

Radial basis functions with compact support for multivariate geostatistics

Emilio Porcu · Daryl J. Daley · Martin Buhmann ·
Moreno Bevilacqua

Published online: 13 November 2012
© Springer-Verlag Berlin Heidelberg 2012

Abstract Matrix-valued radially symmetric covariance functions (also called radial basis functions in the numerical analysis literature) are crucial for the analysis, inference and prediction of Gaussian vector-valued random fields. This paper provides different methodologies for the construction of matrix-valued mappings that are positive definite and compactly supported over the sphere of a d -dimensional space, of a given radius. In particular, we offer a representation based on scaled mixtures of Askey functions; we also suggest a method of construction based on B-splines. Finally, we show that the very appealing convolution arguments are indeed effective when working in one dimension, prohibitive in two and feasible, but substantially useless, when working in three dimensions. We exhibit the statistical performance of the proposed models through simulation study and then discuss the computational gains that come from our constructions when the parameters are estimated via maximum likelihood. We finally apply our constructions to a North American Pacific Northwest temperatures dataset.

Keywords Askey function · Buhmann class · Convolution · Sphere · Splines

E. Porcu (✉) · M. Bevilacqua
Department of Statistics, Universidad de Valparaiso, Valparaiso,
Chile
e-mail: porcu.emilio@yahoo.com; emilio.porcu@uv.cl

D. J. Daley
Department of Mathematics and Statistics, University
of Melbourne, Parkville, VIC, Australia

M. Buhmann
Mathematisches Institut, Justus–Liebig University, Giessen,
Germany

1 Introduction

Radial basis functions have a dual use and interpretation as confirmed by the literature of both approximation theory and statistics. In the spatial statistics field, the construction of cross-covariance models has become an important goal for the analysis and study of multi-variate random fields (RFs). The analysis of spatial and space–time data requires not only the specification of some dependence structure within the phenomenon of interest, but also the dependence between phenomena defined on the same spatial domain. For instance, it is common knowledge that environmental processes have reciprocal influences over space and time. Literature has been focused on this problem for decades and relevant applications can be found, e.g. in meteorology (Banerjee et al. 2004; Berrocal et al. 2008) and atmospheric contamination (Schmidt and Gelfand 2003), amongst others. Relevant methodologies can be found, e.g. in Banerjee and Gelfand (2003) for the mathematical properties of such processes; more recent work in the context of tapering has been done in Furrer et al. (2006), Kaufman et al. (2008) and Du et al. (2009). We refer to Li et al. (2008) concerning the possible simplification of the structure of such multivariate processes. Finally, excellent classic textbook accounts include Chiles and Delfiner (1999), Christakos (2000), Christakos and Hristopoulos (1998), Cressie (1993) and Goovaerts (1997).

Methodological approaches vary considerably, reflecting the diverse routes open to studying the problem. In this paper we focus on approaches that are based on the so-called geostatistical framework: the essence of this lies in the specification of the second-order properties of the process of interest, which translates into covariances and variograms (Cressie 1993). Another approach using radial basis functions is given in Beatson et al. (2009).

Building valid cross-covariance models is a nontrivial matter. The literature on this problem can be traced back to Cramer's (1940) seminal paper. There are very few contributions available and these are sketched below:

1. *Separable covariances* These are obtained by factorizing a positive definite function with a positive definite matrix of coefficients. The construction is easy to implement but not very interesting in terms of interpretability and flexibility, because it assumes that all components of the multivariate RF have the same covariance function.
2. *Linear model with co-regionalization* (Wackernagel 2003) Representing every component of the m -dimensional RF as a linear combination of $r < m$ mutually uncorrelated latent variables, leads to a simple model. Constructive criticism in Gneiting et al. (2010) and Wackernagel (2003) suggests that in many situations this model may be inadequate because every component of the random vector is represented as a linear combination of latent, independent univariate spatial processes. The model is not very flexible, nor does it allow us to recover the smoothness of the latent processes, since the smoothness of the components is dominated by the roughest of the latent components representing them.
3. *Kernel and covariance convolution method* These are proposed in Ver Hoef and Barry (1998) and in Gaspari and Cohn (1999), where details and several examples are offered.
4. *Kernels for vector-valued RFs via div and curl* Such kernels have been proposed in the context of numerical analysis with the purpose to adapt the covariance function to the physical characteristics of the associated RF, which could be, for instance, a velocity vector obeying some physical properties. Such properties are called, Narcowich et al. 2007, divergence- and curl-free (see also Benbourhim and Bouhamidi 2005).
5. *Multivariate Matern covariance structure* Gneiting et al.'s (2010) construction belongs to this class. The Matern class of correlation functions is well known in the geostatistical literature (e.g. p.31 of Stein 1999). We define it here by

$$\text{Mat}(x; \alpha, \nu) = (2^{1-\nu} / \Gamma(\nu)) (\alpha \|x\|)^{\nu} K_{\nu}(\alpha \|x\|), \quad \|x\| \geq 0, \quad (1)$$

$x \in \mathbb{R}^d$ and $\|\cdot\|$ denoting the Euclidean norm, where α and ν are positive parameters and $K_{\nu}(\cdot)$ is the MacDonald or modified Bessel function of order ν (see e.g. p. 373 of Whittaker and Watson 1927); α determines the scale of dependence and ν controls the smoothness and Hausdorff dimension of the associated Gaussian RF. Gneiting et al.'s
6. *Multivariate model through latent dimension.* Apanasovich and Genton (2009) extend Gneiting's (2002a) class of space-time covariance functions to higher dimensional spaces (such a construction is a special case of the class presented by Porcu et al. 2006).
7. *Vector-valued permissibility criteria* This refers to criteria such as those offered in Porcu and Zastavnyi (2011), where several sufficient conditions are given for candidate matrix-valued mappings to represent the covariance matrix function of a vector-valued RF.
8. *Spartan vector-valued RFs* These were proposed recently in Hristopoulos and Porcu (2012). They are multivariate analogues of scalar-valued constructions proposed.

The multivariate Matern model may be not so attractive to practitioners dealing with massive spatial datasets because calculation of the simple kriging predictor then involves the inversion of huge matrices. For univariate RFs, this problem has been comprehensively addressed by Gneiting's (2002a) suggestions of several methods for the straightforward construction of compactly supported correlation functions, that is, functions vanishing outside a finite range, typically the unit sphere \mathbb{S}^{d-1} of \mathbb{R}^d .

This paper contains original contributions for the construction of matrix-valued radial basis mappings having compact support. We use two main arguments: on the one hand, scale mixture arguments offer nice closed forms when working with splines and either Askey or Buhmann functions. On the other hand, we show that convolution arguments can be very effective in one dimension, prohibitive in \mathbb{R}^2 , and feasible but pointless in three or more (odd) dimensions.

Here then is an outline of the rest of the paper: Sect. 2 contains basic facts and notation for covariance functions associated to vector valued RFs, and a small review on compactly supported radial basis functions for scalar-valued RFs, which can be very useful for Sect. 3, which is split into two parts: in the former, scale mixtures are applied to the radial basis functions introduced in Sect. 2.1; in the latter, we illustrate a bridge between scale mixtures and splines. Section 4 is dedicated to convolutions, and Sect. 5 offers a simulation study. Section 6 analyzes a dataset from North American Pacific Northwest temperatures dataset. More technical arguments, the proofs of the results, and routine algebra are deferred to the Appendix.

2 Notation, literature and discussion

Throughout this paper we use notation from the theory of RFs. Our nomenclature may thus be more familiar to statisticians than numerical analysts or students of approximation theory. Specifically, we consider a zero mean Gaussian vector-valued m -dimensional RF with index set in some domain $D \subset \mathbb{R}^d$, i.e.

$$\mathbf{Z}(\xi) = (Z_1(\xi), \dots, Z_m(\xi)), \quad \xi \in D.$$

The assumption of Gaussianity implies that \mathbf{Z} is completely characterized by its cross-covariance structure \mathbf{C} defined as the mapping from $D \times D$ into matrices $M_m \in \mathbb{R}^{m \times m}$ for which

$$\mathbf{C}(\xi_1, \xi_2) = [C_{ij}(\xi_1, \xi_2)]_{i,j=1}^m = [\text{Cov}(Z_i(\xi_1), Z_j(\xi_2))]_{i,j=1}^m, \tag{2}$$

for $\xi_k \in D, k = 1, 2$. The mapping \mathbf{C} must be positive definite, i.e. for any finite collection of points $\xi_1, \dots, \xi_n \in D$ and any m -dimensional complex vectors $\mathbf{c}_1, \dots, \mathbf{c}_n$ for which \mathbf{c}_i has entries $(c_{ik})_{k=1}^m, i = 1, \dots, n$, the inequality

$$\sum_{i,j=1}^n \sum_{k,l=1}^m c_{ik} C_{kl}(\xi_i, \xi_j) \bar{c}_{jl} > 0 \tag{3}$$

holds. Under the assumption of stationarity we have $C_{ij}(\xi_1, \xi_2) =: C_{ij}(x)$, with $x := \xi_1 - \xi_2$ termed the lag, and Cramér’s (1940) theorem gives a complete characterization of \mathbf{C} : it is the Fourier-Stieltjes transform of some Borel measure F with values in $\mathbb{R}^{m \times m}$ such that the matrix $[F_{ij}(A)]_{i,j=1}^m$ is positive definite for every Borel set A of \mathbb{R}^d .

We start by fixing our notation. Let M_m denote the set of all complex $m \times m$ matrices, with entries in \mathbb{C} . The matrix $A \in M_m$ is positive definite if the inequality $z'Az \geq 0$ holds for every $z \in \mathbb{C}^m$; here z is any m -vector with \mathbb{C} -valued elements and z' the transpose of the conjugate elements of z . It is easily shown that $A \in M_m$ is positive definite if and only if it permits a Cholesky factorization $A = C'C$ for some matrix $C \in M_m$. Throughout the paper, we consider the class Φ_d^m of mappings $\varphi := [\varphi_{ij}(\cdot)]_{i,j=1}^m: [0, \infty) \rightarrow M_m$, with $\varphi_{ij}(0) < \infty$, and each $\varphi_{ij}(\cdot)$ continuous, $i, j = 1, \dots, m$, such that there exists an m -variate Gaussian RF \mathbf{Z} defined on \mathbb{R}^d such that the associated covariance \mathbf{C} in Eq. (2) is equal to

$$\mathbf{C}(\xi_1, \xi_2) = \varphi(\|\xi_2 - \xi_1\|) = [\varphi_{ij}(\|\xi_2 - \xi_1\|)]_{i,j=1}^m, \tag{4}$$

$\xi_1, \xi_2 \in \mathbb{R}^d$.

For $m = 1$, write $\Phi_d := \Phi_d^1$ for the set of all continuous functions $f: [0, \infty) \rightarrow \mathbb{R}$ with $f(0) < \infty$ such that $f(\|\cdot\|)$ is positive definite on a d -dimensional Euclidean space. Finally, $\Phi^m := \Phi_0^m$ denotes the class of positive-definite matrices.

2.1 Compactly supported radial basis functions for scalar-valued RFs

The Wendland class of correlation functions has been repeatedly used in applications involving the so-called tapered likelihood (Furrer et al. 2006; Kaufman et al. 2008; Du et al. 2009). We recall here Wendland’s (1995) construction (it is rephrased in Gneiting 2002b). Let

$$\psi_{v,0,\beta}(t) := \left(1 - \frac{t}{\beta}\right)_+^v, \tag{5}$$

$t \in [0, \infty), v \in \mathbb{R}_+$ and $\beta > 0$,

be the truncated power function, also known as the Askey function (1973) when $v \in \mathbb{N}$, although in the remainder of the paper we shall call $\psi_{v,0,\beta}$ an Askey function for any positive v for the sake of simplicity; this function belongs to the class Φ_d when $v \geq \frac{d+1}{2}$, which means that there exists a Gaussian RF $Z(\xi), \xi \in D$, such that

$$\text{Cov}(Z(\xi_1), Z(\xi_2)) = \psi_{v,0,\beta}(\|\xi_2 - \xi_1\|), \quad \xi_1, \xi_2 \in D.$$

This fact explains why the radial part of the Askey function is compactly supported over a sphere \mathbb{S}^{d-1} contained in \mathbb{R}^d and with radius $\beta > 0$. Many applications refer to the exponent $v \in \mathbb{N}$, while the real part of the exponent is a basic characteristic of the associated Sobolev space that determines the regularity properties of a Gaussian RF with such covariance structure. For pertinent results on this topic see Wendland (2005, Chap. 8) and, for the cases not treated there, Schaback’s (2009) complementary contribution.

For $x \in \mathbb{R}^d$, clearly we have that $\psi_{v,0,\beta}(\|x\|)$ is not differentiable at zero. Such inconvenience is overcome through the Wendland–Gneiting construction. For any $g \in \Phi(\mathbb{R}^d)$ for which $\int_{\mathbb{R}_+} ug(u)du < \infty$, Mathéron’s Montée operator I is defined by

$$Ig(t) = \frac{\int_t^\infty ug(u)du}{\int_0^\infty ug(u)du} \quad (t > 0).$$

Wendland (1994) defines $\psi_{v,k,\beta} := I^k \psi_{v,0,\beta}$ via k -fold iterated application of the Montée operator on the Askey function $\psi_{v,0,\beta}(x)$ defined at (5). Wendland proves that $\psi_{v,0,\beta} \in \Phi_d$ for $\psi_{v,k,\beta} \in \Phi_{d-2k}$. The implications in terms of differentiability of $\psi_{v,k,\beta}$ are well summarized by Gneiting (2002b), and Wendland (1995) shows that the degree of the piecewise polynomials is minimal for the given smoothness and dimension for which the radial basis function should be positive definite.

Buhmann (2001) proposed a generalization of the Wendland–Gneiting class, and in the literature on numerical analysis and radial basis function interpolation it is sometimes called the Buhmann class (Zastavnyi 2004). His construction is based on arguments using scale

mixtures. Given Wendland–Gneiting functions $\psi_{v+2k,k,\beta} \in \Phi_d$ for $v \geq \lfloor \frac{1}{2}d \rfloor - 2k + 1$, with $d \geq 2k + 1$ and a class of functions $g(\beta; \alpha, \lambda, \gamma) = \beta^\alpha(1 - \beta^\lambda)_+^\gamma$ compactly supported on $[0, 1]$, Buhmann’s problem corresponds to finding the ranges of the parameters α, λ, γ for which the function

$$\psi_{v,k,\alpha,\lambda,\gamma}(\|x\|^2) := \int_0^1 \psi_{v,k,\beta}(\|x\|^2)g(\beta; \alpha, \lambda, \gamma)d\beta, \tag{6}$$

which is compactly supported in the unit sphere \mathbb{S}^{d-1} of \mathbb{R}^d , is positive definite on \mathbb{R}^d . Misiewicz’s (1989) argument shows that the range of α, λ and γ depends on the dimension of the associated Euclidean space. Buhmann (2001) finds the solution for $d = 1, 2, 3$ and higher (this is indeed useful for geostatistical applications), and gives closed form solutions for some characterizations of the parameters involved in the scale mixture at (6) above. Further results in Zastavnyi (2004) determine the exact order of smoothness for any member of the Buhmann class.

According to Williamson (1956), $t \mapsto \psi_{v+2k,k,\alpha,\lambda,\gamma}(\sqrt{t})$, $t > 0$, is $v + 1$ times monotone, showing that this function belongs to the Pólya–Gneiting class (Gneiting 2001) of positive-definite functions, under the restrictions on the parameters α, λ, γ in Misiewicz (1989) (see (6) above). Moreover, $\psi_{v+2k,k,\alpha,\lambda,\gamma}(\cdot)$ is $(1 + \lfloor 2\alpha \rfloor)$ times-differentiable in all \mathbb{R}^d (this can be checked by inspection of the differentiability at 0 and 1, since within the interval (0,1) this function is infinitely differentiable).

3 Some constructions for multivariate models with compact support

This section presents the main theoretical results. All the proofs are deferred to the Appendix for the sake of a neater exposition.

3.1 Multivariate covariances based on Askey and Buhmann functions

The aim of this section is to use members of either the Askey or the Buhmann classes in such a way that, starting from members of the type $\psi_{v,0,\beta}$ and using straightforward constructions, one easily obtains mappings \mathbf{C} whose members are compactly supported and positive definite on \mathbb{R}^d .

Theorem A below is given here in a form that is useful in giving a neater exposition of our results.

Theorem A (Porcu and Zastavnyi 2011) (A): *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $D = [0, \infty)$. Assume that the family of matrix-valued functions $A(t, \omega) = [A_{ij}(t, \omega)] : D \times \Omega \rightarrow M_m$ satisfies the conditions (I) for every $i, j =$*

1, \dots, m and $t \in D$, the functions $A_{ij}(t, \cdot)$ belong to $L_1(\Omega, \mathcal{F}, \mu)$; and (II) $A(\cdot, \omega) \in \Phi_d^m$ for μ -almost every $\omega \in \Omega$. Let

$$C(t) := \int_{\Omega} A(t, \omega)\mu(d\omega) = \left[\int_{\Omega} A_{ij}(t, \omega)\mu(d\omega) \right], \quad t \in D. \tag{7}$$

Then $C \in \Phi_d^m$.

(B): *Conditions (I) and (II) are satisfied when $A(t, \omega) = f(t, \omega)F(t, \omega)$, where the maps $f(t, \omega) : D \times \Omega \rightarrow \mathbb{C}$ and $F(t, \omega) = [F_{ij}(t, \omega)] : D \times \Omega \rightarrow M_m$ satisfy the conditions*

- (i) *for every $i, j = 1, \dots, m$ and $t \in D$, the function $f(t, \cdot)F_{ij}(t, \cdot)$ belongs to $L_1(\Omega, \mathcal{F}, \mu)$*
- (ii) *$f(\cdot, \omega) \in \Phi_d$ for μ -almost every $\omega \in \Omega$; and*
- (iii) *$F(\cdot, \omega) \in \Phi_d^m$ for μ -almost every $\omega \in \Omega$, or $F(\cdot, \omega) = F(\omega) \in \Phi^m$ for μ -almost every $\omega \in \Omega$.*

Here is a direct application of Theorem A. Let the function $\varphi(\cdot; \beta) : [0, \infty)^2 \mapsto \mathbb{R}$ belong to the class Φ_d for any positive β . Let $G(\beta) \in M_m$ be a positive definite matrix of coefficients for any fixed value of β . Then the mapping $\mathbf{C} : \mathbb{R}^d \mapsto M_m$, defined by

$$\mathbf{C}(t) := \int_0^\infty \varphi(t; \beta)G(d\beta) = \left[\int_0^\infty \varphi(t; \beta)G_{ij}(d\beta) \right]_{i,j=1}^m, \tag{8}$$

is a member of the class Φ_d^m .

The result presented in Theorem 1 below entails a multivariate correlation structure obtained using linear mixtures, over β , of the Askey function $\psi_{v,0,\beta}(\cdot)$ at Eq. (5). Specifically, we propose a multivariate structure for which

$$\varphi_{ij}(t) := c_{ij}\psi_{v+\mu_{ij}}(t) \frac{\Gamma(v)\Gamma(1 + \mu_{ij})}{\Gamma(1 + v + \mu_{ij})} \quad t \in [0, \infty), \tag{9}$$

where $i, j = 1, \dots, m$, the c_{ij} are real coefficients, and $\mu_{ij} \in \mathbb{R}$. Theorem 1 gives sufficient conditions for $\mathbf{C} := [\varphi_{ij}(\|x\|)]$ to be the matrix-valued covariance of an m -variate Gaussian RF.

Theorem 1 *Let the matrix-valued mapping $\mathbf{C} : \mathbb{R}^d \mapsto M_m$ have elements $C_{ij}(x) = \varphi_{ij}(\|x\|)$, for $\varphi_{ij}(\cdot)$ as defined in Eq. (9) with $v \geq \frac{1}{2}d + 2$, and suppose that $\mu_{ii} \leq \mu_{ij} = \mu_{ji}$ for $i, j = 1, \dots, m$. If $c_{ii} \geq \sum_{j \neq i} |c_{ij}|$ and $c_{ij} = c_{ji}$, then $[\varphi_{ij}(\cdot)]_{i,j=1}^m \in \Phi_d^m$. Thus, \mathbf{C} is the matrix-valued covariance of an m -variate Gaussian RF on \mathbb{R}^d .*

The same approach can yield more general structures. For instance, the next result exhibits a multivariate structure $\mathbf{C}(x) = [C_{ij}(x)]_{i,j=1}^m$ of the Buhmann type in which the elements have the form

$$C_{ij}(x) = c_{ij}\psi_{v+2k,k,\alpha_{ij},1,\gamma_{ij}}(\|x\|), \quad x \in \mathbb{R}^d, \quad i, j = 1, \dots, m, \tag{10}$$

where $\psi_{v,k,\alpha,\lambda,\gamma}$ is the Buhmann scale mixture defined in Eq. (6). We give sufficient conditions below for such $\mathbf{C}(x)$ to be the matrix-valued covariance of an m -variate Gaussian RF. The proof uses the same arguments that establish Theorem 1 and is left to the reader.

Theorem 2 *Suppose that $\alpha_{ii} \leq \alpha_{ij} = \alpha_{ji}$ and $\gamma_{ii} \leq \gamma_{ij} = \gamma_{ji}$ for $i, j = 1, \dots, m$. If $c_{ii} \geq \sum_{j \neq i} |c_{ij}|$, then the multivariate structure \mathbf{C} defined by (10) is the matrix-valued covariance of an m -variate Gaussian RF.*

By way of example, a change of variable and computation shows that the choice $g(\cdot; \alpha, 1, \gamma)$ gives the class

$$\begin{aligned} & \frac{\Gamma(1 + \alpha)\Gamma(1 + \gamma)}{\Gamma(2 + \alpha + \gamma)} {}_2F_1(-1 - \alpha - \gamma, -v; -\alpha; \|x\|) \\ & + \|x\|^{1+\alpha} \frac{\Gamma(-1 - \alpha)\Gamma(1 + v)}{\Gamma(-\alpha + v)} {}_2 \\ & F_1(-\gamma, 1 + \alpha - v; 2 + \alpha; \|x\|), \end{aligned}$$

where ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$ is the Gauss hypergeometric function (Eq. (15.1.1) of Abramowitz and Stegun 1964). The special case $\alpha = \gamma$ gives

$$\begin{aligned} & \frac{\sqrt{\pi}\Gamma(1 + \alpha)}{2^{1+2\alpha}\Gamma(\frac{3}{2} + \alpha)} {}_2F_1(-1 - 2\alpha, -v; -\alpha; \|x\|) \\ & + \|x\|^{1+\alpha} \frac{\Gamma(-1 - \alpha)\Gamma(1 + v)}{\Gamma(-\alpha + v)} {}_2 \\ & F_1(-\alpha, 1 + \alpha - v; 2 + \alpha; \|x\|) \end{aligned}$$

The result of Theorem 2 is applicable to both expressions above. A relevant comment is that the structures are not in general smooth at the origin whereas Buhmann functions are. This difference arises from the mixture integral which in (15) has argument $\|x\|$ which Buhmann can (but we cannot) replace by $\|x\|^2$.

3.1.1 Bivariate cases, parsimonious versions and convenient parametrizations

This section discusses a bivariate Gaussian RF that is used throughout the data analysis and simulation study in Sect. 5. From now on we write ψ_v to denote the particular Askey function $\psi_{v,0,1}$; this function is compactly supported on the interval $[0, 1]$.

Proposition 3 *Let $\sigma_i > 0$ for $i = 1, 2$. A sufficient condition for the covariance function*

$$\begin{pmatrix} \sigma_1^2 \psi_{v+\mu_{11}}(\|x\|) & \rho_{12} \sigma_1 \sigma_2 \psi_{v+\mu_{12}}(\|x\|) \\ \rho_{12} \sigma_1 \sigma_2 \psi_{v+\mu_{12}}(\|x\|) & \sigma_2^2 \psi_{v+\mu_{22}}(\|x\|) \end{pmatrix} \tag{11}$$

to be the matrix-valued covariance of a bivariate Gaussian RF on \mathbb{R}^d when $\mu_{12} \leq \frac{1}{2}(\mu_{11} + \mu_{22})$ and $v \geq \lfloor \frac{1}{2}d \rfloor + 2$, is that

$$|\rho_{12}| \leq \frac{\Gamma(1 + \mu_{12})}{\Gamma(1 + v + \mu_{12})} \sqrt{\frac{\Gamma(1 + v + \mu_{11})\Gamma(1 + v + \mu_{22})}{\Gamma(1 + \mu_{11})\Gamma(1 + \mu_{22})}}. \tag{12}$$

For the sake of clarity, we call the model in Proposition 3 a *full* bivariate Askey model, in order to distinguish it from the *parsimonious* Askey model which is obtained by setting $\mu_{12} = \frac{1}{2}(\mu_{11} + \mu_{22})$.

As a second example consider a bivariate RF; we shall obtain sharper conditions from the same representation. It is sufficient that the matrix

$$\begin{pmatrix} c_{11}\beta^{\alpha_{11}}(1 - \beta)^{\gamma_{11}} & c_{12}\beta^{\alpha_{12}}(1 - \beta)^{\gamma_{12}} \\ c_{12}\beta^{\alpha_{12}}(1 - \beta)^{\gamma_{12}} & c_{22}\beta^{\alpha_{22}}(1 - \beta)^{\gamma_{22}} \end{pmatrix}$$

be positive definite for all $\beta \in (0, 1)$. This condition is satisfied when

$$2 \log c_{12} - \log(c_{11}c_{22}) \leq \inf_{0 < \beta < 1} (\theta_1 \log \beta + \theta_2 \log(1 - \beta)),$$

where $\theta_1 = \alpha_{11} + \alpha_{22} - 2\alpha_{12}$ and $\theta_2 = \gamma_{11} + \gamma_{22} - 2\gamma_{12}$, provided that both θ_1 and $\theta_2 < 0$; the sufficient condition for positive definiteness is that $c_{12}^2 / (c_{11}c_{22}) \leq |\theta_1|^{\theta_1} |\theta_2|^{\theta_2} / |\theta_1 + \theta_2|^{\theta_1 + \theta_2}$. This yields a new sufficient condition for a bivariate model to be positive definite on \mathbb{R}^d whenever $v + 2k \geq \frac{1}{2}d + 1$.

3.2 Scale mixtures of one-dimensional splines or B-splines

Further parameter-dependent covariance functions can be created by using univariate splines (de Boor 1981). These are piecewise polynomial functions of degree $k - 1$ (therefore of order k), say, that is, for a given sequence of so-called knots, the functions are polynomials of that degree between each pair of adjacent knots, and they are also required to be continuously differentiable, of one order less, here $k - 2$. Suitable bases of these linear spaces are, for instance, truncated power functions or the celebrated B-splines which enjoy the advantage of minimal compact support. This is especially interesting in our present context. Moreover, they can be made β -dependent by choosing the knots of the B-splines to be multiples of β , where β is the parameter over which mixing takes place.

Our idea is based on the fact that univariate B-splines give rise to totally positive interpolation matrices if the interpolation points are suitably chosen. Notice here that these interpolation points are not necessarily the same as

the knots which define the splines and the B-splines. We make both of them depend on β .

For this purpose, let $B = B_0$ be a univariate, piecewise polynomial B-spline with knots $0, \beta, 2\beta, \dots, k\beta$ and which has piecewise degree $k - 1$ and support in $[0, k\beta]$. Shifts of these B-splines are to be used to construct the functions which are to be mixed.

First, we need a representation for the B-splines. To this end, using truncated power functions, as is common for specifying piecewise polynomial splines, B_0 can be written as

$$B_0(x) = \sum_{\ell=0}^k d_\ell (x - \ell\beta)_+^{k-1},$$

where the coefficients d_ℓ are given by

$$d_\ell = \prod_{\substack{i=0 \\ i \neq \ell}}^k \frac{1}{(i - \ell)\beta} = \frac{(-1)^\ell \beta^{-k}}{\ell!(k - \ell)!}.$$

Alternatively, we may express B-splines as divided differences of truncated powers:

$$[0, \beta, 2\beta, \dots, k\beta](\cdot - x)_+^{k-1}.$$

An interpolation process based on these B-splines using interpolation points $x_j, j = 0, 1, 2, \dots, k + K - 1$, gives rise to a totally positive interpolation matrix as soon as each of the interpolation points is inside the support of one of the corresponding B-splines (this is the celebrated Schoenberg-Whitney interpolation theorem: see e.g. p. 272 of Powell 1981).

The integer K comes from the dimension of the spline-space spanned by the B-splines

$$B_{-k+1}, B_{-k+2}, \dots, B_0, B_1, \dots, B_K,$$

that are restricted to the interval $[0, (K + 1)\beta]$; the dimension of the spline space is $k + K$. To this end, we let $B_j := B_0(\cdot - j\beta)$. It is then notationally convenient to let the interpolation knots just described be strictly ascending with respect to the index and to lie in the interval $[0, (K + 1)\beta]$, and let x_i be in the support $[(i - k + 1)\beta, (i + 1)\beta]$ of B_{i-k+1} intersected with the interval $[0, (K + 1)\beta]$. Then, for instance, suitable knots can be $x_0 = 0, x_1 = \beta, \dots, x_{k+K-1} = (k + K - 1)\beta$, or $x_0 = 0, x_1 = \frac{1}{2}\beta, \dots, x_{k+K-1} = \frac{1}{2}(k + K - 1)\beta$, so long as $k + K - 1 \leq 2K$. The resulting $(k + K) \times (k + K)$ interpolating matrix is

$$A = \left[B_m(x_j) \right]_{\substack{0 \leq j < k+K \\ -k < m \leq K}}$$

Note that this matrix has elements which depend on β . To ensure that it is a symmetric and positive-definite matrix, consider $\Sigma(\beta) := A'A$ which is certainly symmetric, and then positive definite because A is totally positive.

For example, when the knots are as in the first of the two cases just mentioned, the matrix A has elements

$$\begin{aligned} a_{mn} &= \sum_{\ell=0}^k (m\beta - n\beta - \ell\beta)_+^{k-1} \prod_{\substack{i=0 \\ i \neq \ell}}^k \frac{1}{(i - \ell)\beta} \\ &= \sum_{\ell=0}^k \frac{(-1)^\ell (m - n - \ell)_+^{k-1}}{\ell!(k - \ell)! \beta}, \end{aligned}$$

$0 \leq m < k + K, -k < n \leq K$. Symmetrization of such A yields for substitution into Eq. (7) the functions $G_{ij}(\beta) = \sum_m a_{mi} a_{mj}$, namely

$$\begin{aligned} G_{ij}(\beta) &= \frac{1}{\beta^2} \sum_{m=0}^{k+K-1} \sum_{\ell=0}^k \sum_{\ell'=0}^k \frac{(-1)^{\ell+\ell'} (m - i - \ell)_+^{k-1} (m - j - \ell')_+^{k-1}}{\ell!(k - \ell)! \ell'!(k - \ell')!}, \end{aligned}$$

$i, j = -k + 1, \dots, K$, where for given i and j the inner summations occur over non-zero elements only for $0 \leq \ell \leq \min(k, m - i - 1)$ and $0 \leq \ell' \leq \min(k, m - j - 1)$.

Of course, more complicated formulae may be taken for the choice of the interpolation coefficients. These will give rise to different sets of covariance functions C .

4 Convolved functions for multivariate data

More general functions can be used via constructions based on convolutions of Askey functions. The key idea is contained in the Convolution Theorem below. It paraphrases statements in Gneiting et al. (2010) as a variant of other reformulations quoted there, namely, that the corresponding multivariate Gaussian RF allows a representation as a process convolution, with distinct kernel functions C_1, \dots, C_m relative to a common white noise process.

Theorem 4 [Convolution Theorem] *Let $C_1, \dots, C_m : \mathbb{R}^d \mapsto L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, and set*

$$C(x) := \left[C_{ij}(x) := C_i * C_j(x) \right]_{i,j=1}^m.$$

Then C is the matrix-valued covariance of an m -variate Gaussian RF.

We start by offering a result in the one-dimensional case ($d = 1$), for which somewhat simpler expressions are available though of lesser interest for geostatistical applications.

Proposition 5 *For $d = 1$, integer-valued v_i and $i = 1, \dots, m, x \in \mathbb{R}$, let $C_i(x) := \psi_{v_i}(|x|)$ be covariance functions of the Askey type. Define $C_{ij}(x) := C_i * C_j(x)$ so $C_{ij}(x)$ equals*

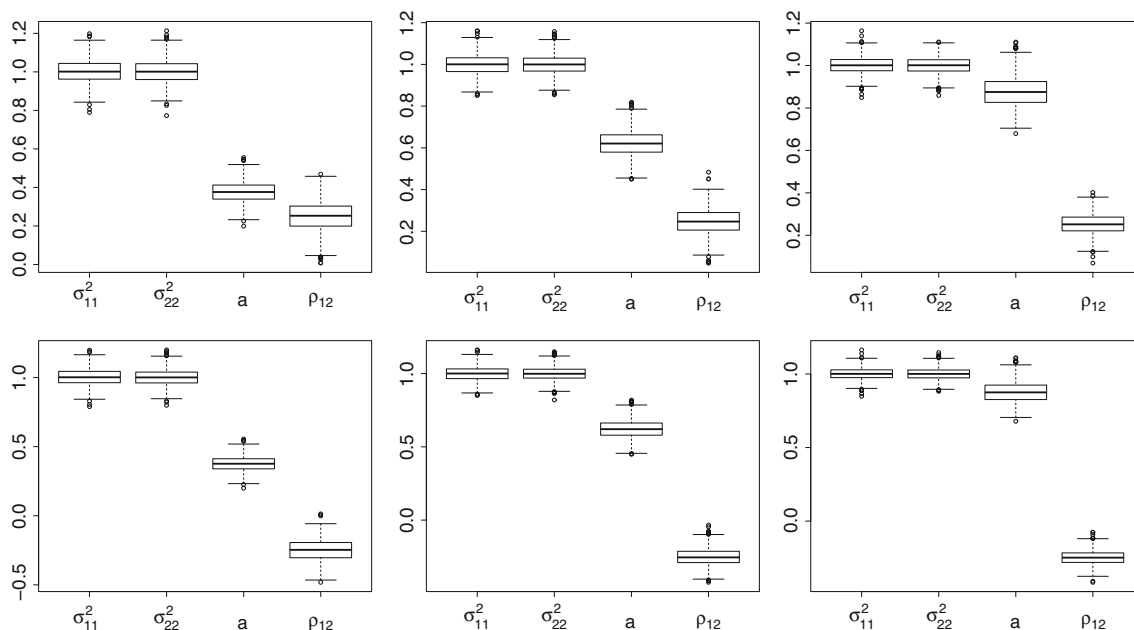


Fig. 1 Boxplots for the simulation study reported according to the described scenarios. The true parameters are $\sigma_{11} = \sigma_{22} = 1$; $a = \frac{\Delta}{8}$; $\sigma_{12} = 0.25$ (first row) and $\sigma_{12} = -0.25$ (second row)

$$\frac{v_i!v_j!}{(v_i + v_j + 1)!} (2 - |x|)^{v_i+v_j+1}, \quad \text{for } 1 \leq |x| \leq 2, \quad (13)$$

$$\begin{aligned} & \frac{v_i!v_j!(-1)^{v_i+1}|x|^{v_j+v_i+1}}{(v_j + v_i + 1)!} + 2 \sum_{k=0}^{\lfloor \frac{1}{2}v_i \rfloor} \frac{v_i!v_j!(1 - |x|)^{v_i-2k}}{(v_i - 2k)!(v_j + 1 + 2k)!} \\ & - 2 \sum_{k=0}^{\lfloor \frac{1}{2}(v_i-1) \rfloor} \frac{v_i!v_j!(1 - |x|)^{v_i+2+2k}}{(v_i - 1 - 2k)!(v_j + 2 + 2k)!}, \quad \text{for } |x| \leq 1. \end{aligned} \quad (14)$$

Then for $\min_i\{v_i\} \geq 2$, $\mathbf{C}(x) := [C_{ij}(x)]_{i,j=1,\dots,m}$ is the covariance of an m -variate Gaussian process on \mathbb{R} .

The derivation of (13) and (14) is given in the Appendix. The constraint involving $\min_i\{v_i\}$ comes from the condition on Askey functions below (5) for the case $d = 1$.

In the second part of the Appendix we exhibit what is possible concerning explicit calculation of convolutions in \mathbb{R}^3 in the isotropic case. These computations are barely tractable, and seem somewhat pointless; Table 2 shows some special cases of the expressions we obtain.

5 Simulation study

We report here results from a simulation study we made with the goal of exploring the performance of the proposed model from the point of view of both statistical and computational efficiency.

Within a quasi-standard setting of increasing domain asymptotics we simulated 1,000 realizations of a Gaussian

RF with the parsimonious bivariate covariance coming from the Askey model. We considered three possible scenarios of 50Δ location sites uniformly distributed over the grid $[-\frac{1}{2}\Delta, \frac{1}{2}\Delta]^2$, for $\Delta = 3, 5, 7$ (i.e. 150, 250, 350 points respectively). We set $\sigma_1^2 = \sigma_2^2 = 1, \rho_{12} = 0.25$ and since we are working under increasing domain asymptotics, we also make a commensurate increase in the support of the Askey functions by setting it equal to $\frac{1}{8}\Delta$. The smoothing parameters are fixed at $\mu_{ij} = 0.5$ for $i, j = 1, 2$.

In Fig. 1 (first row) we report the boxplots for the parameters for the increasing scenarios. From these boxplots it is evident that the variance of the estimates, as expected, decreases as the number of location sites increases. On the other hand the variance associated with the compact support parameter increases slightly: this is not surprising since the strength of correlation increases with the number of location sites.

For the sake of completeness, we repeated this simulation study with the same parameter settings except that we changed the co-location correlation coefficient whose nominal value was now set to -0.25 . The results are reported in the second row of Fig. 1.

Concerning our computations, we used algorithms for sparse matrices in computing likelihoods. We used the R statistical computing environment (R Development Core Team (2007)), using the SPAM package (Furrer and Sain (2010)), coupled with C routines.

Figure 2 shows the computational time when evaluating the likelihood function with or without the use of algorithms for sparse matrices when $\Delta = 3, 5, \dots, 69$, in order

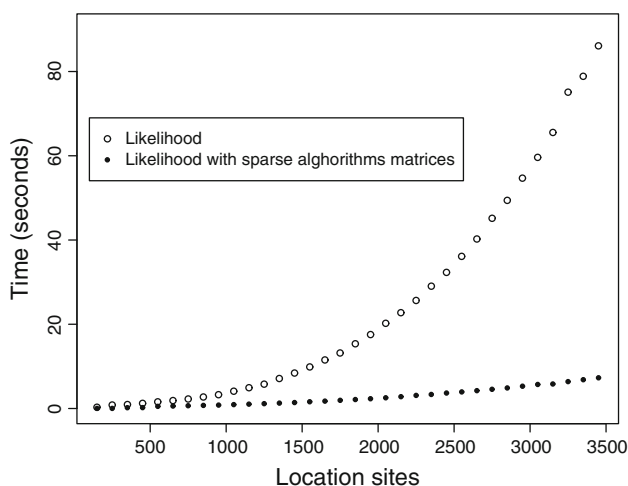


Fig. 2 Time (in seconds) for evaluating the likelihood under the bivariate Askey model with or without sparse algorithm matrices

to get 3,450 points (i.e 6,900 observations). The gain in computational efficiency when using algorithms for sparse matrices is readily apparent.

6 Data analysis

In this section, we use the dataset in Gneiting et al. (2010) so as to compare the performance of the multivariate Matérn model they used with our compactly supported model. We expect this compactly supported model to fit worse than the multivariate Matérn because with compact support there is an obvious loss of information with respect to any model with non-compact support. In our case, the motivation for compact support is the computational gain, as explained in the previous section via simulation.

The data used in Gneiting et al. (2010) consists of temperature and pressure observations and forecasts made at the 157 locations in the North American Pacific Northwest as indicated in Fig. 3 of their paper.

The forecasts are from the GFS member of the University of Washington regional numerical weather prediction ensemble (Eckel and Mass 2005); they were valid on December 18, 2003 at 4 pm local time, with a forecast horizon of 48 h. Gneiting et al. (2010) argue convincingly about the zero mean and the smoothness of this bivariate process. An important remark in their exposition is the frequently noted fact that temperature and pressure are strongly negatively correlated: the co-located empirical correlation coefficient for this dataset is -0.47 . We follow their recommendations and work within the framework of the bivariate Gaussian weakly stationary and zero mean RF $\mathbf{Z}(\xi) = (Z_P(\xi), Z_T(\xi))'$, where the subscripts P and T are used to denote temperature and pressure respectively, and where $\xi \in \mathbb{R}^2$, which is important to keep in mind because

the permissibility of the model in Eq. (9) is obviously related to the dimension of the Euclidean space where the bivariate process is defined.

We discuss here the full version of our model as in Eq. (11) for $x \in \mathbb{R}^2$, namely

$$C_{PP}(x) = \sigma_P^2 \left(1 - \frac{\|x\|}{a}\right)_+^{3+\mu_P} + \tau_P^2 \mathbb{I}(x=0),$$

$$C_{TT}(x) = \sigma_T^2 \left(1 - \frac{\|x\|}{a}\right)_+^{3+\mu_T} + \tau_T^2 \mathbb{I}(x=0),$$

$$C_{PT}(x) = \rho_{PT} \sigma_P \sigma_T \left(1 - \frac{\|x\|}{a}\right)_+^{3+\mu_{PT}},$$

obtained in the special case $\nu = 3$, which is sufficient for this mapping to belong to the class Φ_2^2 by virtue of Proposition 3. We remark that there is no reason to propose a model that has different radii a_P and a_T because it is known from Gneiting et al. (2010) that for this dataset the gain in terms of the likelihood is negligible.

We proceed to compare this full version of our model with the parsimonious and full Matérn models which, as Gneiting et al. showed, outperform the linear model of co-regionalization (LMC). They used the formulae below for the parsimonious model.

$$C_{PP}(x) = \sigma_P^2 \text{Mat}(x, \mu_P, a) + \tau_P^2 \mathbb{I}(x=0),$$

$$C_{TT}(x) = \sigma_T^2 \text{Mat}(x, \mu_T, a) + \tau_T^2 \mathbb{I}(x=0),$$

$$C_{PT}(x) = \rho_{PT} \sigma_P \sigma_T \text{Mat}\left(x, \frac{1}{2}(\mu_P + \mu_T), a\right),$$

and similar expressions for the full Matérn model. The estimates we find are consistent with those in Gneiting et al.'s work and are reported in Table 1. For this dataset, the log-likelihood under the bivariate Askey model is -1266.47 , which is slightly worse than those obtained under the bivariate parsimonious or full Matérn, being respectively equal to $-1,265.76$ and $-1,265.53$ as Gneiting et al. reported. At the same time, using the AIC criterion we have that the order of preference would be parsimonious Matérn (the best), full Askey, and full Matérn. In Fig. 3 we report the maximum likelihood fitting of the empirical covariance and cross-covariance functions for the pressure and temperature, under both the parsimonious Askey and Matérn models.

Acknowledgements D.J. Daley's work was done partly as an Honorary Professorial Associate in the School of Mathematics and Statistics at the University of Melbourne, and partly while visiting the University of Göttingen. Support both in kind and towards living away from home is gratefully acknowledged. The authors thank Professor Victor Léiva for useful discussions during the preparation of this manuscript.

Appendix

Some proofs

Proof of Theorem 1. Direct inspection shows that

Fig. 3 Empirical covariance and cross-covariance functions for the pressure and temperature data, with maximum likelihood fitting under the bivariate Askey (black line) and the bivariate Matérn (red line)

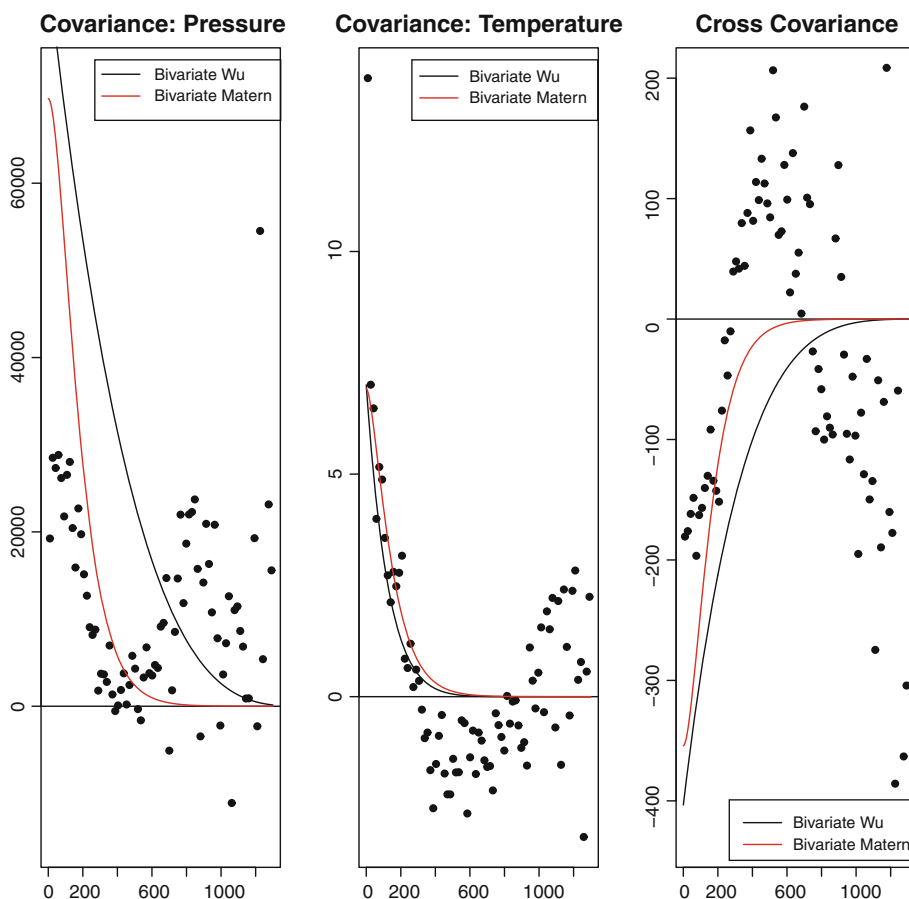


Table 1 Estimates for the bivariate Askey and the full and parsimonious bivariate Matérn models

Model	$\hat{\sigma}_P$	$\hat{\sigma}_T$	$\hat{\rho}_{PT}$	\hat{a}_P	\hat{a}_T	\hat{a}_{PT}	$\hat{\mu}_P$	$\hat{\mu}_T$	$\hat{\mu}_{PT}$	$\hat{\tau}_P$	$\hat{\tau}_T$
Bivariate Askey	290.08	2.64	-0.52	1499.17	1499.17	1499.17	2.16	10.79	6.695	54.02	0
Full Matérn	261.50	2.67	-0.54	99.00	98.40	82.20	1.50	0.59	1.41	68.4	0
Parsimonious Matérn	264.20	2.63	-0.51	92.30	92.30	92.30	1.67	0.60	1.14	70.1	0

$$\psi_{v+\mu}(t) := \frac{\Gamma(1+v+\mu)}{\Gamma(v)\Gamma(1+\mu)} \int_0^1 \psi_{v-1,0,\beta}(t) g(\beta; \mu) d\beta, \quad (15)$$

$t \in [0, \infty)$,

for $\psi_{v-1,0,\beta}(\cdot) \in \Phi(\mathbb{R}^d)$, $v \geq \frac{1}{2}d + 2$ and $\beta \mapsto g(\beta; \mu) := \beta^v(1-\beta)_+^\mu$, $\mu \geq 0$. This is evidently of the form in Eq. (8), being a special case of Theorem A, so that we only need to show that the matrix-valued mapping

$$g(\beta; \mu) := [g(\beta; \mu_{ij})]_{i,j=1}^m$$

belongs to the class Φ^m for all $0 \leq \beta \leq 1$ and $\mu \in \mathbb{R}_+^{m(m+1)/2}$. To do this we appeal to the stronger statement of diagonal dominance and nonnegativity of the elements on the diagonal, which follows from noting that

$$c_{ii}g(\beta; \mu_{ii}) \geq \sum_{j \neq i} |c_{ij}|g(\beta; \mu_{ii}) \geq \sum_{j \neq i} |c_{ij}|g(\beta; \mu_{ij}).$$

Because positive definiteness is preserved under scale mixtures and under Schur products with the positive-definite matrix \mathbf{c} of coefficients c_{ij} used in Eq. (9), the proof is complete.

Proof of Proposition 3. The proof is again constructive. The structure in Eq. (11) is obtained through the Schur product of the matrix based on (9), namely

$$\begin{pmatrix} \sigma_1^2 \frac{\Gamma(2+v+\mu_{11})}{\Gamma(1+v)\Gamma(1+\mu_{11})} & \rho_{12}\sigma_1\sigma_2 \frac{\Gamma(2+v+\mu_{12})}{\Gamma(1+v)\Gamma(1+\mu_{12})} \\ \rho_{12}\sigma_1\sigma_2 \frac{\Gamma(2+v+\mu_{12})}{\Gamma(1+v)\Gamma(1+\mu_{12})} & \sigma_2^2 \frac{\Gamma(2+v+\mu_{22})}{\Gamma(1+v)\Gamma(1+\mu_{22})} \end{pmatrix},$$

with the matrix $\mathbf{C}(\cdot)$ defined in Eq. (9) for $m = 2$, so that we only need to verify that the matrix above is positive

definite; this is indeed the case when the condition in Eq. (12) holds. The condition $\mu_{12} \leq \frac{1}{2}(\mu_{11} + \mu_{22})$ comes from the matrix $C(\cdot)$ in Eq. (11) for the case $m = 2$ via a determinantal inequality.

Proof of Proposition 5. We detail some algebra involving the convolution $(\psi_\nu * \psi_\mu)(t)$, $t \in \mathbb{R}$, of the basic Askey functions $\psi_\nu(t) := (1 - |t|)_+^\nu$ under the restriction $\nu, \mu \in \mathbb{N}$, namely

$$(\psi_\nu * \psi_\mu)(t) := \int_{-1}^1 (1 - |t - u|)_+^\nu (1 - |u|)_+^\mu du. \tag{16}$$

As written in (16) the functions are defined on the interval $[-1, 1]$ in \mathbb{R}^1 ; the essential features are that they are nonnegative, are positive definite for ν, μ greater or equal than two, and have compact support. The convolution at (16) also has compact support, albeit on $[-2, 2]$.

We evaluate (16) for positive integers ν and μ . Observe that the function on the left-hand side is symmetric about the origin so that it is a function of $|t|$. Indeed, inspection of the right-hand side shows that the first factor of the integrand is exactly $(1 - t + u)_+^\nu$ for $1 < t < 2$, and in this range it is nonzero only for $t - 1 < u < 1$. So for such t (and, indeed, for $|t| = t$) the integral equals

$$\begin{aligned} \int_{t-1}^1 (1 - t + u)^\nu (1 - u)^\mu du &= \int_0^{2-t} \nu^\nu (2 - t - \nu)^\mu d\nu \\ &= \int_0^{2-t} \frac{\nu}{\mu + 1} \nu^{\nu-1} (2 - t - \nu)^{\mu+1} d\nu \\ &= \int_0^{2-t} \frac{\nu! \mu!}{(\mu + \nu)!} (2 - t - \nu)^{\mu+\nu} d\nu \\ &= \frac{\nu! \mu!}{(\nu + \mu + 1)!} (2 - t)^{\nu+\mu+1}. \end{aligned} \tag{17}$$

For $0 < t < 1$, the set of values of u making positive contributions to the convolution expands to $-(1 - t) < u < 1$, while

$$(1 - |t - u|)_+ = \begin{cases} 1 - t + u & \text{for } -(1 - t) < u < 0, \\ 1 - t + u & \text{for } 0 < u < t, \\ 1 + t - u & \text{for } t < u < 1. \end{cases}$$

Writing the convolution integral at (16) for this range of u as $\int_{-(1-t)}^1 \dots = (\int_{-(1-t)}^0 + \int_0^t + \int_t^1) \dots$, we evaluate each of these contributions as below:

$$\begin{aligned} &\int_t^1 (1 + t - u)^\nu (1 - u)^\mu du \\ &= \left. \frac{-(1 + t - u)^{\nu+1} (1 - u)^{\mu+1}}{\mu + 1} \right|_t^1 \\ &\quad - \int_t^1 \frac{\nu(1 + t - u)^{\nu-1} (1 - u)^{\mu+1}}{\mu + 1} du \\ &= \frac{(1 - t)^{\mu+1}}{\mu + 1} - \frac{\nu(1 - t)^{\mu+2}}{(\mu + 1)(\mu + 2)} \\ &\quad + \int_t^1 \frac{\nu(\nu - 1)(1 + t - u)^{\nu-2} (1 - u)^{\mu+2}}{(\mu + 1)(\mu + 2)} d\nu \\ &= \sum_{j=0}^{\nu-1} \frac{\nu! \mu! (-1)^j (1 - t)^{\mu+1+j}}{(\nu - j)! (\mu + 1 + j)!} \\ &\quad + (-1)^\nu \int_t^1 \frac{\nu! \mu!}{(\mu + \nu)!} (1 - u)^{\mu+\nu} d\nu \\ &= \sum_{j=0}^{\nu} \frac{\nu! \mu! (-1)^j (1 - t)^{\mu+1+j}}{(\nu - j)! (\mu + 1 + j)!}; \end{aligned} \tag{18}$$

$$\begin{aligned} &\int_0^t (1 - t + u)^\nu (1 - u)^\mu du \\ &= \left. \frac{-(1 - t + u)^\nu (1 - u)^{\mu+1}}{\mu + 1} \right|_0^t \\ &\quad + \int_0^t \frac{\nu(1 - t + u)^{\nu-1} (1 - u)^{\mu+1}}{\mu + 1} du \\ &= \sum_{j=0}^{\nu} \frac{\nu! \mu! [(1 - t)^{\nu-j} - (1 - t)^{\mu+1+j}]}{(\nu - j)! (\mu + 1 + j)!}; \\ &\quad \int_{-(1-t)}^0 (1 - t + u)^\nu (1 + u)^\mu du \\ &= \left. \frac{(1 - t + u)^\nu (1 + u)^{\mu+1}}{\mu + 1} \right|_{-(1-t)}^0 \\ &\quad - \int_{-(1-t)}^0 \frac{\nu(1 - t + u)^{\nu-1} (1 + u)^{\mu+1}}{\mu + 1} du \\ &= \sum_{j=0}^{\nu-1} \frac{\nu! \mu! (-1)^j (1 - t)^{\nu-j}}{(\nu - j)! (\mu + 1 + j)!} + (-1)^\nu \frac{\nu! \mu!}{(\mu + \nu)!} \int_{-(1-t)}^0 (1 + u)^{\mu+\nu} du \\ &= \sum_{j=0}^{\nu} \frac{\nu! \mu! (-1)^j (1 - t)^{\nu-j}}{(\nu - j)! (\mu + 1 + j)!} + (-1)^{\nu+1} \frac{\nu! \mu! \mu^{\mu+\nu+1}}{(\mu + \nu + 1)!}. \end{aligned} \tag{19}$$

$$\begin{aligned} &\int_{-(1-t)}^0 (1 - t + u)^\nu (1 + u)^\mu du \\ &= \left. \frac{(1 - t + u)^\nu (1 + u)^{\mu+1}}{\mu + 1} \right|_{-(1-t)}^0 \\ &\quad - \int_{-(1-t)}^0 \frac{\nu(1 - t + u)^{\nu-1} (1 + u)^{\mu+1}}{\mu + 1} du \\ &= \sum_{j=0}^{\nu-1} \frac{\nu! \mu! (-1)^j (1 - t)^{\nu-j}}{(\nu - j)! (\mu + 1 + j)!} + (-1)^\nu \frac{\nu! \mu!}{(\mu + \nu)!} \int_{-(1-t)}^0 (1 + u)^{\mu+\nu} du \\ &= \sum_{j=0}^{\nu} \frac{\nu! \mu! (-1)^j (1 - t)^{\nu-j}}{(\nu - j)! (\mu + 1 + j)!} + (-1)^{\nu+1} \frac{\nu! \mu! \mu^{\mu+\nu+1}}{(\mu + \nu + 1)!}. \end{aligned} \tag{20}$$

Putting together (18)–(20) gives for this integration, when $0 < t < 1$,

$$2v!\mu! \left(\sum_{j=0}^{\lfloor \frac{v}{2} \rfloor} \frac{(1-t)^{v-2j}}{(v-2j)!(\mu+1+2j)!} - \sum_{j=0}^{\lfloor \frac{v-1}{2} \rfloor} \frac{(1-t)^{\mu+2+2j}}{(v-1-2j)!(\mu+2+2j)!} + \frac{\frac{1}{2}(-1)^{v+1}t^{\mu+v+1}}{(\mu+v+1)!} \right).$$

Finally then, $(\psi_v * \psi_\mu)(t)$ is zero except when $|t| < 2$ where it is as given at (13) and (14) of Proposition 5 for $v_i = v, v_j = \mu$ and $|t| = |x|$.

Derivation of convolution formulae in \mathbb{R}^3 .

The argument sketched below follows the lines of Theorem 3.c.1 in Gaspari and Cohn (1999), and shows that the convolution $(C_i * C_j)(z)$ of two radial functions compactly supported in the unit ball of \mathbb{R}^3 , when their centres are distance z apart, is expressible

$$z := x \mapsto C_{ij}(z) = \begin{cases} \frac{2\pi}{z} \int_0^1 r C_i(r) \int_{|r-z|}^{r+z} s C_j(s) ds dr, & z > 0, \\ \int_0^1 C_i(r) C_j(r) 4\pi r^2 dr, & z = 0. \end{cases} \tag{21}$$

Appealing to the formula at (21), we calculate the convolution $(\psi_v * \psi_\mu)(z)$ of two Wendland–Gneiting functions as

$$\frac{2\pi}{z} \int_0^1 r(1-r)_+^v \int_{|r-z|}^{r+z} s(1-s)_+^\mu ds dr. \tag{22}$$

The inner integrand at (22) equals

$$\int_{\max(r-z, z-r)}^{\min(r+z, 1)} s(1-s)^\mu ds,$$

where the upper limit replaces the truncation in the integrand. For the case $r > z$ this equals

$$\int_{r-z}^{\min(1, r+z)} s(1-s)^\mu ds = \left(\frac{-s(1-s)^{\mu+1}}{\mu+1} - \frac{(1-s)^{\mu+2}}{(\mu+1)(\mu+2)} \right) \Big|_{r-z}^{\min(1, r+z)}$$

after integration by parts, and depending on $r+z >$ or ≤ 1 , this equals

$$\begin{aligned} & \frac{(r-z)(1-r+z)^{\mu+1}}{\mu+1} + \frac{(1-r+z)^{\mu+2}}{(\mu+1)(\mu+2)}, & \text{if } r+z > 1, \\ & \frac{-(r+z)(1-r-z)^{\mu+1} + (r-z)(1-r+z)^{\mu+1}}{\mu+1} + \frac{-(1-r-z)^{\mu+2} + (1-r+z)^{\mu+2}}{(\mu+1)(\mu+2)}, & \text{if } r+z \leq 1. \end{aligned}$$

The case $z > r$ equals

$$\begin{aligned} & \int_{z-r}^{\min(1, r+z)} s(1-s)^\mu ds \\ & = \left(\frac{-s(1-s)^{\mu+1}}{\mu+1} - \frac{(1-s)^{\mu+2}}{(\mu+1)(\mu+2)} \right) \Big|_{z-r}^{\min(1, r+z)}, \\ & = \frac{(z-r)(1-z+r)^{\mu+1}}{\mu+1} + \frac{(1-z+r)^{\mu+2}}{(\mu+1)(\mu+2)}, \quad \text{if } r+z > 1, \\ & = \frac{-(r+z)(1-r-z)^{\mu+1} + (z-r)(1-z+r)^{\mu+1}}{\mu+1} \\ & \quad + \frac{-(1-r-z)^{\mu+2} + (1-z+r)^{\mu+2}}{(\mu+1)(\mu+2)}, \quad \text{if } r+z \leq 1. \end{aligned}$$

Here the case that $r+z > 1$ can be given as the single expression

$$\frac{(1-|z-r|)^{\mu+2}}{(\mu+1)(\mu+2)} + \frac{|z-r|(1-|z-r|)^{\mu+1}}{\mu+1};$$

the case $r+z \leq 1$ also gives a single expression in terms of $|z-r|$.

To evaluate the expression at (22) write $\int_0^1 (\dots) dr = \left(\int_0^{1-z} + \int_{1-z}^1 \right) (\dots) dr$. This gives two integrals

$$\begin{aligned} J_1 : & \\ & = \int_{1-z}^1 r(1-r)^v \left[\frac{(1-|z-r|)^{\mu+2}}{(\mu+1)(\mu+2)} + \frac{|z-r|(1-|z-r|)^{\mu+1}}{\mu+1} \right] dr \end{aligned}$$

and

$$\begin{aligned} J_2 := & \int_0^{1-z} r(1-r)^v \\ & \times \left[\frac{|z-r|(1-|z-r|)^{\mu+1} - (r+z)(1-r-z)^{\mu+1}}{\mu+1} + \frac{(1-|z-r|)^{\mu+2} - (1-r-z)^{\mu+2}}{(\mu+1)(\mu+2)} \right] dr. \end{aligned}$$

To evaluate J_1 we must distinguish the two cases according as $1-z >$ or $< z$, i.e. according as $z < \frac{1}{2}$ or $z > \frac{1}{2}$. For the simpler case $z < \frac{1}{2}$,

$$J_1 = \int_{1-z}^1 \frac{r(1-r)^v}{\mu+1} \left[\frac{(1-r+z)^{\mu+2}}{\mu+2} + (r-z)(1-r+z)^{\mu+1} \right] dr,$$

Table 2 A detailed calculation of the convolution formulae in \mathbb{R}^3 for special values of the parameters μ and ν

Parameters	$0 < z < 1$	$1 < z < 2$
$\mu = \nu = 1$	$\frac{10+24z-30z^2-20z^3+15z^4+12z^5-6z^6}{180z}$	$\frac{(-2+z)^4(-1+4z+2z^2)}{180z}$
$\mu = 2, \nu = 1$	$\frac{35+84z-280z^2+105z^3+126z^4-63z^5+6z^7}{1260z}$	$\frac{(-2+z)^5(-4+11z+6z^2)}{1260z}$
$\mu = 3, \nu = 1$	$\frac{7+16z-28z^2+35z^3+56z^4-21z^5+10z^7-3z^8}{420z}$	$\frac{(-2+z)^6(-1+2z+z^2)}{420z}$
$\mu = 4, \nu = 1$	$\frac{84+180z-336z^2-252z^3+504z^4-171z^5+10z^9}{7560z}$	$\frac{(-2+z)^7(-14+23z+10z^2)}{7560z}$
$\mu = 5, \nu = 1$	$\frac{1}{7560z}(60 + 120z - 240z^2 + 540z^3 - 1260z^4 + 1080z^5 - 315z^6 + 80z^7 - 18z^8 + 18z^9)$	$\frac{(-2+z)^8(-8+19z+20z^2)}{138600z}$
$\mu = \nu = 2$	$\frac{70+192z-728z^3-210z^4+840z^5-672z^6+264z^7-45z^8+4z^9}{5040z}$	$\frac{(-2+z)^6(-4+12z+9z^2)}{5040z}$
$\mu = 3, \nu = 2$	$\frac{42+120z-336z^2-1134z^3+864z^4-306z^5+44z^6+3z^{10}}{7560z}$	$\frac{(-2+z)^7(-2+4z+4z^2)}{5040z}$
$\mu = 4, \nu = 2$	$\frac{42+120z-384z^2+756z^3-1134z^4+864z^5-306z^6+44z^7+3z^{10}}{7560z}$	$\frac{(-2+z)^8(-2+4z+3z^2)}{7560z}$
$\mu = 5, \nu = 2$	$\frac{1}{90} + \frac{1}{252z} - \frac{5z^2}{126} + \frac{2z^4}{15} - \frac{z^5}{4} + \frac{3z^6}{14} - \frac{2z^7}{21} + \frac{19z^8}{756} - \frac{z^9}{360} + \frac{z^{10}}{4620}$	$\frac{(-2+z)^9(-16+27z+18z^2)}{83160z}$
$\mu = \nu = 3$	$\frac{1}{25200z}(126 + 400z - 1440z^2 - 630z^3 + 4536z^4 - 5040z^5 + 2640z^6 - 675z^7 + 120z^8 - 24z^{10})$	$\frac{(-2+z)^8(-3+8z+8z^2)}{25200z}$
$\mu = 4, \nu = 3$	$\frac{1}{138600z}(462 + 1540z - 6600z^2 + 21252z^3 - 34650z^4 + 26400z^5 - 10560z^6 + 2530z^7 - 363z^8 + 20z^{11})$	$\frac{(-2+z)^9(-8+19z+20z^2)}{138600z}$
$\mu = 5, \nu = 3$	$\frac{1}{27720z}(66 + 224z - 1100z^2 + 5016z^3 - 9702z^4 + 8844z^5 - 4620z^6 + 1540z^7 - 297z^8 + 38z^{11} - 6z^{12})$	$\frac{(-2+z)^{10}(-1+2z+2z^2)}{27720z}$
$\mu = 4, \nu = 4$	$\frac{1}{415800z}(924 + 3360z - 17600z^2 + 76824z^3 - 143220z^4 + 126720z^5 - 63360z^6 + 20240z^7 - 4356z^8 + 20240z^9 - 480z^{11} + 25z^{12})$	$\frac{(-2+z)^{10}(-8+20z+25z^2)}{415800z}$
$\mu = 4, \nu = 4$	$\frac{1}{165} + \frac{630z}{27} - \frac{z^2}{27} + \frac{3z^4}{14} - \frac{41z^5}{90} + \frac{7z^6}{15} - \frac{2z^7}{7} + \frac{22z^8}{189} - \frac{2z^9}{63} + \frac{13z^{10}}{2310} - \frac{47z^{11}}{83160} + \frac{z^{12}}{36036}$	$\frac{(-2+z)^{11}(-10+23z+30z^2)}{1081080z}$
$\mu = 5, \nu = 4$	$\frac{2}{429} + \frac{882z}{297} - \frac{10z^2}{21} + \frac{5z^4}{21} - \frac{71z^5}{126} + \frac{97z^6}{147} - \frac{10z^7}{21} + \frac{44z^8}{189} - \frac{2z^9}{63} + \frac{13z^{10}}{693} - \frac{5z^{12}}{18018} - \frac{z^{13}}{28028}$	$\frac{(-2+z)^{12}(-5+12z+18z^2)}{151312z}$

while for the case $z > \frac{1}{2}$ we use $\int_{1-z}^1 = \int_{1-z}^z + \int_z^1$ and find

$$J_1 = \int_{1-z}^z \frac{r(1-r)^v}{\mu+1} \left[\frac{(1-z+r)^{\mu+2}}{\mu+2} + (z-r)(1-z+r)^{\mu+1} \right] dr$$

$$+ \int_z^1 \frac{r(1-r)^v}{\mu+1} \left[\frac{(1-r+z)^{\mu+2}}{\mu+2} + (r-z)(1-r+z)^{\mu+1} \right] dr$$

$$:= J_{11} + J_{12}.$$

The rest of the formula is then deduced through simple, albeit tedious, algebra. Table 2 shows some special cases of the formulae above, obtained for μ and v positive integers.

References

- Abramowitz M, Stegun IA (1964) Handbook of mathematical functions, applied mathematical series 55. National Bureau of Standards, Washington. (reprinted 1968 by Dover Publications, New York.)
- Apanasovich TV, Genton MG (2010) Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika* 97:15–30
- Askey R (1973) Radial characteristic functions, technical report no. 1262. Mathematical Research Center, University of Wisconsin-Madison, Madison (1973)
- Banerjee S, Gelfand AE (2003) On smoothness properties of spatial processes. *J Multivar Anal* 84:85–100
- Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data. Chapman & Hall, Boca Raton
- Beatson RK, zu Castell W, Schroedl S (2009) Kernel based methods for vector-valued data with correlated components. Preprint. Helmholtz-Zentrum, München
- Benbouhaim M, Bouhamidi A (2005) Approximation of vector fields by thin plate splines with tension. *J Approx Theor* 136:198–229
- Berrocal VJ, Raftery AE, Gneiting T (2008) Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *Ann Appl Stat* 2:1170–1193
- de Boor C (1978) A practical guide to splines. Springer, New York
- Buhmann M (2001) A new class of radial basis functions with compact support. *Math Comput* 70:307–318
- Chilés J-P, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York
- Christakos G (2000) Modern spatiotemporal geostatistics. Oxford University Press, New York
- Christakos G, Hristopoulos D (1998) Spatiotemporal environmental health modeling: a tractatus stochasticus. Kluwer, Boston
- Cramér H (1940) On the theory of stationary random functions. *Ann Math* 41:215–230
- Cressie NAC (1993) Statistics for spatial data, revised edition. Wiley, New York
- Du J, Zhang H, Mandrekar V (2009) Infill asymptotic properties of tapered maximum likelihood estimators. *Ann Stat Bull Metrop Insur Co* 37:3330–3361
- Eckel AF, Mass CF (2005) Aspects of effective mesoscale short-range ensemble forecasting. *Weather Forecast* 20:328–350
- Furrer R, Genton M, Nychka D (2006) Covariance tapering for interpolation of large spatial datasets. *J Comput Graph Stat* 15:502–523
- Gaspari G, Cohn SE (1999) Construction of correlation functions in two and three dimensions. *Q J R Meteor Soc* 125:723–757
- Gneiting T (2001) Criteria of pólya type for radial positive definite functions. *Proc Am Math Soc* 129:2309–2318
- Gneiting T (2002a) Nonseparable, stationary covariance functions for space–time data. *J Am Stat Assoc* 97:590–600
- Gneiting T (2002b) Compactly supported correlation functions. *J Multivar Anal* 83:493–508
- Gneiting T, Kleiber W, Schlather M (2010) Matérn cross-covariance functions for multivariate random fields. *J Am Stat Assoc* 105:1167–1177
- Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York
- Goulard M, Voltz M (1992) Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Math Geol* 24:269–282
- Horn R.A., Johnson C.H. (1986) Matrix Analysis. Cambridge: Cambridge University Press
- Hristopoulos D, Porcu E (2012) Spartan vector-valued random fields. Submitted
- Kaufman C, Schervish M, Nychka D (2008) Covariance tapering for likelihood-based estimation in large spatial datasets. *J Am Stat Assoc* 103:1545–1555
- Li B, Genton MG, Sherman M (2008) Testing the covariance structure of multivariate random fields. *Biometrika* 95: 813–829
- Majumdar A, Gelfand AE (2007) Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Math Geol* 39:225–245
- Matérn B (1986) Spatial variation, 2nd edn. Springer, Berlin
- Matheron G (1962) *Traité de Géostatistique appliquée*, tome 1 (1962), tome 2 (1963). Editions Technip, Paris
- Misiewicz J (1989) Positive definite functions on l_∞ . *Statist Probab Lett* 8:255–260
- Narcowich FJ, Ward JD, Wright GB (2007) Divergence-free RBFs on surfaces. *J Fourier Anal Appl* 13:643–663
- Powell MJD (1981) Approximation theory and methods. Cambridge University Press, Cambridge
- Porcu E, Gregori P, Mateu J (2006) Nonseparable stationary anisotropic space–time covariance functions. *Stoch Environ Res Risk Assess* 21:113–122
- Porcu E, Zastavnyi V (2011) Characterization theorems for some classes of covariance functions associated to vector valued random fields. *J Multivar Anal* 102(9):1293–1301
- Schaback R (2011) The missing Wendland functions. *Adv Comput Math* 34(1):67–81
- Schoenberg IJ (1938) Metric spaces and completely monotone functions. *Ann Math Ann* 39:811–841
- Schlather M (2010) Some covariance models based on normal scale mixtures. *Bernoulli* 16:780–797
- Schmidt AM, Gelfand AE (2003) A Bayesian coregionalization approach for multivariate pollutant data. *J Geophys Res Atmos* 108: D24 (article no. 8783)
- Stein ML (1999) Interpolation of spatial data. Springer, New York
- Ver Hoef JM, Barry RP (1998) Constructing and fitting models for cokriging and multivariable spatial prediction. *J Stat Plan Inf* 69:275–294
- Wackernagel H (2003) Multivariate geostatistics, 3rd edn. Springer, Berlin
- Wendland H (1994) Ein Beitrag zur Interpolation mit radialen Basisfunktionen. Diplomarbeit, Göttingen

- Wendland H (1995) Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv Comput Math* 4:389–396
- Wendland H (2005) *Scattered data approximation*. Cambridge monographs on applied and computational mathematics. Cambridge University Press, Cambridge
- Whittaker ET, Watson GN (1927) *A course of modern analysis*, 4th edn. Cambridge University Press, Cambridge
- Williamson RE (1956) Multiply monotone functions and their Laplace transforms. *Duke Math J* 23:189–207
- Zastavnyi VP (2004) On properties of the Buhmann function. *Ukrainian Math J* 58:1184–1208
- Zastavnyi VP (2008) Problems related to positive definite functions. In: Mateu J, Porcu E (eds) *Positive definite functions: from Schoenberg to space–time challenges*. Editorial Universitat Jaume I, Barcelona, p 63–114