave model, $\hat{\gamma}(h)$ is within the shaded envelopes area for all value of h which indicates that the sine-cosine wave fits the data adequately.

6 Conclusion

The covariance function is the key indicator in the analysis of dependent data, including spatial and spatio-temporal data sets, where there is a spatial or spatio-temporal dependence between observations due to dependence among their position in space or in space-time. The higher order kernels have been

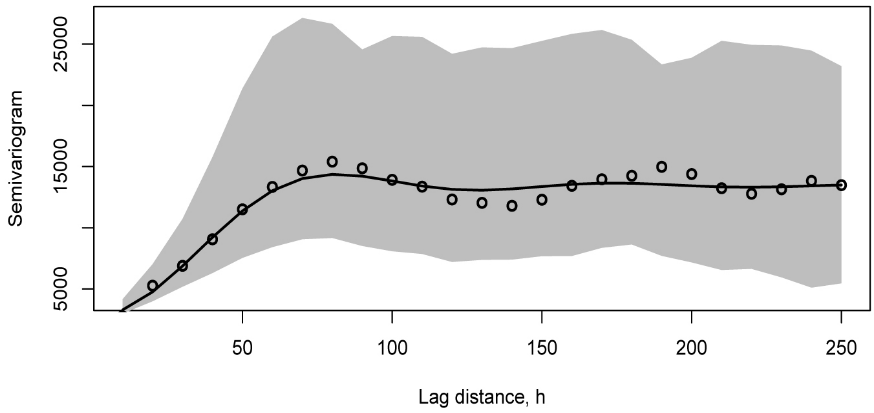

Figure 4. Comparison between the empirical semivariogram of the Swiss rainfall data (circles) and the fitted sine-cosine wave model (solid line) together with 95% simultaneous rank envelopes (shaded areas) calculated from 39 simulations of the fitted model.

proposed to be employed for nonparametric curve estimation offering new insights in providing simple methods for constructing new families of stationary spatial and spatio-temporal stationary covariance functions from the Fourier transform of these kernels. We study a class of higher-order Gaussian stationary spatial and spatio-temporal covariance functions. Further, a class of Müller's stationary covariance functions in spatial and spatio-temporal domains is constructed. Both families have very similar behavior for different orders (r). A parametric version of the model, called the sine-cosine wave model, is applied to the Swiss rainfall data, and comparison between the fitted model and the 95% simultaneous envelopes shows that the sine-cosine wave model fits the data well.

Acknowledgements

MG was financially supported by the Kempe Foundations (JCSMK22-0134) and the Olle Engkvists Foundation (224-0071). We thank the associate editor and the reviewers. Their precise comments and suggestions significantly improved an earlier version of the manuscript.

References

Brown, P.E., Karesen, K.F., Roberts, G.O. and Tonellato, S. (2000). Blur-generated non-separable space-time models, Journal of the Royal Statistical Society. *Series B (Statistical Methodology)*, **62**, 847–860.

Cherian, A., Koniusz, P. and Gould, S. (2017). Higher-order pooling of cnn features via kernel linearization for action recognition, *IEEE Winter Conference on Applications of Computer Vision*, 130–138.

Christakos, G. (2000). *Modern spatiotemporal geostatistics*, Vol. 6, Oxford university press.

Cressie, N.A.C. (1993). *Statistics for Spatial Data*, Second edn, Wiley, New York.

Cressie, N. and Huang, H.C. (1999). Classes of nanseparable, spatio-temporal stationary covariance function, *Journal of the American Statistical Association*, **94**, 1330–1340.

Das, D., Nayak, M. and Pani, S.K. (2019). Missing value imputation–a review, *Int. J. Comput. Sci. Eng.*, **7**, 548–558.

Dubois, G. (1998). Spatial interpolation comparison 97: foreword and introduction, *Journal of Geographic Information and Decision Analysis*, **2**, 1–10.

Fasshauer, G.E. (2007). *Meshfree Approximation Methods with MATLAB*, World Scientific Publishing Company.

Finkenstadt, B., Held, L. and Isham, V. (2007). *Statistical methods for spatio-temporal system*, Taylor & Francis Group, LLC.

Fuentes, M. and Smith, R. (2001). *A new class of nonstationary models*, Technical report, North Carolina State University .

Granovsky, B.L. and Muller, H.G. (1991). Optimizing kernel methods: A unifying variational principle, *International Statistical Review*, **59**, 373–388.

Hansen, B. E. (2005). Exact mean integrated squared error of higher order kernel estimators, *Econometric Theory*, **21**, 1031–1057.

Hurst, S. (1995). *The characteristic function of the student t-distribution.* Financial Mathematics Research Report 006-95, Australian National University, Canberra ACT 0200, Australia.

Jones, R.H. and Zhang, Y. (1997). *Models for continuous stationary space-time processes, Modelling longitudinal and spatially correlated data,* Springer, pp. 289–298.

Lindgren, G. (2012). *Stationary Stochastic Processes: Theory and Applications*, Chapman & Hall/CRC Press.

Ma, C. (2003). Families of spatio-temporal stationary covariance models, *Journal of Statistical Planning and Inference*, **116**, 489–501.

Marron, J. (1994). Visual understanding of higher-order kernels, *Journal of Computational and Graphical Statistics*, **3**, 447–458.

Matern, B. (1960). Spatial variation: stochastic models and their applications to problems in forest surveys and other sampling investigation, *Meddelanden fran Statens Skogforsknings Institut*, **49**.

Matheron, G. (1962). Traite de geostatistique, *Mem. BRGM* (14).

Møller, J. and Waagepetersen, R.P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC, Boca Raton.

Muller, H.G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes, *The Annals of Statistics*, **12**, 766–774.

Powell, M.J.D. (2009). The bobyqa algorithm for bound constrained optimization without derivatives. Research report NA2009/06, Department of Applied Mathematics and Theoretical Physics, Cambridge, England.

Ribeiro Jr, P.J. and Diggle, P.J. (2001). geoR: a package for geostatistical analysis, *R-NEWS*, **1**, 15–18. URL:http://cran.R-project.org/doc/Rnews

Ripley, B.D. (1981). *Spatial Statistics*, Wiley, New York.

Salvi, C., Lemercier, M., Liu, C., Horvath, B., Damoulas, T. and Lyons, T. (2021). Higher order kernel mean embeddings to capture filtrations of stochastic processes, *Advances in Neural Information Processing Systems*, **34**, 16635–16647.

Sampson, P.D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure, *Journal of the American Statistical Association*, **87**, 108–119.

Schucany, W. and Sommers, J.P. (1977). Improvement of kernel type density estimators, *Journal of the American Statistical Association*, **72**, 420–423.

Tsuruta, Y. and Sagae, M. (2017). Higher order kernel density estimation on the circle, *Statistics Probability Letters*, **131**, 46–50.

Wackernagel, H. (2003). *Multivariate Geostatistics*, Springer-Verlag, Berlin Heidelberg.
Wand, M.P. and Schucany, W.R. (1990). Gaussian-based kernels, *The Canadian Journal of Statistics*, **18**, 197–204.

Wikle, C.K., Z.M.A. and Cressie, N. (2019). *Spatio-Temporal Statistics with R*, Boca Raton, FL: Chapman Hall/CRC.

Yaglom, A.M. (1987). *Correlation Theory of Stationary and Related Random Functions. Volume I: Basic Results*, Springer-Verlag, New York Inc.

Ypma, J. (2017). nloptr: R interface to nlopt. R package version 1.0.4.

Nafiseh Vafaei
Department of Computer and Statistics Sciences,
Faculty of Sciences,
University of Mohaghegh Ardabili,
Ardabil, Iran.
email: *n.vafaei@uma.ac.ir*

Jorge Mateu
Department of Mathematics,
Jaume I University,
Castellon, Spain.
email: *mateu@uji.es*

Masoud Ganji
Department of Computer and Statistics Sciences,
Faculty of Sciences,
University of Mohaghegh Ardabili,
Ardabil, Iran.
email: *mganji@uma.ac.ir*

Mohammad Ghorbani
Department of Engineering Sciences and Mathematics,
Lulea University of Technology,
Sweden.
email: *mohammad.ghorbani@ltu.se*