INAUGURAL-DISSERTATION

zur Erlangung der Doktorwürde der

NATURWISSENSCHAFTLICH-MATHEMATISCHEN
GESAMTFAKULTÄT

der

RUPRECHT-KARLS-UNIVERSITÄT
HEIDELBERG

vorgelegt von
Diplom-Mathematiker

Jochen Fiedler

aus Fulda

Tag der mündlichen Prüfung: _____

Distances, Gegenbauer expansions, curls, and dimples: On dependence measures for random fields

Betreuer: **Prof. Dr. Tilmann Gneiting**
Prof. Dr. Donald Richards

Zusammenfassung

Ziel der Arbeit ist es, verschiedene Aspekte von Abhängigkeitsstrukturen stochastischer Prozesse zu untersuchen. Zu diesem Zweck werden neue Abhängigkeitsmaße für räumliche stochastische Prozesse eingeführt, sowie verschiedene analytische Eigenschaften von Korrelationsfunktionen untersucht und mit Merkmalen der Realisierungen der zugehörigen Prozesse verknüpft.

Die neu eingeführten Abhängigkeitsmaße basieren auf der Distanzkorrelation und sind dadurch für größere Klassen von Zufallsfeldern definiert als die bekannte Korrelationsfunktion nach Pearson. Zusätzlich ermöglichen sie das Quantifizieren von nicht-linearen Abhängigkeiten, sowie von Abhängigkeiten in multivariaten stochastischen Prozessen.

Die Untersuchung von analytischen Eigenschaften von Korrelationsfunktionen beschäftigt sich zum Einen mit isotropen, positiv definiten Funktionen auf Sphären verschiedener Dimensionen. Diese werden charakterisiert durch Reihen von Gegenbauerpolynomen mit sogenannten Schönbergkoeffizienten. Der konkrete Zusammenhang zwischen diesen Schönbergkoeffizienten wird hier angegeben.

Zum Anderen beschäftigt sich die Analyse mit der Charakterisierung von Korrelationsfunktionen von zufälligen Vektorfeldern auf der Sphäre, deren Realisierungen wirbel- und quellfrei sind.

Des Weiteren wird die Grübchen-Eigenschaft von räumlich-zeitlichen Zufallsfeldern studiert. Das Grübchen entspricht einem nicht-monotonen Verhalten der Korrelationsfunktion in der Zeit, welches durch sogenannte Transportfelder erzeugt werden kann. Diese Transportfelder und ihre Verbindung zu der Grübchen-Eigenschaft werden genauer analysiert.

Abstract

The aim of this thesis is to analyze several aspects of dependence structures for stochastic processes. To this end, new dependence measures for spatial stochastic processes will be introduced. Further, different analytical properties of correlation functions and their relation to properties of the realizations of the corresponding stochastic processes will be studied.

The newly introduced dependence measures are based on the distance correlation. Thus, they are defined for larger classes of processes than the well-known Pearson correlation function. In addition, the new dependence measures allow to quantify non-linear dependencies, as well as dependencies in multivariate stochastic processes.

On the one hand, the investigation of analytical properties of correlation functions considers isotropic positive definite functions on spheres of different dimensions. These functions are characterized by Gegenbauer expansions involving so-called Schoenberg coefficients. Relationships between those Schoenberg coefficients will be given.

On the other hand, characterizations of correlation functions of random vector fields with almost surely divergence-free and irrotational realizations will be considered.

Further, the dimple property of spatio-temporal random fields will be analyzed. This dimple corresponds to a non-monotonic temporal behaviour of the correlation function, which can be generated by so-called transport fields. These transport fields and their relation to the dimple property will be investigated.

Contents

1 Introduction

Spatial and spatio-temporal random processes play important roles in many research fields, such as meteorology, geology, econometrics, environmetrics, and sociology [16, 18, 33, 111, 24, 36]. These areas deal with phenomena that naturally possess spatial or spatio-temporal dimensions, like temperature or the distribution of ore in soil. Modelling such quantities as stochastic processes often helps for inference or prediction tasks. A crucial property that has to be taken into account is the geometry of the spatial domain of the processes, as this domain might be, for example, a Euclidean space or a curved subset of a sphere.

In order to understand stochastic processes in time and space it is crucial to have meaningful descriptions of their dependence structures. Such a description is attained by associating a dependence measure with every pair of random vectors corresponding to any two points in space or space-time. Then it is important to use a dependence measure which is applicable in the specific situation, and which admits a meaningful interpretation. On the one hand, not all dependence measures are defined for all types of processes. A dependence measure might require the process to possess specific properties, like the existence of second moments. On the other hand, a dependence measure might fail to measure the predominant type of dependence, which could be, for example, highly non-linear. Consequently, the choice of the dependence measure is determined by the nature of the considered process and the particular research question.

In the statistics literature, many concepts of dependence between random variables have been developed, like the well-known Pearson correlation coefficient, Kendall's τ or Spearman's ρ [81, 57, 102]. A particularly important role plays the Pearson correlation coefficient, which is a measure of linear dependencies, cf. [112]. In many applications, researchers assume that the data generating process is Gaussian, and the dependence structure of Gaussian processes is characterized by the Pearson correlation function. The Pearson correlation function is the function that assigns, to any two points in time and space, the Pearson correlation coefficient between the corresponding random variables.

As positive definite functions correspond to Pearson correlation functions, the

study of positive definite functions on domains such as Euclidean spaces or spheres is an important part of spatial statistics. Knowing the properties of classes of positive definite functions helps modelling the properties of realizations of random fields. For example, it is well-known that a Gaussian stochastic process with sample paths of a particular roughness corresponds to classes of positive definite functions with certain smoothness properties, cf. [44, 93, 116]. A more complex case can be found in Scheuerer and Schlather [94], who showed that random vector fields in Euclidean spaces with curl- or divergence-free realizations can be characterized by certain classes of covariance functions. Another example can be given for spatio-temporal random processes: Kent, Mohammadzadeh, and Mosammam [58] showed that the Gneiting class of nonseparable covariance functions [39] possesses a certain non-monotonic behaviour, which they call the dimple property.

This work contributes to the spatial statistics literature in several ways. Motivated by the work of Zhou [117], Chapter 2 introduces the distance correlation function and the distance variogram as dependence measures for spatial processes. Both dependence measures are based on the concept of distance correlation between random vectors developed by Székely, Rizzo, and Bakirov [108]. The distance correlation has the advantage that it is defined in cases where the Pearson correlation coefficient is not, and that it is able to measure non-linear dependence. It is demonstrated that the distance correlation function is a suitable dependence measure for large classes of processes. In particular, stable processes are considered, which constitute a generalization of Gaussian processes. Stable non-Gaussian processes have no finite second moments, so that the Pearson correlation function is not defined. It is shown that the distance variogram for stable processes with finite first moments is a natural generalization of the usual variogram for Gaussian processes. Furthermore, it is proved that the distance correlation function and the distance variogram can be consistently estimated. The finite sample performance of these estimators is analyzed in a simulation study.

Chapter 3 concerns the properties of isotropic positive-definite functions on d-dimensional spheres, where for each dimension there exists a unique Gegenbauer expansion. The relations between the Gegenbauer expansions are studied, and the connection coefficients determining these relations are explicitly given.

In Chapter 4 the work of Scheuerer and Schlather [94] is extended to the spherical domain. Specifically, it is analyzed how curl- and divergence-free Gaussian random vector fields can be characterized by properties of their Pearson cross-covariance functions. It is shown that univariate covariance functions can be used to construct bivariate random processes that correspond to random vector

processes without curls or divergences. To this end, the concept of stochastic differential forms is applied.

The final Chapter 5 deals with the dimple property of certain spatio-temporal correlation functions. This property was first described by Kent, Mohammadzadeh, and Mosammam [58] who considered it to be counterintuitive, as the correlation might be increasing with the time lag. However, there are natural processes with transport effects that possess this behaviour.

2 Distance correlation for random fields

The Pearson correlation coefficient – or the corresponding covariance – is undoubtedly the most frequently used dependence measure in spatial statistics, because of the following reasons:

- The Pearson correlation coefficient has a clear interpretation in the context of assessing a linear relationship between two random variables. That is, the more linear the relationship between the variables, the higher is the absolute value of their Pearson correlation coefficient. This property is widely used for linear interpolation or extrapolation of spatial data to unobserved locations, which is called *kriging* in the geostatistical context (cf., [61, 110]).

- It is well-known that the class of Pearson covariance functions is the same as the class of positive definite functions, as every covariance matrix is symmetric and positive semi-definite. Positive definite functions initially were analyzed by Schoenberg [99], and, since then, they have been studied extensively by researchers from several different areas. As a consequence, there are numerous parametric models available for use as correlation functions, and this provides researchers with much flexibility in modelling.

- In many cases, the *random field (RF)* generating the observed spatial data can be assumed to be a centered Gaussian process.[1] Such a process is completely characterized by its Pearson covariance function, because it determines all finite-dimensional distributions.

Despite its advantages, the Pearson correlation coefficient has properties that limit its applicability, interpretability, and validity.

[1] A stochastic process $\{Z(t) : t \in T\}$ on an index set T is called *Gaussian* if all of its finite-dimensional distributions are Gaussian. In other words, for all integers $n \geq 1$ and $t_1, \ldots, t_n \in T$ the distribution of the random vector $(Z(t_1), \ldots, Z(t_n))^t$ is n-variate Gaussian. In addition, a Gaussian process $\{Z(t) : t \in T\}$ is *centered*, if its mean function is constantly zero, i.e., $\mathbb{E}(Z(t)) = 0$ for all $t \in T$.

- It is not defined for random variables with *heavy-tailed behaviour*, i.e., random variables without *finite second moments*.[2]
- It only measures linear dependence.
- It cannot be used to investigate the strength of the linear dependence between random vectors with more than one component.

Concerning the first point, it is necessary in many cases to assume that the data-generating process has heavy tails. Examples are applications in modeling anomalous diffusion (e.g. [10]), hydrology, where Benson, Wheatcraft, and Meerschaert [8] modeled concentration profiles for tracer plumes in groundwater, or finance, where Mandelbrot [70] found that certain daily price changes seemed to have infinite variance. Further examples can be found in Adler, Feldmann, and Taqqu [3], Uchaikin and Zolotarev [109] or Resnick [85], who investigated heavy-tailed phenomena appearing in physics, finance, computer science, or signal processing.

It is evident that the usual Central Limit Theorem cannot be applied to sums of independent random variables with heavy tails. In other words, if $(Y_n)_{n\in\mathbb{N}}$ is a sequence of *independent and identically distributded (i.i.d.)* random variables with infinite variance, there are no sequences $d_n \in (0,\infty)$ and $a_n \in \mathbb{R}$ such that

$$S_n := \frac{1}{d_n}\sum_{i=1}^{n} Y_i + a_n$$

converges in distribution to a Gaussian random variable. However, if there are $d_n > 0, a_n \in \mathbb{R}$ and a random variable X such that S_n converges in distribution to X, then X necessarily follows a *stable distribution*, whose definition and properties will be given in the following section.

The above convergence property emphasizes the importance of stable distributions as they can be interpreted as generalizations of Gaussian distributions. Further, stable distributions are the fundamental solutions to fractional-order forms of advection-dispersion equations [8] and they have been used to model spatial risks in insurance [103]. Since stable distributions arise as limiting distributions of appropriately normalized sums of random variables with possibly infinite variances,

[2]Note that this is not the most general definition used in the literature. Resnick [84] defined a random variable X to be heavy-tailed if there exists some $\alpha > 0$ such that $P(X > x) = x^{-\alpha}L(x)$ for $x > 0$, where L is a *slowly varying* function, i.e., $\lim_{t\to\infty} L(tx)/L(t) = 1$ for all $x > 0$. This definition implies that X has infinite variance if $\alpha < 2$.

they also have infinite variance in general. This implies that it is never justified to use the Pearson correlation coefficient to investigate the dependence structure between stable non-Gaussian random variables. Moreover, using the empirical Pearson correlation for statistical inference might be misleading, as illustrated in Subsection 2.1.3.

For these reasons, several different approaches were developed to define dependence measures for stable processes. Examples are covariance-like functions (e.g. [6]) or the covariation and the codifference (see [91] or [98, Section 11.2.4]). We will describe in Subsection 2.1.1 the latter two in greater detail.

The aim of this chapter is to introduce the *distance correlation function (DCF)* as a dependence measure for spatial stochastic processes. We show that the DCF has interesting properties for statistical inference for many random fields, in particular, for stable random fields. Distance correlation was introduced by Székely, Rizzo, and Bakirov [108] as a dependence measure between random vectors and, since then, has attracted much interest (cf., [27, 26, 117, 72, 88]). Its key features are:

- The applicability to random vectors with finite first but possibly infinite second moments.
- It equals zero if and only if the random vectors are stochastically independent, implying that it is able to measure non-linear dependencies.
- It is defined for random vectors with arbitrary dimensions. This allows us to study the dependence between random vectors of different dimensions.
- The empirical distance correlation is simple to calculate and provides in many cases a weakly consistent estimator of the population distance correlation.

Consequently, the distance correlation can be applied to random variables which are heavy-tailed, in particular to stable distributions. In generalizing the approach of Zhou [117], we will define and use the DCF to analyze the dependence structure of stationary random fields. Further, we will introduce the *distance variogram* as a natural generalization of the usual variogram in the case of stable random fields. We will prove the consistency of the empirical distance correlation function as an estimator for the DCF and illustrate its applicability in a simulation study.

2.1 Stable processes and some limitations of the Pearson correlation function

That the Pearson correlation coefficient yields no plausible dependence measure for random processes without finite second moments is a well-known problem in the statistical community. This problem has led to lots of research activity with the aim of developing suitable dependence measures for such processes. A particularly large and flexible class of heavy-tailed processes occurring naturally in many situations is the class of stable processes. For this reason, we will use members of this class in examples and simulations in the following sections. The current section serves as preparation, as it contains definitions and results from the theory of stable processes, and it ends with examples of simulated random fields (RFs) where the use of the empirical Pearson correlation function leads to false conclusions.

2.1.1 Stable distributions

The usual Central Limit Theorem shows that Gaussian distributions can be characterized as the limiting distributions of appropriately normalized sums of i.i.d. random vectors with finite second moments. If the second moments do not exist then the limiting distributions are no longer Gaussian; instead, they are stable.

Many results and definitions of this subsection are drawn from the comprehensive work of Samorodnitsky and Taqqu [91], which provides an in-depth treatment of stable distributions and processes.

Here and in the remainder, we will use the notation $U \overset{d}{\sim} V$ to indicate that the random vectors U and V have the same distribution.

Definition 2.1. Let X be a random vector with values in $\mathbb{R}^d$ and let the random vectors $X_1, \ldots, X_n$ be independent copies of X. Then X is called *stable*, if for all integers $n \geq 2$ there exist $b_n > 0$ and $a_n \in \mathbb{R}^d$ such that

$$\sum_{i=1}^{n} X_i \overset{d}{\sim} b_n X + a_n.$$

If in addition $-X \overset{d}{\sim} X$, then the random vector X is called *symmetric stable*.

Corollary 2.1.3 in [91] shows that for every stable random vector X the sequence $(b_n)_{n\in\mathbb{N}}$ appearing in the definition above can always be chosen such that there

exists a unique $0 < \alpha \leq 2$ with $b_n = n^{1/\alpha}$. Hence, a (symmetric) stable random vector is called *(symmetric) α-stable*. Here and in the sequel, the abbreviation $S\alpha S$ refers to *symmetric α-stable*.

The probability density and distribution functions of α-stable random vectors are only in few cases expressible in closed analytic forms. For example, if $\alpha = 2$ then the distribution is Gaussian; if $\alpha = 1$ then it belongs to the Cauchy class. In most other cases, the density and distribution functions are expressible only as series expansions, cf. [109, Chapter 4].

By contrast, the characteristic functions of stable random vectors are always expressible in closed analytic form [91, Theorem 2.3.1], which we now state below in Theorem 2.2. This characterization is particularly important for us because, as we will see in Section 2.2, the distance correlation is defined in terms of characteristic functions.

Here and in the following, we denote by $\langle \cdot, \cdot \rangle$ the *standard scalar product* on the corresponding Euclidean space.

Theorem 2.2. Let $0 < \alpha \leq 2$. A random vector X with values in $\mathbb{R}^d$ is α-stable if and only if there exists a finite measure Λ on the unit sphere $\mathbb{S}^d$ and a vector $\mu \in \mathbb{R}^d$ such that the characteristic function $\varphi_X(t) = \mathbb{E} \exp\left(i\langle t, X\rangle\right), t \in \mathbb{R}^d$, takes the following form:

(a) If $\alpha = 1$ then

$$\varphi_X(t) = \exp\left[-\int_{\mathbb{S}^d} |\langle t, u\rangle| \left(1 + i\frac{2}{\pi}\left(\operatorname{sign}\langle t, u\rangle\right) \log|\langle t, u\rangle|\right) \Lambda(\mathrm{d}u) + i\langle t, \mu\rangle\right].$$

(b) If $\alpha \neq 1$ then

$$\varphi_X(t) = \exp\left[-\int_{\mathbb{S}^d} |\langle t, u\rangle|^\alpha \left(1 - i\left(\operatorname{sign}\langle t, u\rangle\right) \tan(\pi\alpha/2)\right) \Lambda(\mathrm{d}u) + i\langle t, \mu\rangle\right].$$

The *spectral representation* (Λ, μ) is unique.

If we are given jointly α-stable distributed random variables[3] $X_1, \ldots, X_d$ – meaning that the random vector $(X_1, \ldots, X_d)^t$ is α-stable – the theorem above states

[3]Here and in the remainder of this thesis, the term *random variable* refers solely to univariate quantities, i.e., if X is a random variable then it takes values in $\mathbb{R}$. If a quantity is possibly multivariate, we call it a *random vector*.

that the dependence structure between these random variables is completely determined by the *index* α, the *spectral measure* Λ, and μ. We will sometimes write $X \sim \mathcal{S}_\alpha(\Lambda, \mu)$ to indicate that a random vector X is α-stable with spectral representation (Λ, μ).

Remark 2.3. For the case in which X is an $S\alpha S$ random vector, Theorem 2.4.3 in [91] shows that the characteristic function of X takes the simpler form

$$\varphi_X(t) = \exp\left[-\int_{\mathbb{S}^d} |\langle t, u\rangle|^\alpha \, \Lambda(\mathrm{d}u)\right],$$

for all $0 < \alpha \leq 2$.

Now let X be a α-stable random variable with spectral measure Λ. Since X is univariate, Λ is concentrated on $\mathbb{S}^1 = \{-1, 1\}$ and it directly follows from Theorem 2.2 that the characteristic function of X equals, for $\alpha \neq 1$,

$$\varphi_X(t) = \exp\left[-\sigma^\alpha |t|^\alpha \left(1 - i\beta \operatorname{sign}(t) \tan\frac{\pi\alpha}{2}\right) + i\mu t\right]$$

and, for $\alpha = 1$,

$$\varphi_X(t) = \exp\left[-\sigma^\alpha |t|^\alpha \left(1 + i\beta \frac{2}{\pi} \operatorname{sign}(t) \log|t|\right) + i\mu t\right],$$

where

$$\sigma = \big(\Lambda(\{1\}) + \Lambda(\{-1\})\big)^{1/\alpha} \qquad \text{and} \qquad \beta = \frac{\Lambda(\{1\}) - \Lambda(\{-1\})}{\Lambda(\{1\}) + \Lambda(\{-1\})}.$$

Consequently, in this case the spectral measure Λ is characterized by the following three parameters: the *scale parameter* $\sigma \geq 0$, the *skewness parameter* $\beta \in [-1, 1]$, and the *shift parameter* $\mu \in \mathbb{R}$. If X is additionally symmetric then $\beta = 0$.

Similar to the multivariate case, we sometimes will write $X \sim \mathcal{S}_\alpha(\sigma, \beta, \mu)$ to indicate that a univariate random variable X is α-stable with scale σ, skewness β and shift μ. If the univariate α-stable random variable X is additionally symmetric, then $\Lambda(\{1\}) = \Lambda(\{-1\})$, implying $\beta = 0$, and the characteristic function takes for all $0 < \alpha \leq 2$ the form

$$\varphi_X(t) = \exp(-\sigma^\alpha |t|^\alpha).$$

Note that this representation yields immediately that X is Gaussian, resp. Cauchy

distributed in the case $\alpha = 2$, resp. $\alpha = 1$.

Remark 2.4. Property 1.2.16 in [91] shows that stable random variables with $\alpha < 2$ are heavy-tailed. Specifically, let $0 < \alpha < 2$ and $X \sim \mathcal{S}_\alpha(\sigma, \beta, \mu)$. Then

(1) $\mathbb{E}|X|^p < \infty$ for all $0 < p < \alpha$, and

(2) $\mathbb{E}|X|^p = \infty$ for $p \geq \alpha$.

Using characteristic functions, it is straightforward to show that linear combinations of independen stable random variables are again stable [91, Properties 1.2.1 and 1.2.3]. Specifically, let $X_i \sim \mathcal{S}_\alpha(\sigma_i, \beta_i, \mu_i)$ for $i = 1, 2$ be independent. Then $X_1 + X_2 \sim \mathcal{S}_\alpha(\sigma, \beta, \mu)$ with

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}, \quad \beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad \mu = \mu_1 + \mu_2, \tag{2.1}$$

and

$$\begin{aligned} aX_1 &\sim \mathcal{S}_\alpha(|a|\sigma_1,\, \mathrm{sign}(a)\beta_1,\, a\mu_1), && \text{if } \alpha \neq 1, \\ aX_1 &\sim \mathcal{S}_1(|a|\sigma_1,\, \mathrm{sign}(a)\beta_1,\, a\mu_1 - \frac{2}{\pi} a \log|a|\sigma_1\beta_1), && \text{if } \alpha = 1, \end{aligned} \tag{2.2}$$

where $a \in \mathbb{R}$.

Given several jointly α-stable distributed random variables, it is natural to seek to define a meaningful dependence measure between them – a problem tackled by several researchers. Press [83] introduced the *association parameter*, and this was extended by Paulauskas [80] to the *generalized association parameter*. Miller [74] developed the *covariation* for two jointly $S\alpha S$ random variables with $1 < \alpha \leq 2$; this was used by Samorodnitsky and Taqqu [91] to introduce the *codifference*, which itself was generalized to the *signed symmetric covariation coefficient* by Garel and Kodia [35].

Here we are interested in the codifference and the covariation. We note that the codifference for α-stable processes is closely related to the *distance variogram*, which will be introduced in Section 2.4.

Definition 2.5. Let X, Y be jointly distributed $S\alpha S$ random variables, where $1 < \alpha \leq 2$.

(a) The *covariation* of X and Y is defined as

$$[X, Y]_\alpha = \int_{\mathbb{S}^2} u_1 \operatorname{sign}(u_2)|u_2|^\alpha \, \Lambda(\mathrm{d}u),$$

where Λ is the spectral measure of the stable random vector $(X, Y)^t$.

(b) The *covariation norm* of X is defined as

$$|||X|||_\alpha = ([X, X]_\alpha)^{1/\alpha}.$$

Remark 2.6. (1) For $\alpha > 1$, the covariation norm indeed is a seminorm on certain vector spaces of jointly $S\alpha S$ random variables, cf. [91, Section 2.8]. Specifically, on these spaces the covariation norm is *absolutely homogeneous*, i.e., $|||\lambda X|||_\alpha = |\lambda| \, |||X|||_\alpha$ for $\lambda \in \mathbb{R}$; satisfies the *triangle inequality*, i.e., $|||X + Y|||_\alpha \leq |||X|||_\alpha + |||Y|||_\alpha$; and $|||X|||_\alpha = 0$ if and only if $X = 0$ almost surely.

(2) We also use the notation $|||X|||_\alpha$ to denote the scale parameter of an $S\alpha S$ random variable X with $0 < \alpha \leq 2$. Corollary 2.7.6 in [91] shows that this yields no conflict, because the covariation norm of an $S\alpha S$ random variable X with $1 < \alpha \leq 2$ equals its scale parameter σ.

Samorodnitsky and Taqqu [91] introduced the codifference as a dependence measure between two jointly $S\alpha S$ random variables. Since it is based on the scale parameter, it has the advantage over the covariation that α is allowed to be less than one.

Definition 2.7. Let X, Y be jointly distributed $S\alpha S$ random variables, with $0 < \alpha \leq 2$. Then the *codifference* between X and Y is defined as

$$\tau(X, Y) = |||X|||_\alpha^\alpha + |||Y|||_\alpha^\alpha - |||X - Y|||_\alpha^\alpha.$$

Remark 2.8. Here we summarize some properties of the covariation and the codifference, cf. Theorems 11.1 and 11.2 in [98].

(1) Let X, Y, Z be jointly α-stable random variables with $1 < \alpha \leq 2$. Then the following holds.

(i) For $a, b \in \mathbb{R}$

$$[aX + bY, Z]_\alpha = a[X, Z]_\alpha + b[Y, Z]_\alpha,$$

i.e., the covariation is linear in the first argument.

(ii) If X and Y are independent, then $[X, Y]_\alpha = 0$.

(iii) If X and Y are Gaussian, i.e., $\alpha = 2$, we have $2[X, Y]_\alpha = \mathrm{Cov}(X, Y)$.

(2) Let X, Y be jointly α-stable with $0 < \alpha \leq 2$. Then the following holds.

(i) The codifference is symmetric, $\tau(X, Y) = \tau(Y, X)$.

(ii) If X and Y are independent, then $\tau(X, Y) = 0$. The converse statement holds true if and only if $0 < \alpha < 1$ or $\alpha = 2$.

(iii) If X and Y are Gaussian, then $\tau(X, Y) = \mathrm{Cov}(X, Y)$.

(iv) Let $(X, Y)^t$ and $(U, V)^t$ be $S\alpha S$ random vectors with $X, Y, U, V \sim \mathcal{S}_\alpha(\sigma, 0, 0)$ for some $\sigma > 0$. If $\tau(X, Y) \leq \tau(U, V)$ then for any $c > 0$

$$P(|X - Y| > c) \geq P(|U - V| > c).$$

In other words, pairs of random variables with a higher codifference possess a greater probability to have similar values.

2.1.2 Stable random processes

Bearing in mind the definition of stable random vectors, the concept of a stable random process can be defined in a straightforward manner, see [91, Chapter 3], [98, Chapter 11] or [104, Chapter 9].

Definition 2.9. Let $\{Z(t) : t \in T\}$ be a univariate stochastic process with an arbitrary index set T. The process is called *α-stable,* if for every integer $d \geq 1$ the random vector

$$Z = (Z(t_1), \ldots, Z(t_d))^t$$

is α-stable distributed for each choice of $t_1, \ldots, t_d \in T$. If the distribution of Z is symmetric also, then the process is called *symmetric α-stable ($S\alpha S$)*.

An important method for constructing and characterizing certain classes of stable processes is via *stable integrals* of measurable functions f, which are constructed as follows, cf. Section 11.2.3.2 in [98] and Section 3.4 in [91].

Here and in the remainder of this subsection, we refer to $(E, \mathcal{E}, m)$ as a measure space with a σ-finite measure m, and we define $\mathcal{E}_0 := \{A \in \mathcal{E} : m(A) < \infty\}$.

An *independently scattered random measure* M is defined as a function on $\mathcal{E}_0$ such that:

(i) $M(A)$ is a random variable for all $A \in \mathcal{E}_0$;

(ii) for any integer $n > 0$ and pairwise disjoint $A_1, \ldots, A_n \in \mathcal{E}_0$, the random variables $M(A_1), \ldots, M(A_n)$ are independent;

(iii) for any sequence of pairwise disjoint sets $A_1, A_2, \ldots \in \mathcal{E}_0$ with $\bigcup_{i=1}^{\infty} A_i \in \mathcal{E}_0$, there holds $M(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} M(A_i)$ almost surely.

An independently scattered random measure M is called α-*stable*, if there exists a measurable function $\beta : E \to [-1, 1]$, such that

$$M(A) \sim \mathcal{S}_\alpha \left(m(A)^{1/\alpha}, \frac{\int_A \beta(x)\, m(\mathrm{d}x)}{m(A)}, 0 \right) \tag{2.3}$$

for all $A \in \mathcal{E}_0$. The measure m is called the *control measure* of M and β is called its *skewness function*. If the skewness function is identically zero, then M is $S\alpha S$.

Now let $f : E \to \mathbb{R}$ be a simple function, i.e., $f(x) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(x)$ with $A_1, \ldots, A_n \in \mathcal{E}_0$ being a collection of pairwise disjoint sets, $\mathbf{1}_{A_i}$ the indicator function on A_i, and $a_1, \ldots, a_n \in \mathbb{R}$. Then the stable integral of f with respect to an α-stable random measure M is defined as

$$I(f) = \int_E f(x)\, M(\mathrm{d}x) = \sum_{i=1}^{n} a_i M(A_i).$$

For a general function f contained in

$$L^\alpha(E, m) := \left\{ f : E \to \mathbb{R} \text{ such that } \int_E |f(x)|^\alpha\, m(\mathrm{d}x) < \infty \right\}, \tag{2.4}$$

consider a sequence $(f_n)_{n\in\mathbb{N}}$ of simple functions with $f_n(x) \to f(x)$ for $n \to \infty$ and all $x \in E$. Then the integral of f with respect to an α-stable random measure M is defined as the stochastic limit of the sequence $(I(f_n))_{n\in\mathbb{N}}$, which is also denoted by $I(f)$. This construction is well-defined, cf. [91, Section 3.4].

A consequence of this construction is the linearity of the stable integral [91, Property 3.2.3].

Proposition 2.10. Let $0 < \alpha \leq 2, (E, \mathcal{E}, m)$ be a measure space with σ-finite measure m, and M an α-stable random measure with control measure m. Then the stable integral with respect to M is linear, i.e., if $f_1, f_2 \in L^\alpha(E, m)$ then

$$\lambda_1 I(f_1) + \lambda_2 I(f_2) = I(\lambda_1 f_1 + \lambda_2 f_2), \qquad \text{almost surely,}$$

for all $\lambda_1, \lambda_2 \in \mathbb{R}$.

Example 2.11. We derive the distribution of a stable integral of a simple function f. Thus, let $f(x) = \sum_{j=1}^n a_j \mathbf{1}_{A_j}(x)$ with $A_1, \ldots, A_n \in \mathcal{E}_0$ being a collection of pairwise disjoint sets and non-zero $a_1, \ldots, a_n \in \mathbb{R}$. Further, let M be an $S\alpha S$-stable random measure with control measure m. We assume $\alpha \neq 1$, as this simplifies the necessary calculations.

From the construction of the stable integral it follows that $I(f) = \sum_{j=1}^n a_j M(A_j)$, which is a linear combination of independent α-stable random variables. Then (2.3) together with (2.1) and (2.2) yields $I(f) \sim \mathcal{S}_\alpha(\sigma, \beta, 0)$ with scale parameter

$$\sigma = \left[\sum_{j=1}^n |a_j|^\alpha m(A_j)\right]^{1/\alpha} = \left[\int_E |f(x)|^\alpha \, m(\mathrm{d}x)\right]^{1/\alpha}$$

and skewness parameter

$$\begin{aligned}
\beta &= \frac{\sum_{j=1}^n \beta_j \operatorname{sign}(a_j)|a_j|^\alpha m(A_j)}{\sum_{j=1}^n |a_j|^\alpha m(A_j)} \\
&= \frac{\sum_{j=1}^n \operatorname{sign}(a_j)|a_j|^\alpha m(A_j) m(A_j)^{-1} \int_{A_j} \beta(x)\, m(\mathrm{d}x)}{\sum_{j=1}^n |a_j|^\alpha m(A_j)} \\
&= \frac{\int_E \beta(x) \operatorname{sign}(f(x))|f(x)|^\alpha \, m(\mathrm{d}x)}{\int_E |f(x)|^\alpha \, m(\mathrm{d}x)}.
\end{aligned}$$

In the following proposition we will see that it is not only possible to transfer the above example to general integrable functions [91, Property 3.2.2], but also that it is possible to construct multivariate random vectors via stable integrals [91, Proposition 3.4.2].

Proposition 2.12. Let $0 < \alpha \leq 2$ and $(E, \mathcal{E}, m)$ be a measure space with σ-finite measure m. Further, let M be an α-stable random measure with control measure

m and skewness function β.

(1) If $f \in L^\alpha(E,m)$ then for $\alpha \neq 1$

$$I(f) \sim S_\alpha(\sigma_f, \beta_f, 0)$$

with

$$\sigma_f = \left(\int_E |f(x)|^\alpha \, m(\mathrm{d}x) \right)^{1/\alpha}$$

and

$$\beta_f = \frac{\int_E \operatorname{sign}(f(x)) |f(x)|^\alpha \beta(x) \, m(\mathrm{d}x)}{\int_E |f(x)|^\alpha \, m(\mathrm{d}x)}.$$

Now let $\alpha = 1$. If it is additionally assumed that

$$\int_E \left| f(x)\beta(x) \log |f(x)| \right| m(\mathrm{d}x) < \infty, \tag{2.5}$$

then

$$I(f) \sim S_1(\sigma_f, \beta_f, \mu_f).$$

where σ_f and β_f are given as above and

$$\mu_f = -\frac{2}{\pi} \int_E f(x)\beta(x) \log |f(x)| \, m(\mathrm{d}x).$$

In the case in which M additionally is symmetric the stable integral $I(f)$ is an $S\alpha S$ random variable, for any $0 < \alpha \leq 2$.

Note that condition (2.5) is automatically fulfilled if M is an $S\alpha S$ random measure, as this implies $\beta \equiv 0$.

(2) Let $n \geq 1$ be an integer and $f_1, \ldots, f_n \in L^\alpha(E,m)$. If $\alpha = 1$ we additionally assume that each function individually fulfills (2.5). Then the random vector $X = (I(f_1), \ldots, I(f_n))^t$ follows an α-stable distribution, where $I(f_1), \ldots, I(f_n)$ are stable integrals with respect to M. In particular, the characteristic function of X for $\alpha \neq 1$ equals

$$\begin{aligned} &\varphi_X(t_1, \ldots, t_n) \\ &= \exp\left[-\int_E \left| \sum_{j=1}^n t_j f_j(x) \right|^\alpha \left(1 - i\beta(x) \operatorname{sign}\left(\sum_{j=1}^n t_j f_j(x) \right) \tan \frac{\pi\alpha}{2} \right) m(\mathrm{d}x) \right], \end{aligned}$$

and for $\alpha = 1$

$$\varphi_X(t_1, \dots, t_n)$$
$$= \exp\left[-\int_E \left| \sum_{j=1}^n t_j f_j(x) \right| \left(1 + i \frac{2}{\pi} \beta(x) \operatorname{sign}\left(\sum_{j=1}^n t_j f_j(x) \right) \log \left| \sum_{j=1}^n t_j f_j(x) \right| \right) m(\mathrm{d}x) \right].$$

Again, the distribution of X is symmetric if M is symmetric.

The second part of this proposition implies immediately that it is possible to construct stable random processes via stable integrals, as follows. Let T be an index set and $\{f_t : t \in T\}$ a collection of functions in $L^\alpha(E, m)$, where $0 < \alpha \leq 2$. In the case $\alpha = 1$ we also assume that (2.5) holds for each f_t. If M is an α-stable random measure with control measure m then

$$Z(t) := I(f_t) = \int_E f_t(x)\, M(\mathrm{d}x), \qquad t \in T, \tag{2.6}$$

is an α-stable stochastic process.

The representation of stable processes via stable integrals was studied extensively by several researchers, for example by Kuelbs [62], Rosiński [90], Kanter [55], Schilder [95], and Hardin [48]. A natural problem is to determine which α-stable processes can be represented as stable integrals of the form (2.6). Samorodnitsky and Taqqu [91] found several conditions for this to be true; here, we state only their Proposition 3.11.3 (i) as this covers many situations appearing in spatial statistics.

Proposition 2.13. Let $\{Z(t) : t \in T\}$ be an $S\alpha S$ stochastic process on a separable metric space T. If there is a countable subset $T' \subset T$ such that $\{Z(t) : t \in T \backslash T'\}$ is continuous in probability then Z can be represented as (2.6).

Remark 2.14. (1) In the sequel, we will consider *$S\alpha S$ moving average (MA) processes* Z defined as follows. Let $f \in L^\alpha(\mathbb{R}^d, \lambda)$, where λ is the Lebesgue measure on $\mathbb{R}^d$. If M is an $S\alpha S$ random measure with control measure λ then

$$Z(x) = \int_{\mathbb{R}^d} f(x - y)\, M(\mathrm{d}y), \qquad x \in \mathbb{R}^d. \tag{2.7}$$

Note that $S\alpha S$ MA processes are strictly stationary[4] and that their dependence structure is completely determined by the kernel function f and the stable measure M. For these reasons, we will use $S\alpha S$ MA processes in our simulation studies in Section 2.6 to illustrate features of the DCF.

(2) From the construction of the stable integral via simple functions it follows that a stable integral can be approximated with arbitrary precision by a weighted sum of stable random variables. Hence, it is possible to use the integral representation to simulate stable random variables and also to simulate realizations of stable processes.

Based on the covariation and the codifference given in Definitions 2.5 and 2.7, it is possible to define the following dependence measures for stable random fields, cf. [98, Section 11.2.4] and [91, Section 4.7].

Definition 2.15. Consider the strictly stationary $S\alpha S$ process $\{Z(x) : x \in \mathbb{R}^d\}$.

(a) Let $1 < \alpha \leq 2$. The *covariation function* of Z is defined as $[Z(h), Z(0)]_\alpha$ for $h \in \mathbb{R}^d$.

(b) Let $0 < \alpha \leq 2$. Then the *codifference function* of Z is defined as

$$\tau(h) = 2|||Z(0)|||_\alpha^\alpha - |||Z(h) - Z(0)|||_\alpha^\alpha, \qquad h \in \mathbb{R}^d.$$

2.1.3 Motivating examples

Resnick, Davis and Samorodnitsky, among others, demonstrated in several papers how unreliable empirical estimates for the Pearson correlation function can be in measuring the dependence structure of time series [22, 21, 14, 87, 86]. They found two basic reasons for this. First, for certain ARCH (*autoregressive conditional heteroscedastic*) or ARMA (*autoregressive moving average*) processes, the Pearson sample correlation function is not able to measure non-linear dependencies in a meaningful way. Second, the Pearson sample correlation function may converge to a random limit for some processes with heavy tails.

[4] Let V be a vector space. A random field $\{Z(x) : x \in V\}$ is defined to be *strictly stationary* if for any integer $n \geq 1$ and $h, x_1, \ldots, x_n \in V$ the random vectors $(Z(x_1), \ldots, Z(x_n))^t$ and $(Z(x_1 + h), \ldots, Z(x_n + h))^t$ are identically distributed. In other words, the finite dimensional distributions of Z are invariant with respect to shifts in V.

In the following we illustrate these phenomena in a similar manner for spatial processes and also for certain time series. The aim is to motivate the usefulness in data analysis of the empirical distance correlation function (EDCF), which will be defined later in Subsection 2.2.2. In our examples, we will see that the EDCF yields results matching our intuition and the visible properties of data sets much better than the Pearson sample correlation function.

Remark 2.16. Recall the definition of the *Pearson sample correlation function* $\widehat{\rho}$ for a zero mean, second-order stationary[5] stochastic process $\{Z(x) : x \in \mathbb{Z}^d\}$, which is observed on the grid $\Gamma = \{1, \dots, n\}^d$. Let $h \in \Gamma$ be the lag vector of interest and define $N_\Gamma(h) = \{x \in \Gamma : x + h \in \Gamma\}$. Then

$$\widehat{\rho}(h) = \frac{\sum_{x \in N_\Gamma(h)} Z(x)Z(x+h)}{\sum_{x \in N_\Gamma(h)} Z(x)^2}. \tag{2.8}$$

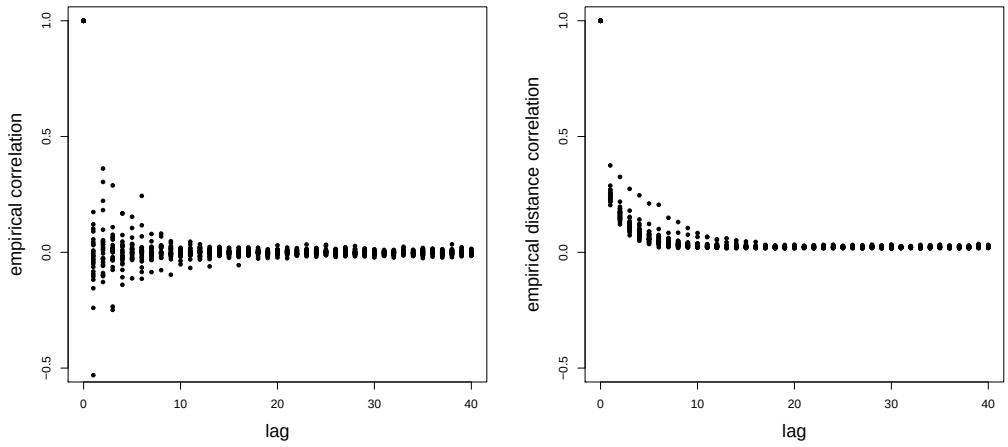
Example 2.17. Consider the following one-dimensional ARCH(1) process, taken from Section 3.2 in [14]. Let $(\varepsilon_n)_{n \geq 1}$ be a sequence of i.i.d. standard normally distributed random variables. Given a random variable Z_0, we define the time series Z_n recursively as

$$Z_n = \varepsilon_n (1 + 0.99 Z_{n-1}^2)^{1/2}, \quad n \geq 1.$$

As Cohen, Resnick, and Samorodnitsky [14] pointed out, only a particular choice of the initial distribution of Z_0 makes this process strictly stationary. In this case, de Haan, et al. [23] proved that the process Z_n exhibits heavy-tailed behaviour, as there exists a constant $c > 0$ such that $P(Z_n > z) \sim cz^{-1.014}$. Cohen, Resnick, and Samorodnitsky [14] dealt with the stationarity issue in their simulations by setting $Z_0 = 0$ and discarding the first $1\,000$ observations, which eliminates the initial transient in the system. We use a similar approach in our simulations: we set $Z_0 = 0$, simulate $11\,000$ observations of Z and discard the first $1\,000$.

Figure 2.1 (a) shows the values of the sample correlation function (2.8) for 25 different realizations of Z. For lags greater than one the values are oscillating randomly around zero. This shows that the Pearson sample correlation function alone cannot explain the dependence structure of the process.

[5] Let V be a vector space. A stochastic process $\{Z(x) : x \in V\}$ is called *second-order stationary* if its mean function is constant, and if $\mathrm{Cov}(Z(x_1), Z(x_2))$ depends only on the lag vector $x_1 - x_2$, for all $x_1, x_2 \in V$. Note that every strictly stationary process with finite second moments is also second-order stationary. Further, every second-order stationary Gaussian process is also strictly stationary.



(a) (b)

Figure 2.1: (a) Values of the empirical Pearson correlation function and (b) the EDCF for 25 different realizations of an ARCH(1) process.

By contrast, Figure 2.1 (b) shows the values of the EDCF for the same realizations, where the EDCF was calculated as in [117]. Its values are oscillating around a rapidly decreasing function, satisfying our expectations on the dependence structure much better than the Pearson sample correlation function. This implies that using the EDCF might lead to a more accurate understanding of the underlying dependence structure than using the empirical Pearson correlation function.

Example 2.18. The example above can be generalized to RFs on $\mathbb{N}^2 = \{1, 2, \dots\}^2$ by using results of Doukhan and Truquet [25]. They gave conditions on a measurable function g, such that

$$Z(j) = g\left(\varepsilon(j-s)_{s\in\mathbb{Z}^d}\right), \tag{2.9}$$

is a strictly stationary autoregressive RF.[6] In particular, it follows from their results

[6] We consider general RFs of type (2.9) in Section 2.5, where we call them transformation fields.

that the RF

$$Z(j) = \varepsilon(j)\Big[1 + Z(j - e_1)/5 - Z(j - e_2)/5 + Z(j - e_1 - e_2)/2\Big], \qquad j \in \mathbb{N}^2, \tag{2.10}$$

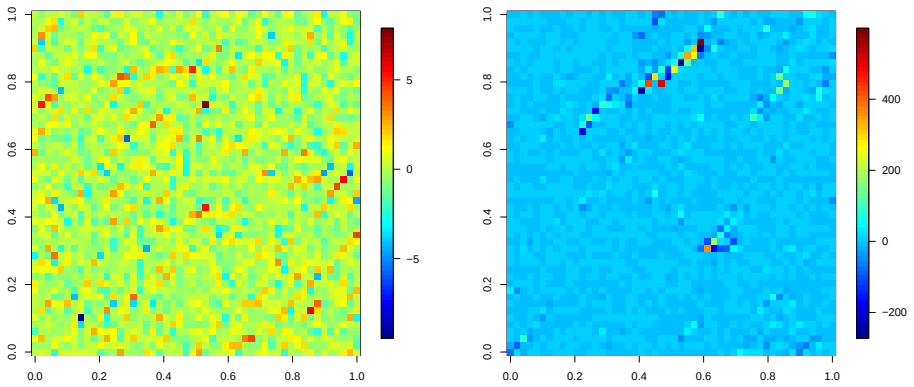
is strictly stationary. Here $\varepsilon(j), j \in \mathbb{Z}^2$, is a collection of i.i.d. random variables, and $e_1 = (1,0)^t, e_2 = (0,1)^t$. For our simulations, we set $Z(0) = Z(e_1) = Z(e_2) = 0$, simulate Z on the grid $\{1,\dots,1050\}^2$ and use the simulated values on $\{1001,\dots,1050\}^2$ as the observation of Z. In what follows, we consider two different distributions for the *innovations* $\varepsilon(j)$.

(1) Let $\varepsilon(j) \sim \mathcal{N}(0,1)$. Figure 2.2 shows a simulated realization of model (2.10) on the grid $\{1,\dots,50\}^2$. We analyze the dependence structure only for the diagonal lag vectors $(0,0)^t, (1,1)^t, \dots, (19,19)^t$ by using the Pearson sample correlation function and the EDCF (see Subsection 2.2.2). Panel (a) in Figure 2.3 shows the Pearson sample correlation function for these lag vectors for 25 different realizations. We see no dependence at all, as the values of the Pearson sample correlation function are very close to zero for all lag vectors different from $(0,0)^t$.

By contrast, the EDCF for the same realizations and lag vectors is significantly greater than zero for the lag vector $(1,1)^t$, as can be seen in panel (b). This indicates strongly that the EDCF is a better tool for detecting dependencies in this case.

(2) Now let $\varepsilon(j)$ be an $S\alpha S$ random variable with $\alpha = 1.5$. Figure 2.2, panel (b) suggests that this model admits stronger dependence for diagonal lag vectors. Again, the Pearson sample correlation function is not able to capture this behaviour, as shown in panel (c) of Figure 2.3. It is clearly visible that the Pearson sample correlation function for the 20 diagonal lag vectors $(0,0)^t, (1,1)^t, \dots, (19,19)^t$ and for 25 different realizations is oscillating randomly around zero.

As opposed to this, the EDCF for the same realizations and lag vectors reflects far better our observations of diagonal dependencies. This can be seen in Figure 2.3 panel (d), where the values for the EDCF decrease to zero for increasing lag vectors.



(a) (b)

Figure 2.2: Realizations of model (2.10) for two different types of innovations. In panel (a), the innovations are standard normally distributed, in panel (b), they are $S\alpha S$ with $\alpha = 1.5$.

2.2 Basic concepts and definitions

The aim of this section is to give a formal definition of the DCF for multivariate stochastic processes and to introduce its empirical version. To this end, we recall in Subsection 2.2.1 the definition of the distance correlation between two random vectors, given in [108], and use it in Subsection 2.2.2 to generalize this concept to distance correlation functions for multivariate RFs. The latter is analogous to the approach of Zhou [117] for multivariate time series.

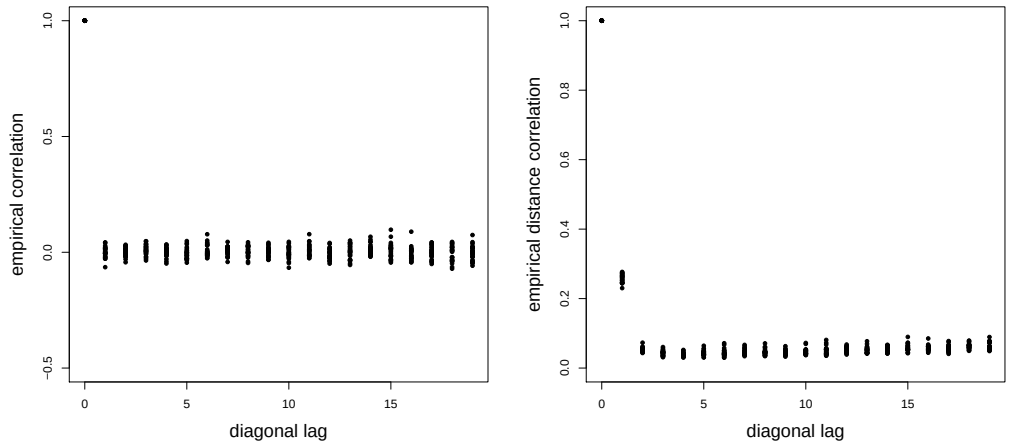
2.2.1 Distance correlation for random vectors

Throughout this subsection, let X and Y be random vectors with finite mean vectors, such that X takes values in $\mathbb{R}^p$ and Y values in $\mathbb{R}^q$, where p, q are positive integers. The characteristic functions for X and Y are given by

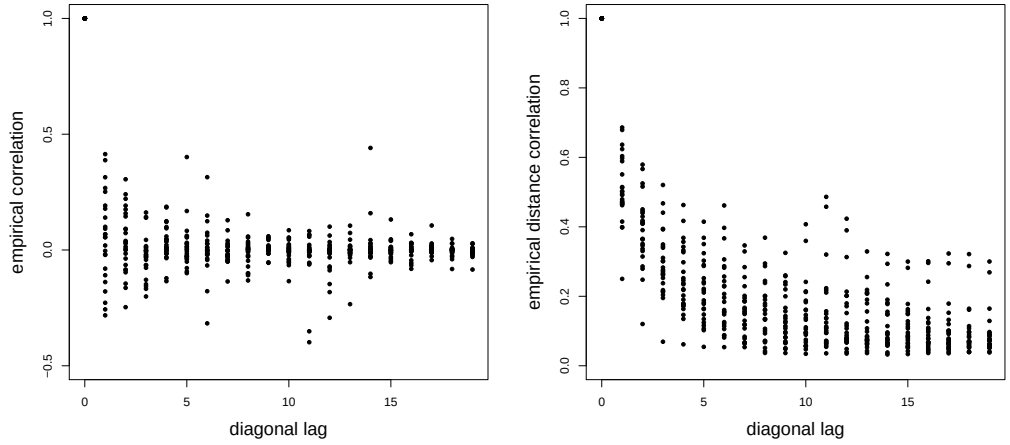
$$\varphi_X(t) = \mathbb{E}\exp(i\langle t, X\rangle), \quad \text{resp.} \quad \varphi_Y(s) = \mathbb{E}\exp(i\langle s, Y\rangle), \quad t \in \mathbb{R}^p, s \in \mathbb{R}^q,$$

respectively, and

$$\varphi_{X,Y}(t,s) = \mathbb{E}\exp\left[i(\langle t, X\rangle + \langle s, Y\rangle)\right],$$



(a) (b)



(c) (d)

Figure 2.3: The Pearson sample correlation function and the EDCF for 25 different realizations of the process (2.10) with two different innovations. Panels (a) and (b) show the Pearson sample correlation function and EDCF for normal innovations, while panels (c) and (d) show the Pearson sample correlation function and EDCF for stable innovations. An integer i on the x-axis corresponds to the diagonal lag vector $(i, i)^t$.

is the characteristic function of the random vector $(X, Y)^t$.

The following definitions are taken from [108].

Definition 2.19. The *distance covariance* $\mathrm{dCov}(X, Y)$ between X and Y is defined via

$$\mathrm{dCov}(X,Y)^2 \;=\; \frac{1}{c_p c_q}\int_{\mathbb{R}^p}\int_{\mathbb{R}^q} \left|\varphi_{X,Y}(t,s) - \varphi_X(t)\varphi_Y(s)\right|^2 \frac{\mathrm{d}t\,\mathrm{d}s}{\|t\|^{p+1}\|s\|^{q+1}}, \quad (2.11)$$

where $c_n = \pi^{(n+1)/2}/\Gamma((n+1)/2)$ is a normalizing constant for integers $n \geq 1$. Analogously to the case of the usual covariance, the *distance variance* of the random vector X is defined as

$$\mathrm{dVar}(X) = \mathrm{dCov}(X, X). \quad (2.12)$$

Further, the *distance correlation* between X and Y is defined as

$$\mathrm{dCorr}(X,Y) = \frac{\mathrm{dCov}(X,Y)}{\sqrt{\mathrm{dVar}(X)\mathrm{dVar}(Y)}}, \quad (2.13)$$

if $\mathrm{dVar}(X)\mathrm{dVar}(Y) \neq 0$, and $\mathrm{dCorr}(X, Y) = 0$ otherwise.

As it is evident from the definition (2.11), the distance covariance measures, in some sense, how strongly the characteristic function $\varphi_{X,Y}$ differs from the product of the marginal characteristic functions $\varphi_X\varphi_Y$. This can be used as a dependence measure, because X and Y are stochastically independent if and only if $\varphi_{X,Y} = \varphi_X\varphi_Y$. However, the distance covariance is not normalized and, hence, not comparable for different pairs of random vectors. By contrast, the distance correlation is appropriately normalized such that $0 \leq \mathrm{dCorr}(X, Y) \leq 1$, with $\mathrm{dCorr}(X, Y) = 0$ if and only if X and Y are stochastically independent [108, Theorem 3]. This indicates that the distance correlation represents a meaningful way to quantitatively assess the dependence between random vectors.

Empirical distance correlation

Suppose we have observed a sample $(\boldsymbol{X}, \boldsymbol{Y}) = \{(X^{(1)}, Y^{(1)}), \ldots, (X^{(n)}, Y^{(n)})\}$ of size n from the random vectors X and Y. We want to approximate empirically the distance covariance. A natural way to define an empirical version of the dis-

tance covariance would be to use (2.11) and substitute the involved characteristic functions by their sample versions

$$\widehat{\varphi}_X^n(t) = \frac{1}{n}\sum_{k=1}^{n} \exp\left(i\langle t, X^{(k)}\rangle\right), \qquad \widehat{\varphi}_Y^n(s) = \frac{1}{n}\sum_{k=1}^{n} \exp\left(i\langle s, Y^{(k)}\rangle\right)$$

and

$$\widehat{\varphi}_{X,Y}^n(t,s) = \frac{1}{n}\sum_{k=1}^{n} \exp\left(i\langle t, X^{(k)}\rangle + i\langle s, Y^{(k)}\rangle\right).$$

The corresponding estimator is denoted as $\widehat{\mathrm{dCov}}_n(\boldsymbol{X}, \boldsymbol{Y})$. This estimator may seem to have the disadvantage that it requires the calculation of an integral, which is only numerically possible in many cases and, hence, makes the estimator unattractive. Interestingly, Theorem 1 in [108] provides the surprising and beautiful result below, showing that $\widehat{\mathrm{dCov}}_n(\boldsymbol{X}, \boldsymbol{Y})$ equals $\mathcal{V}^n(\boldsymbol{X}, \boldsymbol{Y})$, defined as follows.

Definition 2.20. Let $(\boldsymbol{X}, \boldsymbol{Y})^t$ be a sample of size n from $(X, Y)^t$. For $1 \leq k, l \leq n$ define

$$a_{kl} = \|X^{(k)} - X^{(l)}\|, \qquad \overline{a}_{.l} = \frac{1}{n}\sum_{k=1}^{n} a_{kl}, \qquad \overline{a}_{k.} = \frac{1}{n}\sum_{l=1}^{n} a_{kl},$$

and

$$\overline{a}_{..} = \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl} \qquad \text{and} \qquad A_{kl} = a_{kl} - \overline{a}_{.l} - \overline{a}_{k.} + \overline{a}_{..},$$

where $\|\cdot\|$ denotes the Euclidean norm on the corresponding Euclidean space. The quantity B_{kl} is analogously defined using $b_{kl} = \|Y^{(k)} - Y^{(l)}\|$.

Then the estimator $\mathcal{V}^n(\boldsymbol{X}, \boldsymbol{Y})$ is determined via

$$\mathcal{V}^n(\boldsymbol{X}, \boldsymbol{Y})^2 = \frac{1}{n^2}\sum_{k,l=1}^{n} A_{kl}B_{kl}. \tag{2.14}$$

Based on this estimator, an empirical version of the distance variance is obtained as

$$\mathcal{V}^n(\boldsymbol{X})^2 = \mathcal{V}^n(\boldsymbol{X}, \boldsymbol{X})^2 = \frac{1}{n^2}\sum_{k,l=1}^{n} A_{kl}^2,$$

and an empirical version of the distance correlation as

$$\mathcal{R}^n(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\mathcal{V}^n(\boldsymbol{X}, \boldsymbol{Y})}{\sqrt{\mathcal{V}^n(\boldsymbol{X})\mathcal{V}^n(\boldsymbol{Y})}},$$

whenever $\mathcal{V}^n(\boldsymbol{X})\mathcal{V}^n(\boldsymbol{Y}) \neq 0$ – otherwise $\mathcal{R}^n(\boldsymbol{X}, \boldsymbol{Y}) = 0$.[7]

Theorem 2.21. Let $(\boldsymbol{X}, \boldsymbol{Y})$ be an i.i.d. sample of size n of the random vectors X and Y. Then $\widehat{\mathrm{dCov}}_n(\boldsymbol{X}, \boldsymbol{Y}) = \mathcal{V}^n(\boldsymbol{X}, \boldsymbol{Y})$.

Remark 2.22. The following properties of the estimators $\mathcal{V}^n$ and $\mathcal{R}^n$ were shown in [108].

- The empirical distance correlation satisfies $0 \leq \mathcal{R}^n(\boldsymbol{X}, \boldsymbol{Y}) \leq 1$.
- If $\mathcal{R}^n(\boldsymbol{X}, \boldsymbol{Y}) = 1$, then $\boldsymbol{Y}$ is an orthogonal affine version of $\boldsymbol{X}$. In other words, there exist a vector $a \in \mathbb{R}^p$, a real number $b \neq 0$ and an orthogonal matrix $C \in \mathbb{R}^{p \times p}$, such that $\boldsymbol{Y} = a + b\boldsymbol{X}C$.
- Let X and Y be random vectors with finite first moments[8]; and let $(\boldsymbol{X}, \boldsymbol{Y})$ be an i.i.d. sample of X and Y. Then the estimators $\mathcal{V}^n(\boldsymbol{X}, \boldsymbol{Y})$ and $\mathcal{R}^n(\boldsymbol{X}, \boldsymbol{Y})$ are strongly consistent, i.e., it holds almost surely that

$$\lim_{n\to\infty} \mathcal{V}^n(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{dCov}(X, Y) \qquad \text{and} \qquad \lim_{n\to\infty} \mathcal{R}^n(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{dCorr}(X, Y).$$

Remark 2.23. In performing computations, we will use a somewhat simpler form of the empirical distance covariance. With A_{kl} and B_{kl} from Definition 2.20, we have

$$\begin{aligned} A_{kl}B_{kl} =& a_{kl}b_{kl} - a_{kl}\overline{b}_{\cdot l} - a_{kl}\overline{b}_{k\cdot} + a_{kl}\overline{b}_{\cdot\cdot} \\ & - \overline{a}_{\cdot l}b_{kl} + \overline{a}_{\cdot l}\overline{b}_{\cdot l} + \overline{a}_{\cdot l}\overline{b}_{k\cdot} - \overline{a}_{\cdot l}\overline{b}_{\cdot\cdot} \\ & - \overline{a}_{k\cdot}b_{kl} + \overline{a}_{k\cdot}\overline{b}_{\cdot l} + \overline{a}_{k\cdot}\overline{b}_{k\cdot} - \overline{a}_{k\cdot}\overline{b}_{\cdot\cdot} \\ & + \overline{a}_{\cdot\cdot}b_{kl} - \overline{a}_{\cdot\cdot}\overline{b}_{\cdot l} - \overline{a}_{\cdot\cdot}\overline{b}_{k\cdot} + \overline{a}_{\cdot\cdot}\overline{b}_{\cdot\cdot}. \end{aligned}$$

[7] Remark 2 in [108] shows that $\mathcal{V}^n(\boldsymbol{X})\mathcal{V}^n(\boldsymbol{Y}) = 0$ implies that $X^{(1)} = \ldots = X^{(n)}$ or $Y^{(1)} = \ldots = Y^{(n)}$.

[8] We say that an m-variate stochastic process $Z, m \geq 1$, has finite first moments, if for all $x \in \mathbb{R}^d$ all the components of the mean vector $\mathbb{E}(Z(x))$ are finite.

Summation over k and l leads to

$$\sum_{k,l=1}^{n} A_{kl}B_{kl} = \sum_{k,l=1}^{n} a_{kl}b_{kl} - 2n\sum_{l=1}^{n} \overline{a}_{.l}\overline{b}_{.l} + n^2\overline{a}_{..}\overline{b}_{..},$$

where we have used the symmetry of a_{kl} and b_{kl}.

2.2.2 Distance correlation function

We have seen in the preceding subsection that the distance correlation is a measure of the dependence between two random vectors, and which admits a simple consistent estimator. In the next step, we use the distance correlation to define the distance correlation function for univariate and multivariate RFs by generalizing the definition for time series given in [117].

Definition 2.24. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be a univariate, strictly stationary RF with finite first moments. Then the *distance covariance function* of the process Z is

$$\mathrm{C}^{\mathrm{ds}}(h) = \mathrm{dCov}(Z(x), Z(x+h)), \qquad h \in \mathbb{R}^d, \tag{2.15}$$

and the *distance correlation function (DCF)* of Z is

$$\mathrm{R}^{\mathrm{ds}}(h) = \mathrm{dCorr}(Z(x), Z(x+h)), \qquad h \in \mathbb{R}^d. \tag{2.16}$$

Definition 2.25. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be an m-variate, strictly stationary RF with finite first moments. For non-empty subsets $P, Q \subseteq \{1, \ldots, m\}$ with $P = \{i_1, \ldots, i_p\}$ and $Q = \{j_1, \ldots, j_q\}$ define the random processes

$$\{Z_P(x) : x \in \mathbb{R}^d\} = \{(Z_{i_1}(x), \ldots, Z_{i_p}(x))^t : x \in \mathbb{R}^d\}$$

and

$$\{Z_Q(x) : x \in \mathbb{R}^d\} = \{(Z_{j_1}(x), \ldots, Z_{j_q}(x))^t : x \in \mathbb{R}^d\}.$$

Then the *cross-distance covariance function between Z_P and Z_Q* is defined as

$$\mathrm{C}^{\mathrm{ds}}_{P,Q}(h) = \mathrm{dCov}(Z_P(x), Z_Q(x+h)), \qquad h \in \mathbb{R}^d, \tag{2.17}$$

and the *cross-distance correlation function between Z_P and Z_Q* is defined as

$$\mathrm{R}^{\mathrm{ds}}_{P,Q}(h) = \mathrm{dCorr}(Z_P(x), Z_Q(x+h)), \qquad h \in \mathbb{R}^d.$$

Note that these definitions do not depend on $x \in \mathbb{R}^d$, as Z is strictly stationary.

Remark 2.26. One advantage of using the DCF as a dependence measure for RFs is its wide applicability. The DCF is defined for general processes with finite first moments on $\mathbb{R}^d$, including non-Gaussian processes and processes with heavy tails, as well as processes taking continuous or discrete values.

The cross-distance correlation function makes it possible even to analyze the dependence structure of multivariate processes with differing numbers of components. For example, if we have a forecast error field for several weather quantities, we can calculate the cross-distance correlation between the univariate field of precipitation forecast errors with the bivariate field of temperature and wind speed forecast errors.

Given the definition of the DCF, it is interesting to analyze how statistical estimation can be done, because it is clear that finding consistent estimators is more complicated than in the case of i.i.d. samples of pairs of random vectors. Recently, Zhou [117] found consistent estimators for the DCF in the time series case and we will use his approach in the following. Therefore, we restrict ourselves to the estimation of $\mathrm{C}^{\mathrm{ds}}(h)$ in the case where the corresponding process Z is observed on a regular grid $\Gamma \subset \mathbb{R}^d$. For all $h \in \mathbb{R}^d$, define $N_\Gamma(h) = \{x \in \Gamma : x + h \in \Gamma\}$ to be the set of all points in Γ that remain in Γ after being shifted by h.

Definition 2.27. Let $P, Q \subseteq \{1, \dots, m\}$ be non-empty and

$$(\boldsymbol{Z}_P, \boldsymbol{Z}'_Q) = \left\{ \big(Z_P(x), Z_Q(x+h)\big)^t : x \in N_\Gamma(h) \right\},$$

where Z_P, Z_Q are given as in Definition 2.25. Here we make a slight abuse of notation, as $Z_P(x)$ refers to the observed value at the point x, and similarly for Z_Q.

Then the empirical cross-distance covariance function and the empirical cross-distance correlation function are defined as

$$\mathcal{V}^\Gamma_{P,Q}(h) = \mathcal{V}^n(\boldsymbol{Z}_P, \boldsymbol{Z}'_Q) \qquad \text{and} \qquad \mathcal{R}^\Gamma_{P,Q}(h) = \mathcal{R}^n(\boldsymbol{Z}_P, \boldsymbol{Z}'_Q), \tag{2.18}$$

where $n = |N_\Gamma(h)|$. If $P = Q = \{1, \dots, m\}$, we drop the subscripts and write $\mathcal{V}^\Gamma(h)$ and $\mathcal{R}^\Gamma(h)$ and call them the *empirical distance covariance function* and *empirical distance correlation function (EDCF)* , respectively.

Remark 2.28. Using Definition 2.20, we see that the explicit form of the empirical

cross-distance covariance function equals

$$\mathcal{V}^{\Gamma}_{P,Q}(h)^2 = \frac{1}{|N_\Gamma(h)|^2} \sum_{r,l \in N_\Gamma(h)} A_{rl} B_{rl}.$$

Here we have

$$A_{rl} = a_{rl} - \frac{1}{|N_\Gamma(h)|} \sum_{l \in N_\Gamma(h)} a_{rl} - \frac{1}{|N_\Gamma(h)|} \sum_{r \in N_\Gamma(h)} a_{rl} + \frac{1}{|N_\Gamma(h)|^2} \sum_{l,r \in N_\Gamma(h)} a_{rl},$$

where $a_{rl} = \|Z_P(r) - Z_P(l)\|$. The quantity B_{rl} is defined analogously with $b_{rl} = \|Z'_Q(r) - Z'_Q(l)\|$, where the process $Z'_Q(j)$ is the shifted process $Z'_Q(x) = Z_Q(x+h)$, for $x \in N_\Gamma(h)$.

As mentioned earlier, Theorem 1 in [108] shows that we can also write

$$\mathcal{V}^{\Gamma}_{P,Q}(h)^2 = \frac{1}{c_p c_q} \int_{\mathbb{R}^p} \int_{\mathbb{R}^q} \frac{|\widehat{\varphi}^{\Gamma}_h(t,s) - \widehat{\varphi}^{\Gamma}(t)\widehat{\varphi}^{\Gamma}_h(s)|^2}{\|t\|^{p+1}\|s\|^{q+1}} \,\mathrm{d}t\,\mathrm{d}s,$$

with the empirical characteristic functions

$$\begin{aligned}
\widehat{\varphi}^{\Gamma}(t) &= \frac{1}{|N_\Gamma(h)|} \sum_{x \in N_\Gamma(h)} \exp(i\langle t, Z(x)\rangle), \\
\widehat{\varphi}^{\Gamma}_h(s) &= \frac{1}{|N_\Gamma(h)|} \sum_{x \in N_\Gamma(h)} \exp(i\langle s, Z(x+h)\rangle), \qquad \text{and} \\
\widehat{\varphi}^{\Gamma}_h(t,s) &= \frac{1}{|N_\Gamma(h)|} \sum_{x \in N_\Gamma(h)} \exp\big(i\langle t, Z(x)\rangle + i\langle s, Z(x+h)\rangle\big).
\end{aligned} \tag{2.19}$$

Example 2.29. In this example we consider the DCF for stable MA processes. We will use the DCF in Section 2.6.3 to analyze the finite sample performance of the EDCF.

Let $\mathbb{B}^d$ be the Borel-σ-algebra of $\mathbb{R}^d$ and λ the d-dimensional Lebesgue measure. Assume that

$$Z(x) = \int_{\mathbb{R}^d} f(x-y)\, M(\mathrm{d}y), \qquad x \in \mathbb{R}^d,$$

is an $S\alpha S$ MA process with respect to an $S\alpha S$ random measure M with control measure λ. Here $f \in L^\alpha(\mathbb{R}^d, \lambda)$ and $0 < \alpha \leq 2$.

The process Z is strictly stationary, so the DCF is completely determined by

$\varphi_{Z(0)}(t)$ and $\varphi_{Z(0),Z(h)}(t,s)$. By Proposition 2.12, these characteristic functions are calculated as $\varphi_{Z(0)}(s) = \exp(-\sigma^\alpha |s|^\alpha)$, where $\sigma = (\int_{\mathbb{R}^d} |f(x)|^\alpha \,\mathrm{d}x)^{1/\alpha}$,

$$\varphi_{Z(h),Z(0)}(t,s) = \exp\left(-\int_{\mathbb{R}^d} |t\, f(h+x) + s\, f(x)|^\alpha \,\mathrm{d}x\right), \tag{2.20}$$

for $\alpha \neq 1$, and

$$\varphi_{Z(h),Z(0)}(t,s) = \exp\left(-\int_{\mathbb{R}^d} |t\, f(h+x) + s\, f(x)| \log |t\, f(h+x) + s\, f(x)| \,\mathrm{d}x\right),$$

for $\alpha = 1$. These integrals can be evaluated numerically or analytically, depending on the kernel f. The calculation of the EDCF then involves one further numerical integration, which might be non-trivial. In particular, if $\alpha > 1$ is close to one, the convergence of this numerical integration becomes very slow, as we will illustrate in Section 2.6.3.

2.3 Properties of the distance correlation function

One property of the Pearson correlation function which makes it a powerful tool in many applications is its characterization as a positive definite function. Finding a similar characterization for DCFs is a very deep problem, as general properties of the integral in (2.11) are difficult to obtain. This task may become simpler when we restrict the problem to certain classes of distributions (such as the Gaussian distributions, see subsection 2.3.3), but as the involved integral is analytically solvable only in some cases, the problem remains difficult. Moreover, known analytical expressions are also complicated, see the work of Dueck, et al. [27] for the multivariate Gaussian case and Dueck, Edelmann, and Richards [26] for the case of random vectors with Lancaster distributions. Unfortunately, in our particular case of interest, the α-stable distributions, no analytical expressions are known for the distance correlation.

This illustrates the problem of deriving theoretical characterizations or properties of the DCF. Nevertheless, we derive in this section results which can be obtained without the need to characterize or find closed-form expressions for the DCF.

2.3.1 Evenness of the distance correlation function

As in the case of the Pearson correlation function, the DCF is even if the underlying RF is strictly stationary.

Theorem 2.30. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be an m-variate, strictly stationary centered RF and $P \subseteq \{1, \dots, m\}$ non-empty. Then

$$\mathrm{R}^{\mathrm{ds}}_{P,P}(h) = \mathrm{R}^{\mathrm{ds}}_{P,P}(-h),$$

for all $h \in \mathbb{R}^d$.

Proof. Let $P = \{i_1, \dots, i_p\}$ and $Z_P(x) = (Z_{i_1}(x), \dots, Z_{i_p}(x))^t$, for all $x \in \mathbb{R}^d$. Using (2.15) and (2.11) yields for $h \in \mathbb{R}^d$,

$$\begin{aligned}&\Big[\mathrm{C}^{\mathrm{ds}}(-h)\Big]^2 \\ &= \frac{1}{c_p^2} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \left|\varphi_{Z_P(-h),Z_P(0)}(t,s) - \varphi_{Z_P(-h)}(t)\varphi_{Z_P(0)}(s)\right|^2 \frac{\mathrm{d}t\,\mathrm{d}s}{\|t\|^{p+1}\|s\|^{p+1}}. \quad (2.21)\end{aligned}$$

The strict stationarity of Z implies $\varphi_{Z_P(-h)}(t) = \varphi_{Z_P(0)}(t)$ as well as $\varphi_{Z_P(0)}(s) = \varphi_{Z_P(h)}(s)$. This, together with

$$(Z_P(-h), Z_P(0))^t \overset{d}{\sim} (Z_P(0), Z_P(h))^t,$$

implies that

$$\varphi_{Z_P(-h),Z_P(0)}(t,s) = \varphi_{Z_P(0),Z_P(h)}(t,s) = \varphi_{Z_P(h),Z_P(0)}(s,t),$$

where the last equality is a consequence from the definition of the characteristic function. Substituting these results into (2.21) yields the evenness of the distance covariance function and, hence, of the DCF. □

Remark 2.31. Under the assumptions of the foregoing theorem, if $P, Q \subseteq \{1, \dots, m\}$ with $P \neq Q$ then $\mathrm{C}^{\mathrm{ds}}_{P,Q}$ is in general not even, which is illustrated by the following example.

Let $\{Y(x) : x \in \mathbb{R}\}$ be a bivariate, second-order stationary, centered Gaussian process with *cross-covariance function*

$$\mathsf{C}_Y(h) := \mathbb{E}\left(Y(h)^t \cdot Y(0)\right) = \begin{pmatrix} C_{11}(h) & C_{12}(h) \\ C_{21}(h) & C_{22}(h) \end{pmatrix}, \qquad h \in \mathbb{R},$$

where C follows a *parsimonious bivariate Matérn model* introduced by Gneiting,

Kleiber, and Schlather [43]. In particular, let

$$C_{11}(h) = C_{22}(h) = \frac{1}{8}|h|^3 K_3(|h|),$$

and

$$C_{12}(h) = C_{21}(h) = \frac{1}{16}|h|^3 K_3(|h|),$$

where K_3 is a modified Bessel function of the second kind with parameter 3, cf. [79, Section 10.25]. Theorem 3 in [43] shows that C is a valid cross-correlation function. We define the bivariate process Z as

$$Z(x) = (Z_1(x), Z_2(x))^t = (Y_1(x), Y_2(x-1))^t, \qquad x \in \mathbb{R},$$

which is a Gaussian process, too. Following Li and Zhang [65], Z is second-order stationary and its cross-correlation function equals

$$\mathsf{C}_Z(h) = \begin{pmatrix} C_{11}(h) & C_{12}(h-1) \\ C_{21}(h+1) & C_{22}(h) \end{pmatrix}, \qquad h \in \mathbb{R}.$$

Now let $P = \{1\}$ and $Q = \{2\}$. Székely, Rizzo, and Bakirov [108] showed that the distance correlation of two jointly Gaussian random variables is determined by their Pearson correlation coefficient, see also Lemma 2.34 in Section 2.3.3 below. Part (ii) of that lemma implies that two different non-negative Pearson correlation coefficients yield two different distance correlation values. Then $1/2 = C_{12}(0) \neq C_{12}(-2) \approx 0.32$ implies $\mathrm{C}^{\mathrm{ds}}_{P,Q}(1) \neq \mathrm{C}^{\mathrm{ds}}_{P,Q}(-1)$, as $\mathrm{C}^{\mathrm{ds}}_{P,Q}(1)$ is determined by $C_{12}(0)$ and $\mathrm{C}^{\mathrm{ds}}_{P,Q}(-1)$ by $C_{12}(-2)$.

2.3.2 Isotropy of the distance correlation function

An RF $\{Z(x) : x \in \mathbb{R}^d\}$ is called *strictly isotropic* if its finite-dimensional distributions are rotation invariant, i.e., if for all rotation matrices $O \in \mathrm{SO}(d)$[9] and all $x_1, \ldots, x_n \in \mathbb{R}^d$ we have

$$(Z(x_1), \ldots, Z(x_n))^t \overset{d}{\sim} (Z(Ox_1), \ldots, Z(Ox_n))^t.$$

In the context of covariance functions, a weaker concept of isotropy is often used, which is called *isotropy in the wide sense* or *isotropy in second order*: an RF $\{Z(x) :$

[9] $\mathrm{SO(d)}$ denotes the *special orthogonal group* of $\mathbb{R}^d$, i.e., the group of all orthogonal matrices in $\mathbb{R}^{d\times d}$ with determinant equal to one.

$x \in \mathbb{R}^d\}$ is called isotropic in the wide sense, if for all pairs $x_1, x_2 \in \mathbb{R}^d$ and rotation matrices $O \in \mathrm{SO}(d)$ we have

$$\mathbb{E}(Z(x_1)) = \mathbb{E}(Z(Ox_1)) \qquad \text{and} \qquad \mathrm{Cov}(Z(x_1), Z(x_2)) = \mathrm{Cov}(Z(Ox_1), Z(Ox_2)).$$

It is clear that every strictly isotropic RF is also isotropic in the wide sense.

A slightly modified concept of wide-sense isotropy can be defined in the context of DCFs.

Definition 2.32. A RF $\{Z(x) : x \in \mathbb{R}^d\}$ is called *pairwise isotropic*, if

$$(Z(x_1), Z(x_2))^t \overset{d}{\sim} (Z(Ox_1), Z(Ox_2))^t$$

for all pairs of points $x_1, x_2 \in \mathbb{R}^d$ and all rotation matrices $O \in \mathrm{SO}(d)$.

Obviously, every strictly isotropic field is pairwise isotropic and every pairwise isotropic field is isotropic in the wide sense. Further, in the case of Gaussian RFs, every wide-sense isotropic field is also strictly isotropic.

It is evident that pairwise isotropy is the weakest type of isotropy which implies the DCF to be isotropic.

Remark 2.33. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be a strictly stationary centered RF. If Z is pairwise isotropic then its distance covariance function is isotropic, i.e.,

$$\mathrm{C}^{\mathrm{ds}}(h) = \mathrm{C}^{\mathrm{ds}}(Oh)$$

for all $O \in \mathrm{SO}(d)$ and $h \in \mathbb{R}^d$.

2.3.3 Positive definiteness in the Gaussian case

As mentioned at the beginning of this section, it is difficult to find a useful characterization of the class of admissible distance correlation functions for general RFs. However, in the case of Gaussian RFs, it is possible to find such a characterization, which relies on the following connection between the Pearson correlation coefficient and distance correlation for the case of bivariate Gaussian random vectors, given in [108].

Lemma 2.34. Let $(X, Y)^t$ be a bivariate, centered, normally distributed random

vector with Pearson correlation coefficient $\rho = \rho(X, Y)$ and variances $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$. Then the distance correlation $\mathrm{dCorr}(X, Y)$ satisfies

(i) $|\mathrm{dCorr}(X, Y)| \leq |\rho|$, and

(ii) $\mathrm{dCorr}^2(X, Y) = \xi^2(\rho)$, where

$$\xi^2(\rho) = \frac{\rho \arcsin \rho - \rho \arcsin(\rho/2) + \sqrt{1-\rho^2} - \sqrt{4-\rho^2} + 1}{1 + \pi/3 - \sqrt{3}}.$$

We claim that ξ^2 is a positive definite function on $[-1, 1]$. To this end, we use the following series expansion, proven in [26] (Proposition 5.1)

$$\zeta^2(\rho) := \mathrm{dCov}(X, Y)^2 = \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{((2k-3)!!)^2}{(2k)!} \left(1 - 2^{-(2k-1)}\right) \rho^{2k},$$

where

$$n!! = \begin{cases} n(n-2)(n-4)\cdots 2, & \text{if } n \in \mathbb{N} \text{ is even,} \\ n(n-2)(n-4)\cdots 1, & \text{if } n \in \mathbb{N} \text{ is odd,} \\ 1, & \text{if } n = -1, \text{ or } n = 0, \end{cases} \tag{2.22}$$

is the *double factorial*.

Now all coefficients in the series expansion of ζ^2 – and, hence, in the series expansion of ξ^2 – are non-negative, implying that ξ^2 is the pointwise limit of a sequence of positive definite functions (compare [9, Section 3.1]). Consequently, in the case of a Gaussian RF, the square of the DCF can be characterized as the positive definite function $\xi^2 \circ C$, where C is positive definite.

2.3.4 Smoothness of the distance correlation function

The association between smoothness of the Pearson correlation function and the mean-square smoothness of a RF $\{Z(x) : x \in \mathbb{R}^d\}$ with finite second moments is well understood. This motivates us to examine if there is a similar relation between smoothness properties of the DCF and the RF in question.

To this end, we recall the definition of mean-square continuity of a process (see Section 1.4 in [116]).

Definition 2.35. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be an m-variate, centered stochastic process, that is not necessarily strictly stationary. For $r > 0$, the process Z is called *continuous in r-th mean* at $x_0 \in \mathbb{R}^d$ if

$$\lim_{x \to x_0} \mathbb{E}\|Z(x_0) - Z(x)\|^r = 0.$$

If this holds for all $x_0 \in \mathbb{R}^d$, the process Z is called continuous in mean of order r.

In particular, Z is called *mean continuous*, resp. *mean-square continuous* if $r = 1$, resp. $r = 2$.

Remark 2.36. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be a univariate, centered RF with finite second moments. Then the continuity of

$$C(x, y) = \operatorname{Cov}(Z(x), Z(y)), \qquad x, y \in \mathbb{R}^d,$$

on the diagonal of $\mathbb{R}^d \times \mathbb{R}^d$ is sufficient and necessary for Z to be mean-square continuous, cf. [1, Theorem 2.1].

Now let Z be a strictly stationary centered process. In the following, we will show that mean continuity of Z and continuity of the DCF are indeed related, but the relation is weaker than in the case of the Pearson correlation function.

Theorem 2.37. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be an m-variate, strictly stationary centered process. If Z is mean continuous then the distance covariance function $\mathrm{C}^{\mathrm{ds}}(h)$ is continuous in zero.

Proof. Define

$$\mathrm{d}w(t, s) = (c_m^2 \|t\|^{m+1} \|s\|^{m+1})^{-1} \,\mathrm{d}t \,\mathrm{d}s$$

and the function $f : \mathbb{R}^{2m} \times \mathbb{R}^d \to \mathbb{R}$ via

$$f(t, s, h) = \left|\varphi_{Z(h),Z(0)}(t, s) - \varphi_{Z(h)}(t)\varphi_{Z(0)}(s)\right|^2.$$

Then, $\mathrm{C}^{\mathrm{ds}}(h) = \int_{\mathbb{R}^{2m}} f(t, s, h) \,\mathrm{d}w(t, s)$. First, we claim that

$$\lim_{h \to 0} \int_{\mathbb{R}^{2m}} f(t, s, h) \,\mathrm{d}w(t, s) = \int_{\mathbb{R}^{2m}} \lim_{h \to 0} f(t, s, h) \,\mathrm{d}w(t, s), \tag{2.23}$$

i.e., it is admissible to interchange the limit and the integral. For this, it is sufficient

to show that

(i) the function $(t,s) \mapsto \sup_{h\in\mathbb{R}^d} |f(t,s,h)|$ is integrable with respect to $\mathrm{d}w(t,s)$, and

(ii) the function $h \mapsto f(t,s,h)$ is continuous in $h=0$, for almost all $(t,s) \in \mathbb{R}^{2m}$.

(Compare for example Theorem 6.27 in [60]). For item (i), we have

$$f(t,s,h) \leq \left(1-|\varphi_{Z(h)}(t)|^2\right)\left(1-|\varphi_{Z(0)}(0)|^2\right),$$

which follows from the Cauchy inequality, cf. [108]. As Z is strictly stationary, the above inequality implies

$$f(t,s,h) \leq \left(1-|\varphi_{Z(0)}(t)|^2\right)\left(1-|\varphi_{Z(0)}(s)|^2\right)$$

and, hence,

$$\sup_{h\in\mathbb{R}^d} |f(t,s,h)| \leq \left(1-|\varphi_{Z(0)}(t)|^2\right)\left(1-|\varphi_{Z(0)}(s)|^2\right).$$

The right-hand side of this inequality is integrable with respect to $\mathrm{d}w(t,s)$, which is a consequence of Lemma 1 in [108], Fubini's Theorem and the existence of the first moments of Z (cf. [108, (2.5)]). Therefore, (i) holds.

As Z is strictly stationary, we have $\varphi_{Z(h)}(t) = \varphi_{Z(0)}(t)$. To show (ii), it remains to prove that

$$\lim_{h\to 0} \varphi_{Z(h),Z(0)}(t,s) = \varphi_{Z(0),Z(0)}(t,s). \tag{2.24}$$

But this holds true for all $t,s \in \mathbb{R}^m$, because of the continuity theorem (for example, Theorem 3 (e) in [31]) and the fact that

$$(Z(h),Z(0))^t \xrightarrow{d} (Z(0),Z(0))^t, \qquad \|h\| \to 0,$$

since Z is mean continuous. □

Remark 2.38. Note that the mean continuity of Z in the above theorem is not a necessary assumption. It suffices if Z is *continuous in distribution*, i.e.,

$$(Z(h),Z(0))^t \xrightarrow{d} (Z(0),Z(0))^t, \qquad \|h\| \to 0,$$

which does not imply mean continuity in general. As continuity in distribution is a weak type of continuity for RFs, it is unclear whether there are cases in which continuity of the DCF implies a.s. continuity of the sample paths of Z. Such a connection would be interesting, as the continuity of sample paths might represent important physical properties in applications.

Remark 2.39. For univariate, second-order stationary Gaussian RFs Z on $\mathbb{R}^d$ with covariance function C we have

$$\mathrm{R}^{\mathrm{ds}}(h) = \xi(C(h)), \qquad h \in \mathbb{R}^d,$$

where ξ is the function defined in Lemma 2.34. As ξ is continuous, the continuity of $\mathrm{R}^{\mathrm{ds}}(h)$ is equivalent to the continuity of $C(h)$. This underlines again that in the case of Gaussian processes, the distance correlation provides almost the same information about the process as the Pearson correlation. Note that it cannot provide exactly the same information, as the DCF is always non-negative.

2.4 Distance variogram

In addition to the dependence measures mentioned in Subsection 2.1.1, there is another interesting one for stable processes, introduced by Astrauskas, Lévy, and Taqqu [6]. Their aim was to quantify dependence for certain univariate stable time series $\{Z(x) : x \in \mathbb{R}\}$ with strictly stationary increments (so-called *linear fractional Lévy motions*). To that end, they considered the increment process

$$Y(x) = Z(x+1) - Z(x), \qquad x \in \mathbb{R}, \tag{2.25}$$

and studied the behaviour of the function

$$r(t,s,x) = \mathbb{E}\left[\exp(itY(x) + isY(0))\right] - \mathbb{E}[\exp(itY(x))]\mathbb{E}[\exp(isY(0))],$$

for $t, s \in \mathbb{R}$. Note that this function corresponds to the kernel of the distance covariance between the random variables $Y(x)$ and $Y(0)$. Unlike the time series case, there is no distinguished lag vector for spatial processes $\{Z(x) : x \in \mathbb{R}^d\}$ with strictly stationary increments. Hence, we cannot define canonically an increment process that corresponds to Y given in (2.25), nor can we study its DCF in the spatial case.

Nevertheless, to circumvent this problem we consider the general spatial incre-

ment process $Y(x,h) = Z(x+h) - Z(x)$ for lags $h \in \mathbb{R}^d$, and study its *distance variance*. This approach is similar to the construction of the usual variogram γ, which is defined as one-half of the usual variance of the increment process $Y(x,h)$, i.e.,

$$\gamma(h) = \frac{1}{2}\mathrm{Var}(Y(x,h)), \qquad x,h \in \mathbb{R}^d,$$

provided that Z is *intrinsically stationary*, i.e., $\mathbb{E}(Y(x,h)) = 0$ for all $x,h \in \mathbb{R}^d$ and $\mathrm{Var}(Y(x,h))$ is independent of $x \in \mathbb{R}^d$, for all $h \in \mathbb{R}^d$.[10] We shall call the new dependence measure the *distance variogram*.

2.4.1 Definition

As with the usual variogram, the distance variogram can be defined with a weaker assumption than strict stationarity of the process because the definition requires only that the increment process is strictly stationary.

Definition 2.40. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be an m-variate RF with constant mean and let $h \in \mathbb{R}^d$. Then Z is called *intrinsically strictly stationary* if the distribution of the increment process

$$Y(x,h) = Z(h+x) - Z(x), \qquad h \in \mathbb{R}^d,$$

is independent of x.

If $\{Z(x) : x \in \mathbb{R}^d\}$ is an intrinsically strictly stationary RF, we omit the dependence on x and write $Y(h)$ for $Y(x,h)$ whenever this causes no confusion.

Remark 2.41. It is evident that every strictly stationary process is also intrinsically strictly stationary, but the opposite is not true. For example, consider the (Lévy) fractional isotropic Brownian motion $\{B(x) : x \in \mathbb{R}^d\}$ [16, Section 5.4], with covariance function

$$\mathrm{Cov}(B(x), B(y)) = \|x\| + \|y\| - \|x-y\|, \qquad x,y \in \mathbb{R}^d.$$

This implies directly that B is not strictly stationary. Its variogram γ is given by $\gamma(h) = \|h\|$ for each $h \in \mathbb{R}^d$, showing that the process is an intrinsically stationary

[10] Note that an intrinsically stationary process need not be strictly stationary, but every strictly stationary process is intrinsically stationary.

process. Moreover, as B is Gaussian, it is also intrinsically strictly stationary.

Definition 2.42. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be a univariate, intrinsically strictly stationary RF and let $\{Y(h) : h \in \mathbb{R}^d\}$ be the corresponding increment process. Then the *distance variogram* is defined as

$$\mathrm{V}^{\mathrm{ds}}(h) = \mathrm{dVar}(Y(h)), \qquad h \in \mathbb{R}^d,$$

whenever this distance variance exists.

Analogous to the cross-distance correlation function, it is possible to generalize the distance variogram to multivariate processes.

Definition 2.43. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be an m-variate, intrinsically strictly stationary RF and let $P, Q \subseteq \{1, \dots, m\}$ be non-empty, with $P = \{i_1, \dots, i_p\}$ and $Q = \{j_1, \dots, j_q\}$. For processes Z_P and Z_Q given as in Definition 2.25, we consider the increment processes

$$Y_P(h) = Z_P(h) - Z_P(0) \quad \text{and} \quad Y_Q(h) = Z_Q(h) - Z_Q(0), \qquad h \in \mathbb{R}^d.$$

Then the *cross-distance variogram between Z_P and Z_Q* is defined as

$$\mathrm{V}^{\mathrm{ds}}_{P,Q}(h) = \mathrm{dCov}(Y_P(h), Y_Q(h)), \qquad h \in \mathbb{R}^d,$$

provided it exists.

A sufficient condition for the existence of the (cross-) distance variogram is the finiteness of the first moments of Z, which certainly is not a necessary condition: If $Z(x) \equiv Z$ for some random variable Z with infinite or undefined first moments,[11] then $Z(x) - Z(y) = 0$ for all $x, y \in \mathbb{R}^d$, implying the existence of the distance variogram.

In the case of random processes with infinite first moments, the finiteness of the distance variogram is not guaranteed. However, in cases of strictly stationary $S\alpha S$ processes, it is possible to define a normalized distance variogram which exists for all $0 < \alpha \leq 2$, see the following subsection.

[11] An example for a distribution with infinite mean is any *Pareto distribution* with shape parameter $\alpha \leq 1$; an example for a distribution with undefined mean is the Cauchy distribution.

Remark 2.44. Consider a univariate, centered and second-order stationary Gaussian RF $\{Z(x) : x \in \mathbb{R}^d\}$ with covariance function $C(h)$ and $C(0) = 1$. Its increment process fulfills

$$Y(h) = Z(h) - Z(0) \overset{d}{\sim} \mathcal{N}(0, 2\gamma(h)),$$

where $\gamma(h) = 1 - C(h)$ denotes the usual variogram for intrinsically stationary processes [38]. With this we can calculate the distance variogram via

$$\mathrm{V}^{\mathrm{ds}}(h)^2 = \frac{1}{\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left(e^{-2(t+s)^2\gamma(h)^2} - e^{-2t^2\gamma(h)^2} e^{-2s^2\gamma(h)^2} \right)^2 \frac{\mathrm{d}t}{t^2} \frac{\mathrm{d}s}{s^2},$$

which equals zero if $\gamma(h) = 0$. Otherwise, if $\gamma(h) \neq 0$ then we replace (t, s) by $(\sqrt{2}\gamma(h)t, \sqrt{2}\gamma(h)s)$ and obtain

$$\mathrm{V}^{\mathrm{ds}}(h)^2 = \kappa^2 \gamma(h)^2,$$

with

$$\kappa^2 = \frac{2}{\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left(e^{-(t+s)^2/2} - e^{-t^2/2} e^{-s^2/2} \right)^2 \frac{\mathrm{d}t}{t^2} \frac{\mathrm{d}s}{s^2}$$

being a positive constant. Hence, the distance variogram is proportional to the usual variogram.

Note that κ^2 can be calculated explicitly: It is straightforward to show that

$$\kappa^2 = 4\,\mathrm{dVar}(X)^2,$$

where X is a standard normal random variable. The proof of Theorem 7 (ii) in [108] implies

$$\mathrm{dVar}(X)^2 = \frac{4}{\pi} \left(1 + \frac{\pi}{3} - \sqrt{3} \right),$$

such that $\kappa^2 \approx 1.605$.

2.4.2 The distance variogram for $S\alpha S$-processes

The Gaussian case in Remark 2.44 indicates that the distance variogram might be a generalization of the usual variogram. Indeed, our investigations of the distance variogram for intrinsically strictly stationary $S\alpha S$ processes will confirm this assumption. Additionally, we will show that the distance variogram for this class of processes possesses further interesting properties which are similar to the properties of the usual variogram.

Theorem 2.45. Let Z be a univariate, intrinsically strictly stationary $S\alpha S$ RF on $\mathbb{R}^d$ with $1 < \alpha \leq 2$. Then the distance variogram exists and equals

$$\mathrm{V}^{\mathrm{ds}}(h) = \kappa(\alpha) \| Z(h) - Z(0) \|_\alpha,$$

where

$$\kappa(\alpha) = \frac{1}{\pi} \left(\int_{\mathbb{R}} \int_{\mathbb{R}} \left\{ \exp\left[-|t+s|^\alpha \right] - \exp\left[-(|t|^\alpha + |s|^\alpha) \right] \right\}^2 \frac{\mathrm{d}t}{t^2} \frac{\mathrm{d}s}{s^2} \right)^{1/2},$$

and $\|\cdot\|_\alpha$ denotes the covariation norm, cf. Definition 2.5.

Proof. Since $(Z(h), Z(0))^t$ is a bivariate $S\alpha S$ random vector, Theorem 2.2 yields the characteristic function

$$\begin{aligned} \mathbb{E} \exp[itZ(h) - itZ(0)] &= \exp\left[-\int_{\mathbb{S}^2} |tu_1 - tu_2|^\alpha \, \Lambda_h(\mathrm{d}u) \right] \\ &= \exp\left[-|t|^\alpha \int_{\mathbb{S}^2} |u_1 - u_2|^\alpha \, \Lambda_h(\mathrm{d}u) \right], \end{aligned}$$

where Λ_h is the spectral measure of the random vector $(Z(h), Z(0))^t$. The last equality implies immediately that the scale parameter of $Z(h) - Z(0)$ is given by

$$\| Z(h) - Z(0) \|_\alpha^\alpha = \int_{\mathbb{S}^2} |u_1 - u_2|^\alpha \, \Lambda_h(\mathrm{d}u).$$

Consequently, we have

$$\begin{aligned} \mathrm{V}^{\mathrm{ds}}(h)^2 = \frac{1}{\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \Big\{ &\exp\left[-|t+s|^\alpha \| Z(h) - Z(0) \|_\alpha^\alpha \right] \\ &- \exp\left[-(|t|^\alpha + |s|^\alpha) \| Z(h) - Z(0) \|_\alpha^\alpha \right] \Big\}^2 \frac{\mathrm{d}t}{t^2} \frac{\mathrm{d}s}{s^2}. \end{aligned}$$

The change of variables

$$t \rightsquigarrow t \| Z(h) - Z(0) \|_\alpha \qquad \text{and} \qquad s \rightsquigarrow s \| Z(h) - Z(0) \|_\alpha$$

finally gives

$$\mathrm{V}^{\mathrm{ds}}(h)^2 = \kappa(\alpha)^2 \| Z(h) - Z(0) \|_\alpha^2,$$

with

$$\kappa(\alpha)^2 = \frac{1}{\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left\{ \exp\left[-|t+s|^{\alpha} \right] - \exp\left[-(|t|^{\alpha} + |s|^{\alpha}) \right] \right\}^2 \frac{\mathrm{d}t}{t^2} \frac{\mathrm{d}s}{s^2}. \tag{2.26}$$

□

Remark 2.46. If $\alpha = 2$, the number $\kappa(2)$ given by (2.26) equals the constant κ from Remark 2.44. This demonstrates that the distance variogram is a natural generalization of the usual variogram in the case of $S\alpha S$ RFs.

A strength of the result above lies in the fact that $\kappa(\alpha)$ does not depend on h but on α only. This implies that the result might be useful for investigating connections between properties of the covariation norm and properties of the empirical distance variogram. Moreover, it allows us to remove the dependence on $\kappa(\alpha)$ by introducing the normalized distance variogram. Normalizing the distance variogram by $\gamma^{\mathrm{ds}}(0)$ yields

$$\frac{\mathrm{V}^{\mathrm{ds}}(h)}{\mathrm{C}^{\mathrm{ds}}(0)} = \left[\frac{\| Z(h) - Z(0) \|_{\alpha}}{\| Z(0) \|_{\alpha}} \right]^{1/\alpha}, \qquad h \in \mathbb{R}^d,$$

for RFs Z that satisfy the assumptions of the theorem above. In other words, the constant $\kappa(\alpha)$ cancels and hence it is possible to extend the definition of the normalized distance variogram to $S\alpha S$ RFs with $\alpha \leq 1$, i.e., cases where (2.26) might be infinite and the distance variogram does not exist.

Definition 2.47. The *normalized distance variogram* for a univariate, strictly stationary $S\alpha S$ RF $\{Z(x) : x \in \mathbb{R}^d\}$ is defined as

$$\gamma^{\mathrm{ds}}(h) = \left[\frac{\| Z(h) - Z(0) \|_{\alpha}}{\| Z(0) \|_{\alpha}} \right]^{1/\alpha}, \qquad h \in \mathbb{R}^d,$$

provided it exists.

The following relationship between the codifference and γ^{ds} for $S\alpha S$ RFs with $0 < \alpha \leq 2$ illustrates that γ^{ds} is a meaningful dependence measure of these RFs.

Remark 2.48. By Definitions 2.15 and 2.47 it follows

$$\gamma^{\mathrm{ds}}(h)^2 = \left(2 - \frac{\tau(h)}{|||Z(0)|||_\alpha^\alpha}\right)^{2/\alpha}, \qquad h \in \mathbb{R}^d, \tag{2.27}$$

for strictly stationary, univariate $S\alpha S$ RFs Z with $0 < \alpha \leq 2$.

Using the above remark, we can derive an asymptotic result for the normalized distance variogram in the case of univariate, strictly stationary $S\alpha S$ MA processes as defined in Section 2.1.2.

Theorem 2.49. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be a univariate, strictly stationary $S\alpha S$ MA process. Then

$$\gamma^{\mathrm{ds}}(h) \to 2^{1/\alpha}, \qquad \|h\| \to \infty.$$

Proof. For one-dimensional processes, this result was stated as Theorem 4.7.3 in [91]. After making straightforward modifications, the proof there also applies to the higher-dimensional case $d > 1$, as follows.

First, assume the function f, corresponding to the MA process Z, to be compactly supported. For $h \in \mathbb{R}^d$, we define the function $f_h(x) := f(x+h), x \in \mathbb{R}^d$. If $\|h\|$ is sufficiently large then the functions f and f_h have disjoint supports, which implies

$$\begin{aligned} |||Z(x) - Z(0)|||_\alpha^\alpha &= \int_{\mathbb{R}^d} |f_h(x) - f(x)|^\alpha \,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} |f_h(x)|^\alpha \,\mathrm{d}x + \int_{\mathbb{R}^d} |f(x)|^\alpha \,\mathrm{d}x \\ &= 2|||Z(0)|||_\alpha^\alpha. \end{aligned}$$

Thus, for $\|h\| \to \infty$ we obtain $\tau(h) \to 0$ and, by (2.27), $\gamma^{\mathrm{ds}}(h) \to 2^{1/\alpha}$.

Suppose that f does not have compact support. Let $\varepsilon > 0$ and choose a closed ball $B_\varepsilon \subset \mathbb{R}^d$ such that $\int_{\mathbb{R}^d} |f(x)|^\alpha \mathbf{1}_{B_\varepsilon^c}(x) \,\mathrm{d}x < \varepsilon$, where $\mathbf{1}_{B_\varepsilon^c}$ is the indicator function of the complement of B_ε. For $h \in \mathbb{R}^d$, we define the function $(f\mathbf{1}_{B_\varepsilon})_h(x) := f(x+h)\mathbf{1}_{B_\varepsilon}(x+h), x \in \mathbb{R}^d$. As $f\mathbf{1}_{B_\varepsilon}$ has compact support, the first part of the proof shows $\tau_\varepsilon(h) = 0$, for $\|h\|$ sufficiently large, where τ_ε is obtained from τ by replacing f by $f\mathbf{1}_{B_\varepsilon}$. Now Lemma 4.7.2 in [91] shows in our case that there is a

finite $C > 0$ such that

$$\int_{\mathbb{R}^d} \big| |a(x)|^\alpha - |b(x)|^\alpha \big| \, \mathrm{d}x \leq C \left(\int_{\mathbb{R}^d} |a(x) - b(x)|^\alpha \, \mathrm{d}x \right)^{\min\{1, 1/\alpha\}},$$

for functions $a, b \in L^\alpha(\mathbb{R}^d, \lambda)$; cf. (2.4). With this inequality, together with straightforward calculations, we obtain

$$|\tau(h)| = |\tau(h) - \tau_\varepsilon(h)| \leq 2\varepsilon + (4C\varepsilon)^{\min\{1, 1/\alpha\}}.$$

Consequently, $\tau(h) \to 0$, respectively $\gamma^{\mathrm{ds}}(h) \to 2^{1/\alpha}$, for $\|h\| \to \infty$. □

Remark 2.50. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be a univariate, strictly stationary $S\alpha S$ process with $0 < \alpha < 1$. Then the theorem above together with Remark 2.8 implies that if $\gamma^{\mathrm{ds}}(h) = 2^{1/\alpha}$ for $h \in \mathbb{R}^d$ then $Z(0)$ and $Z(h)$ are independent.

It is well-known that the usual variogram is a conditionally negative definite function.[12] The next result shows that a similar statement holds for the distance variogram for $S\alpha S$ processes if $1 \leq \alpha \leq 2$.

Theorem 2.51. For $1 \leq \alpha \leq 2$, let $\{Z(x) : x \in \mathbb{R}^d\}$ be a univariate, intrinsically strictly stationary $S\alpha S$ RF. Then the normalized distance variogram is a conditionally negative definite function.

Proof. Let $n \geq 2$ be an integer. Property 2.10.7 in [91] states that if $(X_1, \ldots, X_n)^t$ is an $S\alpha S$-random vector, then the matrix $T \in \mathbb{R}^{n \times n}$ with entries

$$T_{ij} = \|\!| X_i |\!\|_\alpha^\alpha + \|\!| X_j |\!\|_\alpha^\alpha - \|\!| X_i - X_j |\!\|_\alpha^\alpha$$

is positive definite. Consequently, the codifference function

$$\tau(h) = \|\!| Z(0) |\!\|_\alpha^\alpha + \|\!| Z(h) |\!\|_\alpha^\alpha - \|\!| Z(0) - Z(h) |\!\|_\alpha^\alpha$$

is positive definite, which implies immediately that $2 - \tau(h)/\|\!| Z(0) |\!\|_\alpha^\alpha$ is conditionally negative definite. Now Corollary 2.10 in [9] shows that if $\psi : \mathbb{R} \to [0, \infty)$ is a conditionally negative definite function with $\psi(0) \geq 0$, then ψ^β is again condition-

[12] Let $n \geq 1$ be an integer. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *conditionally negative definite* if, for all $x_1, \ldots, x_n \in \mathbb{R}^d$ and $a_1, \ldots, a_n \in \mathbb{R}$ with $\sum_{i=1}^n a_i = 0$, it holds that $\sum_{i=1}^n \sum_{j=1}^n a_i a_j f(x_i - x_j) \leq 0$.

ally negative definite for $\beta \in [0, 1]$. Hence, it follows that

$$\frac{\mathrm{V}^{\mathrm{ds}}(h)}{\mathrm{C}^{\mathrm{ds}}(0)} = \left(2 - \frac{\tau(h)}{\|Z(0)\|_\alpha^\alpha}\right)^{1/\alpha}$$

is conditionally negative definite. □

Evidently, the proof is not applicable in the case $0 < \alpha < 1$, indicating that we only know that the normalized distance variogram is a power of a conditionally negative definite function in that case.

Remark 2.52. Theorem 2.51 implies that under the assumption that the data follows an $S\alpha S$ process with $1 \leq \alpha \leq 2$, the choice of fitting a parametric curve to the empirical distance variogram is restricted to the case of conditionally negative definite functions. This narrows the potential functions for fitting and establishes further possibilities to check if the data fulfills the assumptions.

Unfortunately, we do not know if every member of the class of non-negative conditionally negative definite functions yields a valid normalized distance variogram for $S\alpha S$ processes with $1 \leq \alpha < 2$. All we know is that for every conditionally negative definite function γ there is some intrinsically stationary Gaussian process with variogram γ (a consequence of Theorem 6.1.9 in [92]) and, hence, every conditionally negative definite function is a valid normalized distance variogram of an $S\alpha S$ process with $\alpha = 2$. However, we expect that the class of admissible positive definite functions shrinks with decreasing α.

We conclude this subsection by investigating the calculation of the normalized distance variogram in the case of univariate $S\alpha S$ MA RFs. Remark 2.48 shows that the computation of the normalized distance variogram only involves the calculation of $\|\!|Z(h) - Z(0)|\!\|_\alpha$ and $\|\!|Z(0)|\!\|_\alpha$. If $\{Z(x) : x \in \mathbb{R}^d\}$ is a univariate $S\alpha S$ MA RF, then

$$Z(x) = \int_{\mathbb{R}^d} f(x - y)\, M(\mathrm{d}y)$$

for some admissible kernel $f : \mathbb{R}^d \to \mathbb{R}$. By the linearity of the stable integral (Proposition 2.10),

$$Z(h) - Z(0) = \int_{\mathbb{R}^d} (f(h - y) - f(y))\, M(\mathrm{d}y), \qquad \text{almost surely},$$

which again is an $S\alpha S$ random variable for each $h \in \mathbb{R}^d$. In the sequel, we assume

the control measure of M to be the Lebesgue measure. Then Proposition 2.12 implies that

$$|||Z(0)|||_\alpha^\alpha = \int_{\mathbb{R}^d} |f(y)|^\alpha \,\mathrm{d}y, \qquad h \in \mathbb{R}^d,$$

and analogously it follows that

$$|||Z(h) - Z(0)|||_\alpha^\alpha = \int_{\mathbb{R}^d} |f(h-y) - f(y)|^\alpha \,\mathrm{d}y, \qquad h \in \mathbb{R}^d.$$

Lemma 2.53. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be a univariate $S\alpha S$ MA process with kernel function f. Assume the control measure of M to be the Lebesgue measure. Then

$$\gamma^{\mathrm{ds}}(h) = \left[\frac{\int_{\mathbb{R}^d} |f(h-y) - f(y)|^\alpha \,\mathrm{d}y}{\int_{\mathbb{R}^d} |f(y)|^\alpha \,\mathrm{d}y} \right]^{1/\alpha}, \qquad h \in \mathbb{R}^d.$$

Note that the calculation of the normalized distance variogram of a univariate $S\alpha S$ MA process involves the integration of expressions depending only on the kernel function of the process. This is, in practice, often simpler as the calculation of the distance correlation, as the evaluation of the latter might require the integration of a function with a singularity at the origin. It might even be possible to find analytical expressions for γ^{ds} for certain kernel functions, which might provide more insight into the structure of the class of distance variograms. However, this is beyond the scope of this work, although it may be interesting for future research.

2.5 Consistency of the empirical distance correlation function

We have provided, so far, only the definitions of the EDCF and the empirical distance variogram, but we have written nothing about their reliability as estimators. This will be dealt with in the current section, where we show that these estimators are consistent under certain assumptions. Here, we are guided by the work of Zhou [117] who proved the consistency of the EDCF in the case of multivariate time series. Throughout this section, Z denotes an m-variate, strictly stationary RF on $\mathbb{Z}^d$, which is called a *transformation field* if it is representable as

$$Z(j) = g\left(\varepsilon(j-s) : s \in \mathbb{Z}^d\right), \qquad j \in \mathbb{Z}^d, \tag{2.28}$$

where $g : \mathbb{R}^{\mathbb{Z}^d} \to \mathbb{R}^m$ is a measurable function and $(\varepsilon(j))_{j\in\mathbb{Z}^d}$ are i.i.d. real-valued random variables.

Remark 2.54. We will analyze consistency in the remainder of this section only for transformation fields. Thus, one might wonder how strong is this restriction. For $S\alpha S$ RFs, we can provide at least a qualitative answer. Each realization of an $S\alpha S$ RF on $\mathbb{R}^d$ can be approximated with arbitrary precision by a realization of a certain $S\alpha S$ RF $\{Z(j) : j \in \mathbb{Z}^d\}$. Now Remark 2.14 (2) indicates that realizations of Z can be approximated by sums of independent $S\alpha S$ random variables, i.e., by realizations of $S\alpha S$ transformation fields. Consequently, it is not overly restrictive to consider transformation fields in practical applications.

It is well-known that proofs of the consistency of estimators based on observing a single realization of some RF always require assumptions on the RF, ensuring that it provides a mixing structure or short-range dependence. In our case, we will assume a short-range dependence assumption based on the so-called *physical dependence measure*. This concept was introduced by Wu [114] for time series of the form (2.28) (i.e., $d = 1$) and it was generalized recently by Machkouri, Volný, and Wu [68] to transformation fields with $d > 1$. The basic idea is to measure the influence of the input variable $\varepsilon(0)$ on the transformation field.

Definition 2.55. Let the RF $\{Z(j) : j \in \mathbb{Z}^d\}$ be a transformation field, i.e., it meets (2.28) for a collection of i.i.d. random variables $(\varepsilon(j))_{j\in\mathbb{Z}^d}$ and some measurable function g. Further, let $(\varepsilon^*(j))_{j\in\mathbb{Z}^d}$ be another collection of i.i.d. random variables defined via

$$\varepsilon^*(j) = \begin{cases} \varepsilon(j), & j \neq 0, \\ \varepsilon'(0), & j = 0, \end{cases}$$

where $\varepsilon'(0)$ is an independent copy of $\varepsilon(0)$. Then, for $r > 0$ and each $j \in \mathbb{Z}^d$, the *physical dependence measure* $\delta_{j,r}$ is defined as

$$\delta_{j,r} = \|Z(j) - Z^*(j)\|_r = \left[\mathbb{E}\left(|Z(j) - Z^*(j)|^r\right)\right]^{1/r}, \tag{2.29}$$

with $Z^*(j) = g\left(\varepsilon^*(j - s) : s \in \mathbb{Z}^d\right)$.

Based on this, we say that a transformation field $\{Z(j) : j \in \mathbb{Z}^d\}$ fulfills the

q-short-range criterion,[13] if there is a $q > 0$ such that

$$\sum_{j \in \mathbb{Z}^d} \delta_{j,q} < \infty. \tag{2.30}$$

This means, roughly speaking, that the effect of a single input variable on a point of the transformation field vanishes at infinite distances and the cumulated effects are bounded.

The statement and the proof of the consistency of the EDCF use the following notation. First, similar to Zhou [117], the functions α and β_k are defined as

$$\alpha(t, j) = \exp(itZ(j)) \qquad \text{and} \qquad \beta_k(s, j) = \exp(isZ(j + k)),$$

for $j, k \in \mathbb{Z}^d$ and $t, s \in \mathbb{R}^m$. Second, referring to Machkouri, Volný, and Wu [68], the projection operator $\mathcal{P}_i$ is defined as follows: Let $\tau : \mathbb{N} \to \mathbb{Z}^d$ be a bijection and $i \in \mathbb{Z}$ and $j \in \mathbb{Z}^d$. Then

$$\mathcal{P}_i Z(j) = \mathbb{E}(Z(j)|\mathcal{F}_i) - \mathbb{E}(Z(j)|\mathcal{F}_{i-1}),$$

where $\mathcal{F}_i = \sigma(\varepsilon(\tau(l)) : l \leq i)$ is the σ-algebra generated by the set of random variables $\{\varepsilon(\tau(l)) : l \leq i\}$. [14,15]

Theorem 2.56. Let $\{Z(x) : x \in \mathbb{R}^d\}$ be a univariate, strictly stationary transformation field, such that Z fulfills the q-short-range criterion (2.30) and $\|Z(x)\|_q = (\mathbb{E}|Z(x)\|^q)^{1/q} < \infty$ for some $q > 1$. Further, assume that we have samples of Z on the lattice

$$\Gamma = \{-n, -n + 1, \ldots, n - 1, n\}^d, \qquad n \in \mathbb{N}.$$

Then

$$\mathcal{V}^\Gamma(k) \xrightarrow{p} \mathrm{C}^{\mathrm{ds}}(k), \qquad n \to \infty,$$

for all $k \in \mathbb{Z}^d$. In other words, the empirical distance covariance function is a weakly consistent estimator for the distance covariance function.

In the proof we will use the following Lemma.

[13] Machkouri, Volný, and Wu [68] called RFs fulfilling this criterion q-stable. As this notation could be confused with the term 'α-stable', we use a different term.

[14] With the properties of the conditional expectation and $\mathcal{F}_{i-1} \subset \mathcal{F}_i$ it is straightforward to check that $P_i(P_i(Z(j))) = P_i(Z(j))$, showing that $\mathcal{P}_i$ is indeed a projection operator.

[15] Note that $\mathcal{F}_i = \sigma\{\emptyset\}$ is the trivial σ-algebra for any integer $i < 0$.

Lemma 2.57. Let $\tau : \mathbb{N} \to \mathbb{Z}^d$ be bijective. For all $j \in \mathbb{Z}^d, l \in \mathbb{Z}, \eta \in (0,1]$ and $q \geq 1$, the following inequalities hold.

$$\begin{aligned}\|\mathcal{P}_l \alpha(t,j)\|_q &\leq 2^{2-\eta}|t|^\eta [\delta(j-\tau(l),\eta q)]^\eta \\ \|\mathcal{P}_l \alpha(t,j)\beta_k(s,j)\|_q &\leq 2^{2-\eta}\left(|t|^\eta [\delta(j-\tau(l),\eta q)]^\eta + |s|^\eta [\delta(j+k-\tau(l),\eta q)]^\eta\right)\end{aligned}$$

Proof of Lemma 2.57: Note that Lemma 1 in [68] implies

$$\|\mathcal{P}_l \alpha(t,j)\|_q \leq \|\alpha(t,j) - \alpha^*(t,j)\|_q,$$

where $\alpha^*(t,j) = \exp(itZ^*(j))$. With this the proof is completely analogous to the proof of Lemma 1 in [117]. □

Proof of Theorem 2.56. The proof follows the lines of the proofs of Theorem 1 in [117] and Theorem 2 in [108]. Note that $k \in \mathbb{Z}^d$ is fixed but arbitrary and Γ is assumed to be 'big' enough, such that $k \in \Gamma$. Further, we assume without restriction that $1 < q \leq 2$. In the following, we use the notations $\mathrm{d}w = (c_1^2|t|^2|s|^2)^{-1}\,\mathrm{d}t\,\mathrm{d}s$, and

$$\begin{aligned}\widehat{\zeta}^\Gamma(t,s) &= \widehat{\varphi}_k^\Gamma(t,s) - \widehat{\varphi}_k^\Gamma(t)\widehat{\varphi}^\Gamma(s), \\ \zeta(t,s) &= \varphi_{Z(k+j),Z(j)}(t,s) - \varphi_{Z(k+j)}(t)\varphi_{Z(j)}(s),\end{aligned}$$

with the empirical characteristic functions given as in (2.19). For $\Delta > 0$ the set $D(\Delta)$ is defined to be

$$D(\Delta) = \{(t,s) \in \mathbb{R} \times \mathbb{R} : \Delta \leq |t|, |s| \leq 1/\Delta\},$$

and the empirical and theoretical distance covariances 'restricted' to $D(\Delta)$ are defined as

$$\mathcal{V}_\Delta^\Gamma(k) = \int_{D(\Delta)} |\widehat{\zeta}^\Gamma(t,s)|^2\,\mathrm{d}w \qquad \text{and} \qquad \mathrm{C}_\Delta^{\mathrm{ds}}(k) = \int_{D(\Delta)} |\zeta(t,s)|^2\,\mathrm{d}w.$$

For all $t, s \in \mathbb{R}$ we have

$$\begin{aligned}\left||\widehat{\zeta}^\Gamma(t,s)|^2 - |\zeta(t,s)|^2\right| &= \left|\left(|\widehat{\zeta}^\Gamma(t,s)| + |\zeta(t,s)|\right)\left(|\widehat{\zeta}^\Gamma(t,s)| + |\zeta(t,s)|\right)\right| \\ &\leq 4\left|\widehat{\zeta}^\Gamma(t,s) - \zeta(t,s)\right| \\ &= 4\Bigg[\widehat{\varphi}_k^\Gamma(t,s) - \varphi_{Z(k+j),Z(j)}(t,s) + \varphi_{Z(k+j)}(t)\varphi_{Z(j)}(s)\end{aligned}$$

$$
\begin{aligned}
&\quad + \widehat{\varphi}^{\Gamma}(s)\varphi_{Z(k+j)}(t) - \widehat{\varphi}^{\Gamma}(s)\varphi_{Z(k+j)}(t) - \widehat{\varphi}_k^{\Gamma}(t)\widehat{\varphi}^{\Gamma}(s)\Big] \\
&\leq 4\,[T_1 + T_2 + T_3],
\end{aligned}
$$

where

$$
\begin{aligned}
T_1 &= \left|\widehat{\varphi}_k^{\Gamma}(t,s) - \varphi_{Z(k+j),Z(j)}(t,s)\right| \\
T_2 &= \left|\widehat{\varphi}^{\Gamma}(s)\left(\widehat{\varphi}_k^{\Gamma}(t) - \varphi_{Z(k+j)}(t)\right)\right| \\
T_3 &= \left|\varphi_{Z(k+j)}(t)\left(\widehat{\varphi}^{\Gamma}(s) - \varphi_{Z(j)}(s)\right)\right|.
\end{aligned}
$$

Note that the first inequality above is a consequence of the reverse triangle inequality and the fact that $|\widehat{\varphi}_k^{\Gamma}(t,s)| \leq 2$ and $|\zeta(t,s)| \leq 2$, for all $t, s \in \mathbb{R}$ and all integers $n > 0$.

In a first step, we show that $|\mathcal{V}_{\Delta}^{\Gamma}(k) - \mathrm{C}_{\Delta}^{\mathrm{ds}}(k)| \xrightarrow{p} 0$ and, to this end, it suffices to show that $\|\mathcal{V}_{\Delta}^{\Gamma}(k) - \mathrm{C}_{\Delta}^{\mathrm{ds}}(k)\|_q \to 0$.

Let $\tau : \mathbb{N} \to \mathbb{Z}^d$ be a bijection and define

$$
\Psi_{\Gamma,l}(t,s) := \sum_{j \in N_{\Gamma}(k)} \mathcal{P}_{\tau^{-1}(j+k-l)}\left(\alpha(t,j)\beta_k(s,j)\right).
$$

We show that the summands of $\Psi_{\Gamma,l}(t,s)$ form a martingale difference sequence for each $l \in \mathbb{Z}^d$. Consider the sequence

$$
\left\{\mathcal{P}_{\tau^{-1}(j+k-l)}\left(\alpha(t,j)\beta_k(s,j)\right), \mathcal{F}_{\tau^{-1}(j+k-l)}\right\}_{\tau^{-1}(j+k-l)}. \tag{2.31}
$$

Since τ is bijective, it is possible to arrange the indices $\tau^{-1}(j+k-l)$ for all $j \in N_{\Gamma}(k) := \{j \in \Gamma : j + k \in \Gamma\}$ in increasing order and rename them by the following indices

$$
\iota_1 < \iota_2 < \cdots < \iota_{n_k},
$$

where $n_k := |N_{\Gamma}(k)|$. With this ordering, sequence (2.31) becomes adapted, meaning that $\mathcal{P}_{\iota}$ is $\mathcal{F}_{\iota}$-measurable for every $\iota_1 \leq \iota \leq \iota_{n_k}$. Furthermore, from the definition of $\mathcal{P}_{\iota}$ and the Law of Total Expectation it follows that

$$
\mathbb{E}\left(\mathcal{P}_{\iota_r}\left(\alpha(t,\iota_r)\beta_k(s,\iota_r)\right)|\mathcal{F}_{\iota_{r-1}}\right) = 0, \qquad r \in \{2,\ldots,n_k\},
$$

and hence the summands of $\Psi_{\Gamma,l}(t,s)$ form a martingale difference sequence. Ap-

plying Burkholder's inequality (cf. equation (1.1) in [89]) yields the existence of a finite positive constant A_q depending only on q, such that

$$\|\Psi_{\Gamma,l}(t,s)\|_q^q \leq A_q \mathbb{E}\left(\sum_{i=1}^{n_k} \left|\mathcal{P}_{\iota_i}(\alpha(t,\iota_i)\beta_k(s,\iota_i))\right|^2\right)^{q/2}.$$

This in concert with $q \leq 2$ yields

$$\begin{aligned}\frac{\|\Psi_{\Gamma,l}(t,s)\|_q^q}{A_q} &\leq \mathbb{E}\left(\sum_{i=1}^{n_k} \left|\mathcal{P}_{\iota_i}(\alpha(t,\iota_i)\beta_k(s,\iota_i))\right|^2\right)^{q/2} \\ &\leq \mathbb{E}\sum_{i=1}^{n_k} \left|\mathcal{P}_{\iota_i}(\alpha(t,\iota_i)\beta_k(s,\iota_i))\right|^q,\end{aligned}$$

which, by Lemma 2.57, is less than or equal to

$$\left(\frac{2}{\Delta}\right)^q n_k\big[\delta(l-k,q)+\delta(l,q)\big]^q.$$

Consequently, the assumption that Z satisfies the q-short range criterion together with the Minkowski inequality imply that the series $\left|\sum_{l\in\mathbb{Z}^d}\Psi_{\Gamma,l}(t,s)\right|$ converges absolutely in q-th mean. Moreover, the limit is $n_k T_1$, which can be seen as follows. We have

$$\mathbb{E}\big(\alpha(t,j)\beta_k(s,j)|\mathcal{F}_{-1}\big) = \mathbb{E}\big(\alpha(t,j)\beta_k(s,j)\big) = \varphi_{Z(k+j),Z(j)}(t,s),$$

since $\mathcal{F}_{-1}$ is the trivial σ-algebra, and

$$\mathbb{E}\big(\alpha(t,j)\beta_k(s,j)|\mathcal{F}_\infty\big) = \widehat{\varphi}_k^\Gamma(t,s).$$

Thus, a reordering of the series shows

$$\sum_{l\in\mathbb{Z}^d}\Psi_{\Gamma,l}(t,s) \overset{q}{=} n_k\mathbb{E}\big(\alpha(t,j)\beta_k(s,j)|\mathcal{F}_\infty\big) - n_k\mathbb{E}\big(\alpha(t,j)\beta_k(s,j)|\mathcal{F}_{-1}\big),$$

meaning equality of both sides in the q-th mean sense. Therefore, we find that

$$n_k\|T_1\|_q \leq \kappa n_k^{1/q},$$

that is,

$$\|T_1\|_q \le \kappa n_k^{1/q-1}.$$

Here and in the remainder of the proof, κ denotes a positive constant with possibly different values in different contexts.

It follows in a similar manner that

$$\|T_2\|_q \le \kappa n_k^{1/q-1} \qquad \text{and} \qquad \|T_3\|_q \le \kappa n_k^{1/q-1},$$

which results in

$$\|\mathcal{V}_\Delta^\Gamma(k) - \mathrm{C}_\Delta^{\mathrm{ds}}(k)\|_q \le 4 \int_{D(\Delta)} \big(\|T_1\|_q + \|T_2\|_q + \|T_3\|_q\big)\,\mathrm{d}w \le \kappa n_k^{1/q-1}, \tag{2.32}$$

yielding $|\mathcal{V}_\Delta^\Gamma(k) - \mathrm{C}_\Delta^{\mathrm{ds}}(k)| \xrightarrow{p} 0$ for all k.

Since $|\mathrm{C}_\Delta^{\mathrm{ds}}(k) - \mathrm{C}^{\mathrm{ds}}(k)| \to 0$ for $\Delta \to 0$, it suffices to show that in probability

$$\limsup_{\Delta\to 0} \limsup_{n_k\to\infty} |\mathcal{V}^\Gamma(k) - \mathcal{V}_\Delta^\Gamma(k)| = 0, \qquad k \in \mathbb{Z}^d,$$

i.e., the restricted empirical distance covariance uniformly converges in probability to the unrestricted one. For $\Delta > 0$,

$$\begin{aligned}
|\mathcal{V}_\Delta^\Gamma(k) - \mathcal{V}^\Gamma(k)| &= \left| \int_{D(\Delta)} |\widehat{\zeta}^\Gamma(t,s)|^2\,\mathrm{d}w - \int_{\mathbb{R}^2} |\widehat{\zeta}^\Gamma(t,s)|^2\,\mathrm{d}w \right| \\
&\le \int_{|t|<\Delta} |\widehat{\zeta}^\Gamma(t,s)|^2\,\mathrm{d}w + \int_{|t|>1/\Delta} |\widehat{\zeta}^\Gamma(t,s)|^2\,\mathrm{d}w \\
&\quad + \int_{|s|<\Delta} |\widehat{\zeta}^\Gamma(t,s)|^2\,\mathrm{d}w + \int_{|s|>1/\Delta} |\widehat{\zeta}^\Gamma(t,s)|^2\,\mathrm{d}w.
\end{aligned}$$

Consider the first summand on the right-hand side. By using exactly the same arguments as in the proofs of Theorem 2 in [108] and Theorem 1 in [117] it can be shown that there is a constant $\kappa > 0$ such that

$$\begin{aligned}
\int_{|t|<\Delta} |\widehat{\zeta}^\Gamma(t,s)|^2\,\mathrm{d}w \le \kappa \Delta^{r/2} \frac{2}{n_k} &\left(\sum_{j\in N_\Gamma(k)} \big(|Z(j+k)| + \mathbb{E}|Z(0)|\big) \right) \\
&\times \frac{2}{n_k} \left(\sum_{j\in N_\Gamma(k)} \big(|Z(j+k)|^{1+r/2} + \mathbb{E}|Z(0)|^{1+r/2}\big) \right)
\end{aligned} \tag{2.33}$$

with $r = q - 1$. Similar arguments as in the proof of (2.32) lead to

$$\frac{2}{n_k} \sum_{j \in N_\Gamma(k)} |Z(j+k)| \xrightarrow{p} \mathbb{E}|Z(0)|$$

and

$$\frac{2}{n_k} \sum_{j \in N_\Gamma(k)} |Z(j+k)|^{1+r/2} \xrightarrow{p} \mathbb{E}|Z(0)|^{1+r/2},$$

implying together with (2.33) that there is a constant $\kappa > 0$ such that in probability

$$\limsup_{n_k \to \infty} \int_{|t|<\Delta} |\widehat{\zeta}^\Gamma(t,s)|^2 \, \mathrm{d}w \leq \kappa \Delta^{r/2} \mathbb{E}|Z(0)| \mathbb{E}|Z(0)|^{1+r/2}.$$

Consequently, in probability

$$\limsup_{\Delta \to 0} \limsup_{n_k \to \infty} \int_{|t|<\Delta} |\widehat{\zeta}^\Gamma(t,s)|^2 \, \mathrm{d}w = 0.$$

Now consider $\int_{|t|>1/\Delta} |\widehat{\zeta}^\Gamma(t,s)|^2 \, \mathrm{d}w$. The proof of Theorem 2 in [108] shows that this can be bounded by

$$16\Delta \frac{2}{n_k} \sum_{j \in N_\Gamma(k)} \big(|Z(j+k)| + \mathbb{E}|Z(0)| \big),$$

which yields in probability

$$\limsup_{\Delta \to 0} \limsup_{n_k \to \infty} \int_{|t|>1/\Delta} |\widehat{\zeta}^\Gamma(t,s)|^2 \, \mathrm{d}w = 0,$$

by using the same arguments as above.

The uniform convergence of the other two summands in (2.33) can be shown in the same way. Therefore, the proof is complete. □

Corollary 2.58. Under the assumptions of Theorem 2.56, the empirical distance correlation is a consistent estimator.

Remark 2.59. (1) Making appropriate adjustments in the proof of Theorem 2.56 implies that the empirical cross-distance covariance function is a consistent

estimator.

(2) Note that we have proved the consistency of the EDCF for a certain case of *increasing domain asymptotics*. That is, we assumed that the observed area increases with the number of observed points, while the distances between those points remained the same. However, the proof of Theorem 2.56 does not imply the consistency of the EDCF under *infill asymptotics*, where the observed area remains the same but the number of observed points within this area increases. The results of Section 2.6.2 will show that the EDCF indeed is not consistent under infill asymptotics.

Consistency of the empirical distance variogram

If $\{Z(j) : j \in \mathbb{Z}^d\}$ is a transformation field, i.e., it satisfies (2.28), then the increment process $Y(j) = Z(k+j) - Z(k), j \in \mathbb{Z}^d$, also is a transformation field. Although the normalized distance variogram may be defined for RFs with infinite or undefined first moments, Theorem 2.56 implies the consistency of the empirical distance variogram only for RFs with q-short-range increment process and finite central moments of order q for some $q > 1$.

Corollary 2.60. Let $q > 1$. The empirical distance variogram is a consistent estimator for univariate, strictly stationary transformation fields $\{Z(j) : j \in \mathbb{Z}^d\}$ with q-short-range increment process Y satisfying $\|Y(0)\|_q < \infty$.

2.6 Finite sample performance

The applicability of the DCF to RFs without finite second moments or with a highly non-linear dependence structure is appealing for several research areas, such as hydrology or climate sciences [59]. In fact, the DCF can reveal structural dependencies that might be overlooked if one uses other dependence measures, as we have illustrated in Section 2.1.3. A crucial measure of the usefulness of the distance correlation function and the distance variogram is the finite sample performance of their estimators. For this reason, we apply the empirical distance correlation function (EDCF) and the empirical distance variogram to simulated data from white noise, Gaussian, $S\alpha S$, and max-stable fields, and we analyze the dependence of their performance on the sample sizes. In particular, a goal of this section is to find heuristic rules to determine when the EDCF approximates the population DCF.

2.6.1 White noise processes

A spatial white noise process $\{Z(x) : x \in \mathbb{R}^d\}$ possesses a very simple dependence structure. This process consists of a collection of independent and identically distributed random variables, which implies that $Z(x)$ is independent of $Z(y)$ for each $x \neq y$, and that Z is strictly stationary. In other words, the DCF is given by

$$\mathrm{R}^{\mathrm{ds}}(h) = \begin{cases} 1, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

Although it is relatively straightforward to analyze white noise processes via the EDCF, it helps understand its finite sample performance.

In the following example, we simulate Gaussian white noise processes Z on grids $\Gamma = \{1, \ldots, n\}^2$ for different integers $n > 1$ and calculate the EDCF. Note that it makes no difference for white noise processes if we use increasing domain or infill asymptotics: both types are equivalent in this case because there is no stochastic dependence between different points. Hence, although we formally use infill asymptotics, we expect the estimates to get more reliable for increasing n.

As a spatial white noise process is strictly stationary and isotropic, we display in Figure 2.4 the EDCF as a function of the distance. To obtain a single value for each distance, we calculate the mean of all values of the EDCF corresponding to lag vectors with the same length.[16] It can be seen that, for all grids, the empirical distance correlation:

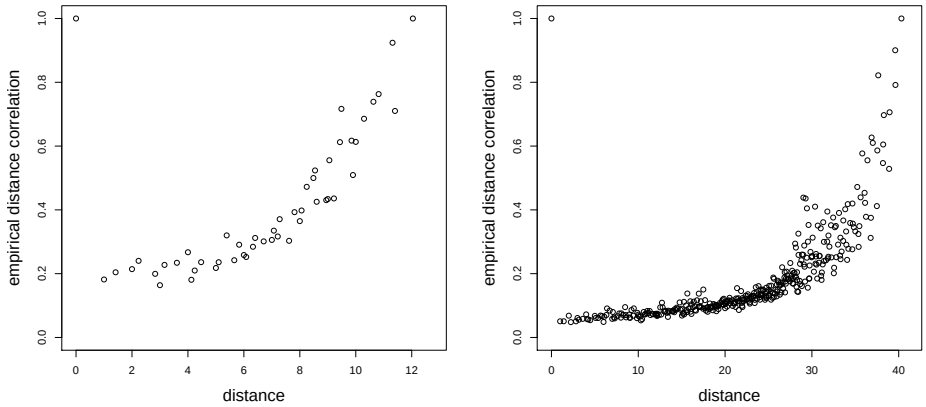
(i) equals one for distance zero;

(ii) is close to zero for sufficiently small non-zero distances;

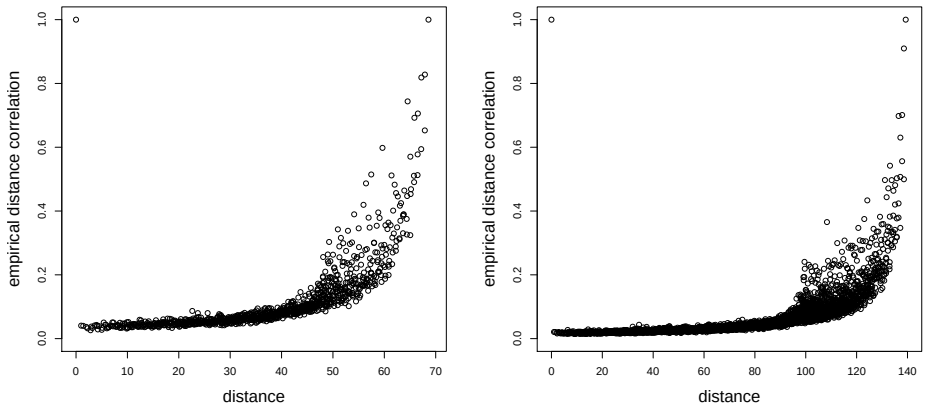
(iii) and increases with increasing distance.

Point (i) holds by construction and Point (ii) can be explained as follows. The EDCF is always non-negative, and, for a given lag vector, equals zero only if all observations separated by this lag vector are equal, cf. Subsection 2.2.1. Consequently, we expect the EDCF for samples of white noise processes to have a certain level greater than zero for non-zero distances. In our observations, the level of the EDCF for small non-zero distances depends on n and decreases for increasing n. For $n = 10$ it is approximately 0.2, which differs strongly from the theoretical value zero, showing the poor finite sample performance in this case. For $n = 30, n = 50$

[16] We will use this procedure in the remainder of this section for all isotropic processes.



(a) $n = 10$ (b) $n = 30$



(c) $n = 50$ (d) $n = 100$

Figure 2.4: Empirical isotropic distance correlation, plotted against the Euclidean distance, $\|h\|$, for Gaussian white noise processes simulated on four different grids $\{1, \ldots, n\}^2$.

and $n = 100$ the level equals approximately $0.07, 0.04$, and 0.02, which is in all cases sufficiently close to zero.

Point (iii) seems to be a contradiction, as we do not expect the EDCF to increase with increasing distance. However, the behaviour can be explained by the fact that for increasing distances the number of data points used to calculate the empirical distance correlation decreases. Hence, the EDCF for large distances is calculated by using only few sample points, making the EDCF an unreliable estimator for large distances. For distances greater than zero and smaller than $n/2$ the empirical distance correlation remains roughly the same. As a consequence, we consider the EDCF an unreliable estimator for distances greater than $n/2$, if we have samples of isotropic fields on the grid $\{1, \dots, n\}^2$.

2.6.2 Gaussian fields

As mentioned in Subsection 2.3.3, the DCF for Gaussian RFs can be expressed in closed analytical form. This implies that Gaussian RFs constitute a large class of RFs with non-trivial dependence structure and explicitly given DCF – a fact we will use to investigate the finite sample performance of the EDCF in cases of increasing domain and infill asymptotics. These two types are compared, as we wish to obtain deeper insight into the behaviour of the EDCF. We expect that the EDCF converges to the true DCF in the case of increasing domain asymptotics, as stated in Theorem 2.56. In the case of infill asymptotics, we do not expect convergence, analogously to the case of the usual variogram.

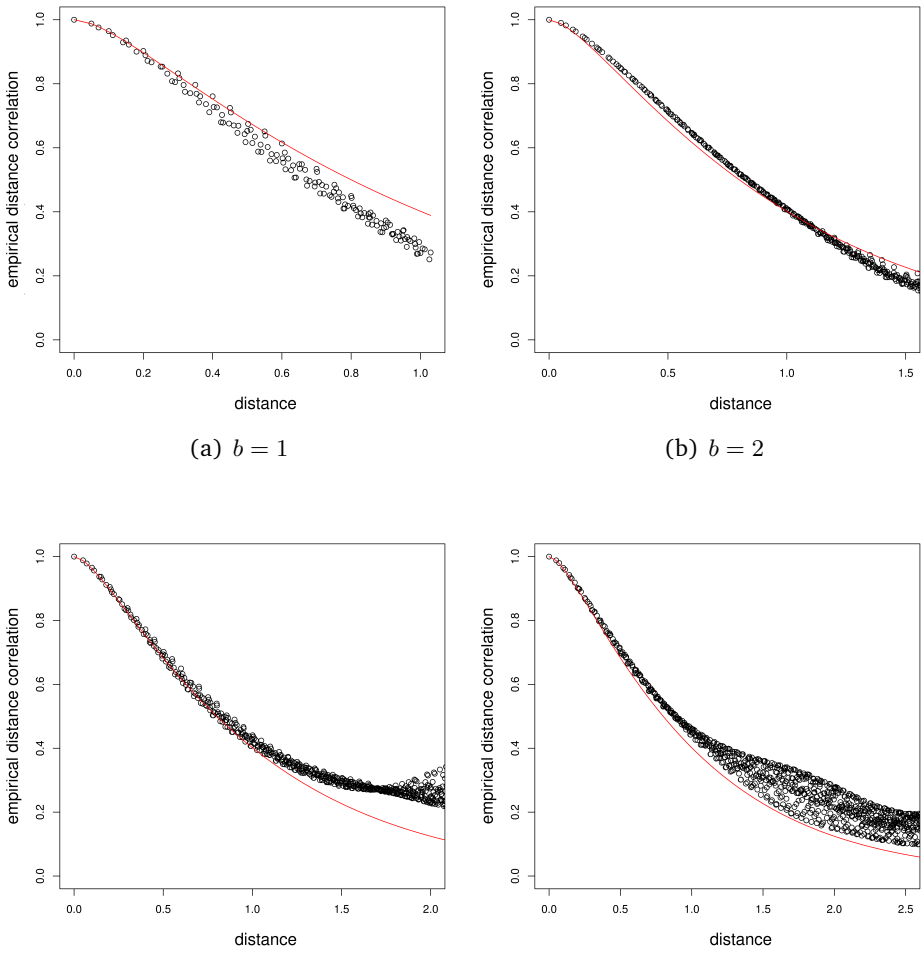
By investigating these asymptotics, we obtain an understanding of how the behaviour of the EDCF depends on the ratio of the length scale of the underlying process to the size of the observed area. We believe that this behaviour may be useful for judging the accuracy of an estimate of the DCF in a given application.

Increasing domain asymptotics

For the proposed investigation, we calculate the EDCF for a simulated centered Gaussian process possessing the Matérn correlation function with the parameterization introduced in Handcock and Wallis [46],

$$C(h) = \|\sqrt{2}h\| \, K_1(\|\sqrt{2}h\|), \qquad h \in \mathbb{R}^2,$$

where K_1 is the modified Bessel function of the second kind with parameter 1. The corresponding simulation takes place on the regular grid Γ_b, with fixed grid spacing



(a) $b = 1$ (b) $b = 2$

(c) $b = 3$ (d) $b = 4$

Figure 2.5: Empirical distance correlation as a function of the Euclidean distance, $\|h\|$, for Gaussian RFs with Matérn correlation function C, simulated on grids $\Gamma_b = \{-b, -b + 0.05, \ldots, b - 0.05, b\}^2$. The red curves display the true DCFs.

0.05×0.05 but increasing area. Specifically, we choose

$$\Gamma_b = \{-b, -b + 0.05, -b + 0.1, \ldots, b - 0.1, b - 0.05, b\}^2.$$

The results are plotted in Figure 2.5. Note that the range of the x-axis increases from $[0, 1]$ in panel (a) to $[0, 2.5]$ in panel (d). Bearing this in mind, we see that the estimates become more reliable with increasing domains. Panels (c) and (d) show particularly good estimates for $|\Gamma_3| = 121^2$, resp. $|\Gamma_4| = 161^2$, observed points.

Infill asymptotics

In this case, we consider realizations of a centered Gaussian process Z with the exponential covariance function

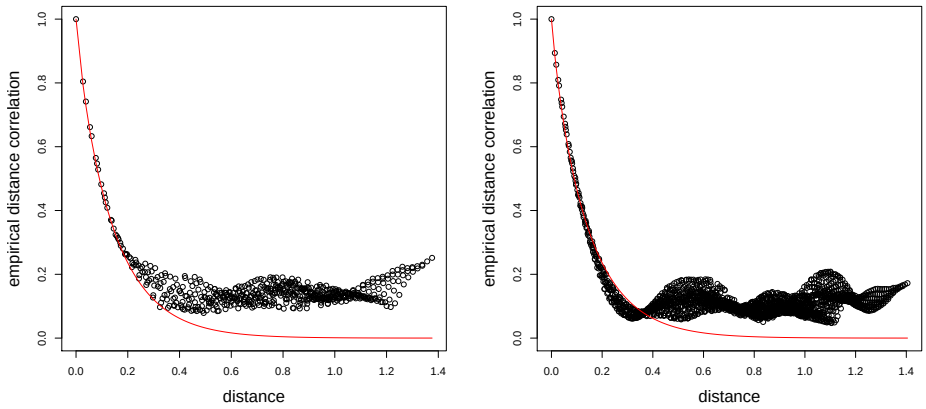
$$C(h) = \exp(-\|h\|/a), \qquad h \in \mathbb{R}^2,$$

where $a > 0$ is the scale parameter. This process is simulated on regular grids $\Gamma \subset [-1, 1]^2$ with grid spacing $2/(n-1) \times 2/(n-1)$, depending on a variable integer $n \geq 2$. Figure 2.6 shows plots of the EDCF against distance, calculated for simulations of Z using different grid spacings and scale parameters. We see that the EDCF resembles the true DCF in a similar acceptable manner for both grid spacings in the case $a = 0.15$.

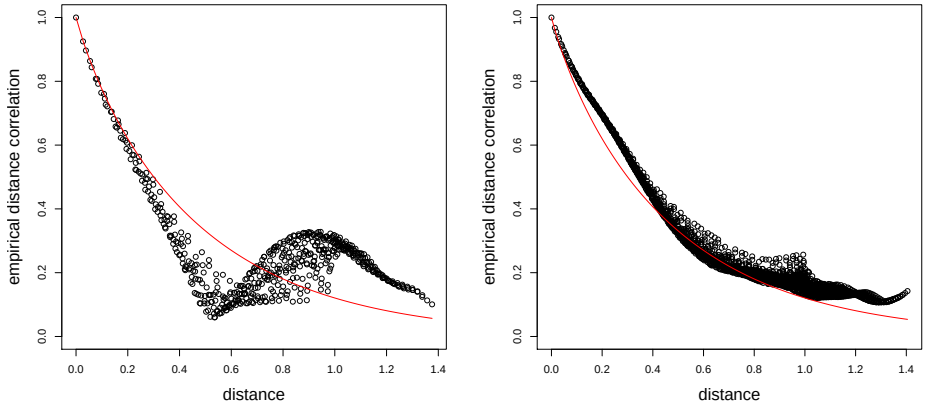
If $a = 0.5$ then the situation changes. In the case $n = 75$, we see an unacceptable estimate of the ECDF. If $n = 150$, the estimate looks slightly better, but the estimate differs significantly from the true DCF also for small distances. This indicates that the EDCF should be handled with care if the range of the dependence of the process is large in comparison with the range of the observed region. However, this is not surprising, as the estimator for the Pearson sample correlation function also lacks consistency in this case. We illustrate this in Figure 2.7, where we use (2.8) to calculate the Pearson sample correlation function, for the realization depicted in Figure 2.6 (d). It can be seen that the Pearson sample correlation function differs significantly from the theoretical correlation function $\exp(-2\|h\|)$.

2.6.3 Stable random fields

In the following, we study the applicability of the EDCF and the empirical distance variogram to simulated $S\alpha S$ RFs. The simulation of general $S\alpha S$ RFs is not trivial and only approximate simulation algorithms are known. One algorithm focusing on the simulation of $S\alpha S$ MA processes will be investigated in more detail below.



(a) $n = 75, a = 0.15$ (b) $n = 150, a = 0.15$



(c) $n = 75, a = 0.5$ (d) $n = 150, a = 0.5$

Figure 2.6: Empirical distance correlation as a function of the Euclidean distance, $\|h\|$. The underlying random processes are Gaussian RFs with exponential correlation function, and are simulated for different length scales a on grids with different sizes n. The true distance correlation functions are displayed as red curves.

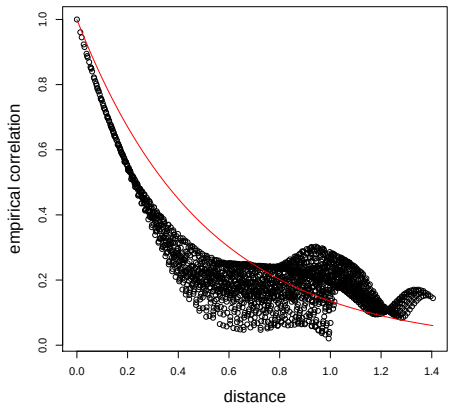


Figure 2.7: Empirical Pearson correlation function as a function of the Euclidean distance, $\|h\|$, for the simulated realization used in Figure 2.6 (d). The theoretical correlation function is displayed as the red curve.

In particular, we separate the class of $S\alpha S$ MA processes into *discrete* and *general* MA $S\alpha S$ RFs, because realizations of the latter can be approximated by realizations of the first.

The subsection closes with a treatment of sub-Gaussian processes, which are simple types of stable, but not necessarily MA, processes.

Linearly transformed $S\alpha S$ fields on a grid

By Theorem 2.56, the EDCF is a consistent estimators for the DCF for realizations of $S\alpha S$ processes that are q-short-range transformation fields. The simplest way of simulating such q-short-range transformation fields is to directly construct discrete moving average $S\alpha S$ processes with suitable weights on subsets of $\mathbb{Z}^d$. This case is of particular interest, because realizations of general MA $S\alpha S$ processes can be approximated by discrete sums of independent $S\alpha S$ random variables, which is a consequence of the construction of the stable integral.

Let $(\varepsilon(j))_{j\in\mathbb{Z}^d}$ be a collection of i.i.d. random variables, such that $\varepsilon(j)$ is $S\alpha S$ for some $\alpha > 1$. In other words, $\varepsilon(j)$ has characteristic function

$$\varphi_{\varepsilon(j)}(t) = \exp(-\sigma^\alpha |t|^\alpha), \qquad t \in \mathbb{R},$$

for some $\sigma > 0$. Consequently, the discrete MA process $\{Z(j) : j \in \mathbb{Z}^d\}$ defined as

$$Z(j) = \sum_{s \in \mathbb{Z}^d} a_s \varepsilon(j - s), \qquad a_s \in \mathbb{R}, \tag{2.34}$$

is a strictly stationary transformation field, if $\sum_{j \in \mathbb{Z}^d} |a_j|^\alpha = c < \infty$, i.e., if it fulfills (2.28). Now the characteristic function of the random variable $Z(j)$ is under the previous assumptions given by

$$\varphi_{Z(j)}(t) = \exp(-c\sigma^\alpha |t|^\alpha) \qquad \text{for } j \in \mathbb{Z}^d, \tag{2.35}$$

implying that the process Z additionally is $S\alpha S$. In this case, the physical dependence measure $\delta_{j,q}$ for $j \in \mathbb{Z}^d$ and $1 < q < \alpha$ equals

$$|a_j| \, \|\varepsilon_0 - \varepsilon_0'\|_q,$$

where ε_0' is an i.i.d. copy of ε_0. Thus, Z meets the q-short-range criterion (2.30) for any $1 < q < \alpha$ if the a_j are absolutely summable:

$$\sum_{j \in \mathbb{Z}^d} |a_j| < \infty.$$

We summarize these findings in the following remark.

Remark 2.61. Let $\{Z(j) : j \in \mathbb{Z}^d\}$ be a RF of the form (2.34) for $1 < \alpha \leq 2$ and let $1 < q < \alpha$. Then Z is a strictly stationary $S\alpha S$ RF satisfying the q-short-range criterion if

$$\sum_{j \in \mathbb{Z}^d} |a_j| < \infty.$$

Comparing the EDCF with the DCF is difficult for general $S\alpha S$ processes Z of the form (2.34), as illustrated in the following. In order to calculate the distance correlation, we need the characteristic functions (2.35) of the single random variable $Z(0)$, and the characteristic functions of the bivariate random vector $(Z(0), Z(k))^t$. Here, the latter can be expressed as

$$\exp\left(itZ(0) + irZ(k)\right) = \exp\left(it \sum_{s \in \mathbb{Z}^d} a_s \varepsilon(-s) + ir \sum_{s \in \mathbb{Z}^d} a_s \varepsilon(k - s)\right)$$

$$= \exp\left(it \sum_{s\in\mathbb{Z}^d} a_s \varepsilon(-s) + ir \sum_{s\in\mathbb{Z}^d} a_{s+k} \varepsilon(-s) \right)$$
$$= \exp\left(i \sum_{s\in\mathbb{Z}^d} (ta_s + ra_{s+k}) \varepsilon(-s) \right).$$

Because of the independence of the $\varepsilon(j)$, the characteristic function factorizes, i.e.,

$$\begin{aligned}\varphi_{Z(0),Z(k)}(t,r) &= \prod_{s\in\mathbb{Z}^d} \varphi_{\varepsilon(-s)}(ta_s + ra_{s+k}) \\ &= \prod_{s\in\mathbb{Z}^d} \exp\left(-|ta_s + ra_{s+k}|^\alpha\right) \\ &= \exp\left(- \sum_{s\in\mathbb{Z}^d} |ta_s + ra_{s+k}|^\alpha \right).\end{aligned}$$

Therefore, the square of the distance covariance function is proportional to

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \left[\exp\left(- \sum_{s\in\mathbb{Z}^d} |ta_s + ra_{s+k}|^\alpha \right) - \exp\left(- \sum_{s\in\mathbb{Z}^d} |a_s|^\alpha (|t|^\alpha + |r|^\alpha) \right) \right]^2 \frac{\mathrm{d}t\,\mathrm{d}r}{t^2 r^2}.$$

In some circumstances, by truncation of the involved infinite sums, it might be possible to simulate RFs of the form (2.34) with suitable accuracy. However, a precise numerical evaluation of the above integral could be infeasible in these cases, as the finite sums in the involved exponentials could still contain too many summands for fast computation.

To circumvent this problem, we consider more general $S\alpha S$ MA processes in the following subsection. Although these RFs can be approximated with arbitrary precision by linearly transformed fields on a grid, it is simpler to numerically evaluate their theoretical DCF, as compared to the theoretical DCF in the case of general linearly transformed fields. Nevertheless, we will see that such a numerical evaluation might still be a difficult problem.

General $S\alpha S$ moving average processes

For the analysis of the finite sample performance of the EDCF for general $S\alpha S$ MA processes we consider in the following two different compactly supported kernels.

The evaluation is based on a comparison to the numerically calculated DCF based on our considerations in Example 2.29. Assume in the following that the kernel function f is compactly supported. Considering compactly supported kernels has two advantages: first, realizations of $S\alpha S$ MA processes determined by them are easier to simulate and, secondly, we can theoretically identify the distance where the stochastic dependence becomes zero.

First, we turn to the simulation of $S\alpha S$ processes. Karcher, Scheffler, and Spodarev [56] showed that realizations of $S\alpha S$ RFs that are representable as stable integrals can be approximated as realizations of linearly transformed fields. We illustrate their approach with the example of an $S\alpha S$ MA RF $\{Z(x) : x \in \mathbb{R}^d\}$, determined by some compactly supported kernel function f as in (2.7).[17] Suppose we wish to simulate a realization of Z on the cube $[-T,T]^d \subset \mathbb{R}^d, T > 0$. Let $A > 0$, such that

$$\bigcup_{y\in[-T,T]^d} \operatorname{supp}(f_y) \subset [-A,A]^d,$$

where $f_y(x) = f(x-y)$ and $\operatorname{supp}(f_y) = \{x \in \mathbb{R}^d : f_y(x) \neq 0\}$ is its support. Further, define for each integer $n \geq 1$ and $k = (k_1, \ldots, k_d)^t \in \mathbb{Z}^d$ with $-n \leq k_1, \ldots, k_d < n$ the vector

$$\xi_k = \frac{A}{n} \cdot k,$$

and the set

$$\Delta_k = \left[k_1\frac{A}{n}, (k_1+1)\frac{A}{n}\right) \times \cdots \times \left[k_d\frac{A}{n}, (k_d+1)\frac{A}{n}\right).$$

Then with the help of

$$\widehat{f}_y^{(n)}(x) = \sum_{-n\leq k_1,\ldots,k_d<n} f_y(\xi_k)\mathbf{1}_{\Delta_k}(x),$$

a realization of $Z(x)$ can be approximated by

$$\widehat{Z}^{(n)}(x) := \int_{[-A,A]^d} \widehat{f}_y^{(n)}(x)\, M(\mathrm{d}x) = \sum_{-n\leq k_1,\ldots,k_d<n} f_y(\xi_k) M(\Delta_k).$$

From the properties of M and the fact that $\Delta_k \cap \Delta_j = \emptyset$ for $k \neq j$ it follows

[17] Karcher, Scheffler, and Spodarev [56] provided slightly modified algorithms that also work for kernel functions without compact support.

that $M(\Delta_k)$ are independent $S\alpha S$ random variables for $-n \leq k_1, \ldots, k_d < n$. Consequently, the field $\widehat{Z}^{(n)}$ is a linearly transformed $S\alpha S$ RF, whose simulation only requires the simulation of independent $S\alpha S$ random variables and evaluations of the kernel f.

Naturally, it is interesting to investigate the goodness of such an approximate realization. To this end, Karcher, Scheffler, and Spodarev [56] provided error bounds for kernels f_y that are *Hölder continuous*,

$$|f_y(x) - f_y(x_0)| \leq C\|x - x_0\|^\delta, \qquad x, x_0 \in [-T, T]^d,$$

for some $C > 0$ and $0 < \delta \leq 1$. They showed in this case that the error in p-th mean ($0 < p < \alpha$) is for $0 < \alpha < 1$ bounded via

$$\left(\mathbb{E}|Z(x) - \widehat{Z}^{(n)}(x)|^p\right)^{1/p} \leq C 2^{d/\alpha} \left(\frac{d}{1+\alpha\delta}\right)^{1/\alpha} A^{\delta+d/\alpha} n^{-\delta_y}, \qquad x \in [-T, T]^d,$$

and for $1 < \alpha \leq 2$ via

$$\left(\mathbb{E}|Z(x) - \widehat{Z}^{(n)}(x)|^p\right)^{1/p} \leq C 2^{d} \left(\frac{d}{1+\alpha\delta}\right)^{1/\alpha} A^{\delta+d/\alpha} n^{d-d/\alpha-\delta}, \qquad x \in [-T, T]^d.$$

As one can see, the error bound for $0 < \alpha < 1$ converges to zero for $n \to \infty$, while the error bound for the case $1 < \alpha \leq 2$ converges only to zero if

$$\frac{\delta\alpha}{\alpha - 1} > d. \tag{2.36}$$

This has the following practical implications. Assume that $1 < \alpha \leq 2$ and suppose that we want to simulate a realization of an $S\alpha S$ MA RF Z on a domain D. If the error in the p-th mean converges to zero then also the mean absolute error $\mathbb{E}|Z(x) - \widehat{Z}^{(n)}(x)|$ for each $x \in D$. Hence, by using an n large enough, we can assume that the errors coming from the approximate simulation can be neglected.

Concerning the second advantage mentioned above, we have the following remark.

Remark 2.62. Let $f \in L^\alpha$ be a kernel function whose support is contained in the Ball $B_r(0) = \{x \in \mathbb{R}^2 : \|x\| \leq r\}$ and let $\{Z(x) : x \in \mathbb{R}^2\}$ be an $S\alpha S$ MA RF generated by f. Then $Z(x)$ and $Z(x + h)$ are stochastically independent for all $x \in \mathbb{R}^2$ if $\|h\| > 2r$.

This is evident from (2.20), because $|t\,f(h+x)+s\,f(x)| = |t\,f(h+x)|+|s\,f(x)|$ for all $x \in \mathbb{R}^2$ if $\|h\| > 2r$.

After these considerations, we now turn our attention towards actual examples of simulations and the application of the EDCF as well as the empirical distance variogram.

Example 2.63. In this example, $\{Z(x) : x \in \mathbb{R}^2\}$ is always an $S\alpha S$ MA RF generated by the isotropic Epanechnikov kernel

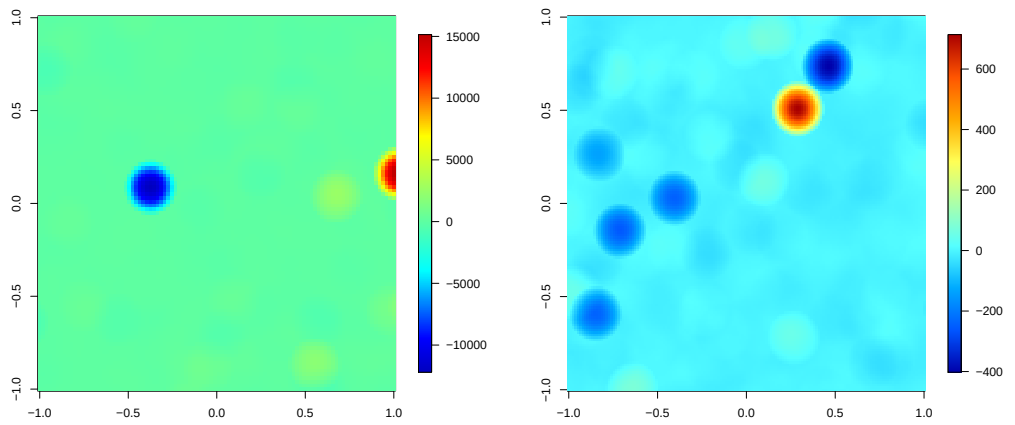
$$f(x) = \begin{cases} b(a^2 - \|x\|^2), & \|x\| \le a, \\ 0, & \|x\| > a, \end{cases}$$

whose support is the ball $B_a(0)$. Hence, the preceding remark yields $\mathrm{C}^{\mathrm{ds}}(h) = 0$ if $\|h\| > 2a$. Another feature that makes the Epanechnikov kernel suitable for simulation purposes is its Hölder-continuity. As noted in [56], it is Hölder-continuous with parameter $C = 2ab$ and $\delta = 1$, implying via (2.36) that the error in the p-th mean converges to zero in the case $d = 2$ and $1 < \alpha < 2$.

(1) Assume $a = 0.15, b = 1$, and $\alpha = 1.1$. To compare the theoretical values for the DCF with the empirical values, we simulate a realization of Z by using the approximate simulation algorithm described above. The simulated values are located on the grid

$$\Gamma = \{-1, -1+\Delta, -1+2\Delta, \ldots, 1-2\Delta, 1-\Delta, 1\}^2, \qquad \Delta = \frac{2}{99}.$$

Figure 2.8 panel (a) shows a realization for Z. There one can see circular structures, caused by the isotropic nature of the Epanechnikov kernel. It is striking that the realization admits values with very high absolute values. Evidently, this is due to the heavy-tail property of the underlying stable distribution. The corresponding EDCF is displayed in Figure 2.9. In panel (a) it is shown as a function of the lag vectors. It shows the expected behaviour: the empirical distance correlation equals one at the origin, is radially symmetric (due to the isotropy of the field) and decreases towards zero for increasing distances. In panel (b), the values of the EDCF are plotted against Euclidean distances. The solid line shows the theoretical DCF, which was numerically evaluated using `Matlab`. We see that the empirical values match the theoretical ones with a satisfying accuracy. The fluctuations in the theoretical



(a) (b)

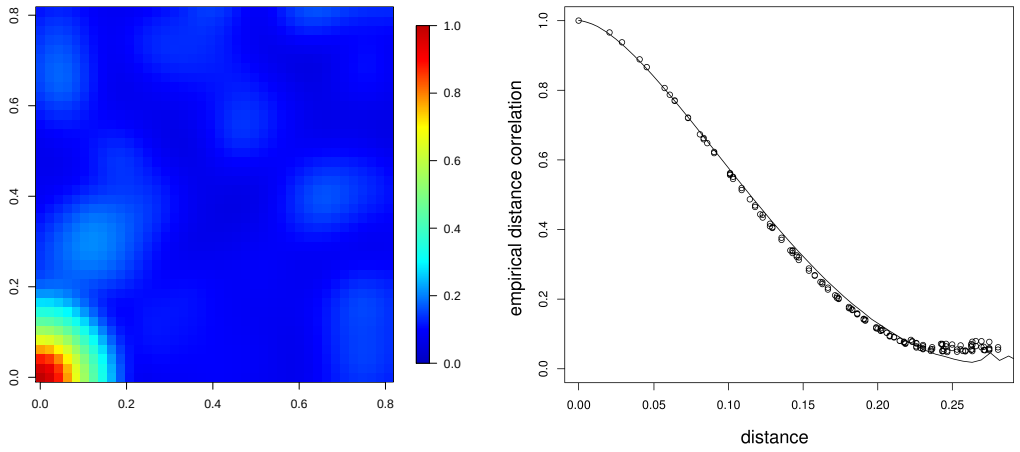
Figure 2.8: Two realizations of an $S\alpha S$ MA RF Z with (a) index $\alpha = 1.1$ and (b) index $\alpha = 1.3$.

DCF for larger distances are due to the fact that numerical evaluation of the theoretical DCF is difficult for the case $\alpha = 1.1$ because the integrator admits a singularity at the point $(t, s)^t = (0, 0)^t$. To make the numerical integration precise in this case, the algorithm needs to explore the singularity carefully, which is not trivial and requires lengthy computation time.

(2) After comparing the theoretical DCF with the EDCF in part (1), we study more closely in this part the distance variogram. We want to determine whether it is possible to derive some properties of the underlying process if we assume only that it is an $S\alpha S$ MA process with index $1 < \alpha \leq 2$. To this end, we simulate a realization of Z using parameters the same parameters $a = 0.15$ and $b = 1$ as in the previous part, but now we take the slightly higher index $\alpha = 1.3$. Further, the simulated values are located on the denser grid

$$\Gamma = \{-1, -1 + \Delta, -1 + 2\Delta, \ldots, 1 - 2\Delta, 1 - \Delta, 1\}^2, \qquad \Delta = \frac{2}{149}.$$

Note that the empirical distance variogram is a consistent estimator in this



(a) (b)

Figure 2.9: Empirical distance correlation function depending on (a) the lag vector and (b) the Euclidean distance for the realization shown in Figure 2.8 (a).

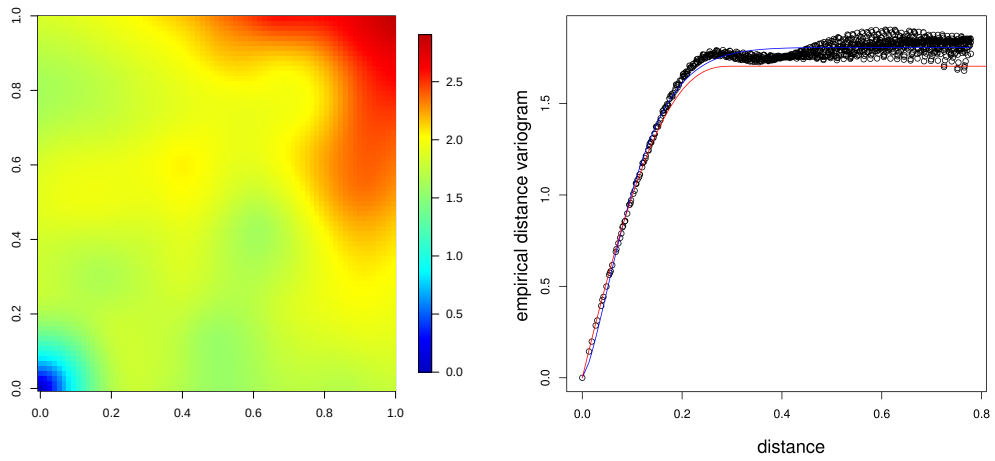
case since the kernel function has a finite support.

Panel (b) in Figure 2.8 displays the realization. It shows structures similar to the realization in panel (a), but with smaller amplitudes caused by the higher index.

Now we assume only that the underlying process is an $S\alpha S$ MA process. We calculate its normalized empirical distance variogram, which is depicted in Figure 2.10. In panel (a), the empirical normalized distance variogram is displayed for different lag vectors and from this it is plausible that the underlying increment process is pairwise isotropic. Panel (b) shows the empirical normalized distance variogram plotted against Euclidean distance. In addition, in panel (b) we fitted a powered exponential variogram model

$$\sigma \left(1 - \exp\left(-\|h/a\|^{\theta}\right)\right), \qquad h \in \mathbb{R}^2,$$

by using least squares estimation. Note that this is a conditionally negative definite function for $\sigma > 0, a > 0$, and $\theta \in (0, 2]$ and, thus, fitting this model



(a) (b)

Figure 2.10: Empirical distance variogram depending on (a) the lag vector and (b) the Euclidean distance, $\|h\|$, for the stable RF in Figure 2.8. The red curve in panel (b) shows the fitted powered exponential model.

is in accordance with Theorem 2.51. The fitted parameters are $\sigma = 1.81, a = 0.11$, and $\theta = 1.42$. Using Theorem 2.49 it is possible to estimate the index of the underlying process, yielding in this case 1.17. This is smaller than the real index 1.3 but not very far off. Furthermore, the estimated scale parameter is close to the true parameter $a = 0.15$. However, the estimated power θ is difficult to interpret and we do not know of any satisfying interpretation.

It must be noted that the procedure in this example is not fully theoretically justified, but it also does not contradict our theoretical results. A comparison with the true normalized distance variogram in panel (b) of Figure 2.10 shows that the fitted distance variogram differs from the true one only slightly for small distances, but it overestimates the limiting value significantly. Nevertheless, given enough data, it is possible to infer properties of the underlying process if one assumes that it is an $S\alpha S$ MA process.

In the following example we consider an anisotropic kernel function.

Example 2.64. Let $\{Z(x) : x \in \mathbb{R}^2\}$ be an $S\alpha S$ MA RF with index $\alpha = 1.6$ and

kernel $g : \mathbb{R}^2 \to \mathbb{R}$ given by

$$g(x) = \begin{cases} (0.15 - |x_1|)(0.15 - |x_2|), & |x_1| \leq 0.15,\ |x_2| \leq 0.15, \\ 0, & \text{else.} \end{cases}$$

Karcher, Scheffler, and Spodarev [56] noted that g is Hölder-continuous with parameters $C = \sqrt{0.6}$ and $\delta = 1$. Thus, the error in p-th mean of the simulated values of Z converges to zero in the case $d = 2$ and $1 < \alpha < 2$, analogously to the previous example. A simulated realization of Z is depicted in panel (a) of Figure 2.11. The anisotropy of the RF can be seen from the angular-shaped structures, caused by the anisotropy of the kernel g. Panel (b) shows the empirical normalized distance variogram depending on the lag vectors. Panels (c) and (d) show the true normalized distance variogram and the difference between the true and the empirical normalized distance variogram, respectively. The true normalized distance variogram was calculated using Lemma 2.53 and the `R` package `cubature`. It can be seen that the anisotropy makes the estimation less precise, compared to the isotropic case. The estimation is best for small distances and there it is also possible to recognize the anisotropy of the process.

Sub-Gaussian RFs

Sub-Gaussian RFs are simple types of α-stable RFs and have been studied by several researchers, see for example [91, Section 3.7] or [98, Chapter 11]. These processes are defined in the following way. Let A be a stable random variable with index $\alpha/2$, scale parameter $\cos(\pi\alpha/4)^{2/\alpha}$, skewness parameter 1 and shift parameter 0. If $\{G(x) : x \in \mathbb{R}^d\}$ is a Gaussian RF independent of A, then the RF

$$Z(x) = A^{1/2} G(x), \qquad x \in \mathbb{R}^d,$$

is called *sub-Gaussian*. The definition implies that for every $x_1, \ldots, x_m \in \mathbb{R}^d$ the random vector $(Z(x_1), \ldots, Z(x_m))^t$ has an $S\alpha S$ distribution.

Assume that the underlying Gaussian field G is strictly stationary with zero mean, covariance function C and correlation function ρ, such that $\rho(h) = C(h)/C(0) = C(h)/\sigma^2$ for some $\sigma > 0$. Then Proposition 2.5.2 in [91] shows that the characteristic functions equal

$$\varphi_{Z(x),Z(x+h)}(t, s) = \exp\left(-\frac{1}{2}\sigma^\alpha |t^2 + 2\rho(h)ts + s^2|^{\alpha/2}\right),$$

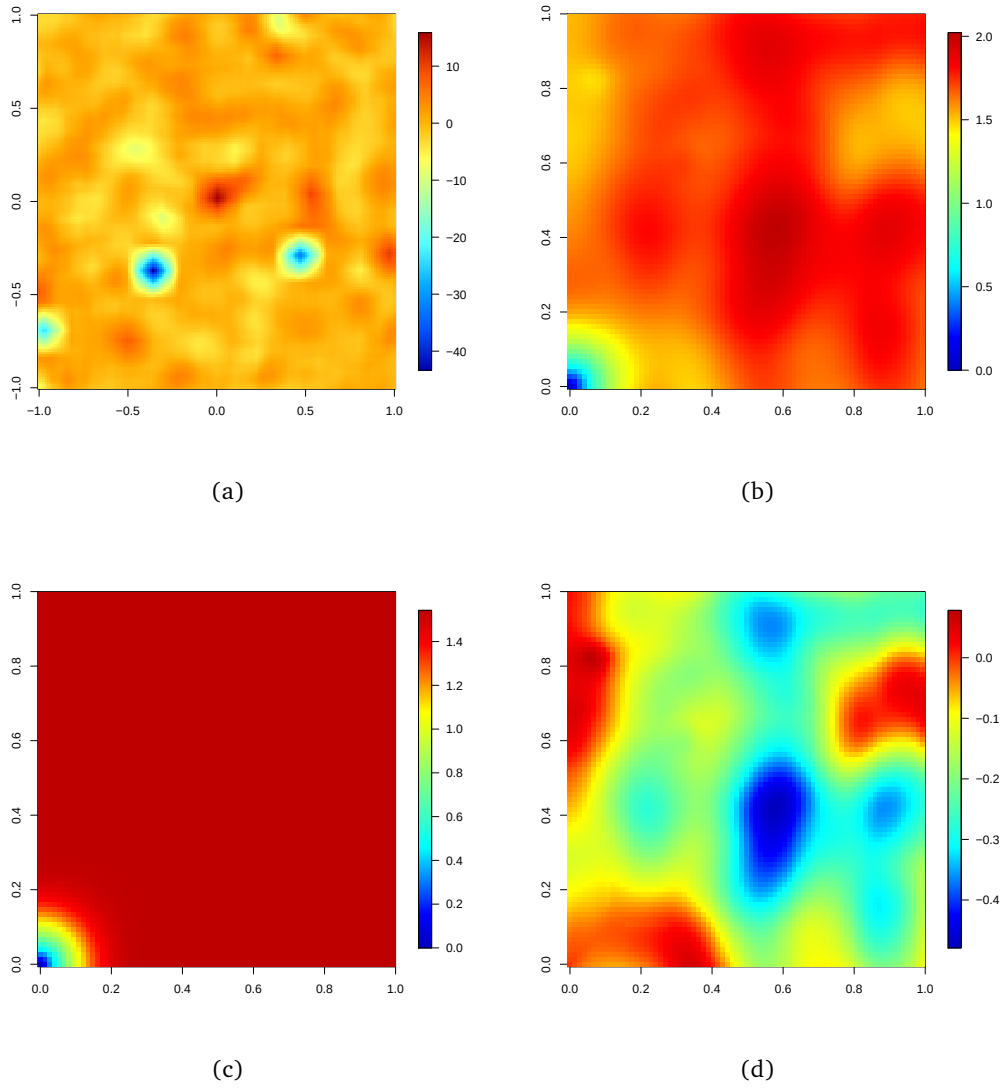


Figure 2.11: Realization of an anisotropic $S\alpha S$ MA RF in panel (a) and its corresponding empirical normalized distance variogram in panel (b). Panels (c) and (d) show the theoretical normalized distance variogram and the absolute value of the difference between theoretical and empirical normalized distance variogram.

$$\varphi_{Z(x)}(t) = \exp\left(-\frac{1}{2}(\sigma|t|)^\alpha\right),$$

and

$$\varphi_{Z(x+h)}(s) = \exp\left(-\frac{1}{2}(\sigma|s|)^\alpha\right).$$

Consequently, the distance covariance function equals

$$\begin{aligned}\mathcal{V}(h)^2 = \frac{1}{c_m}\int_{\mathbb{R}}\int_{\mathbb{R}}\Bigg|&\exp\left(-\frac{1}{2^{\alpha/2}}\sigma^\alpha|t^2+2\rho(h)ts+s^2|^{\alpha/2}\right)\\ &-\exp\left(-\frac{1}{2^{\alpha/2}}\left((\sigma|t|)^\alpha+(\sigma|s|)^\alpha\right)\right)\Bigg|^2\frac{\mathrm{d}t}{t^2}\frac{\mathrm{d}s}{s^2}\\ =\frac{\sigma^2}{c_m}\int_{\mathbb{R}}\int_{\mathbb{R}}\Bigg|&\exp\left(-\frac{1}{2^{\alpha/2}}|t^2+2\rho(h)ts+s^2|^{\alpha/2}\right)\\ &-\exp\left(-\frac{1}{2^{\alpha/2}}\left(|t|^\alpha+|s|^\alpha\right)\right)\Bigg|^2\frac{\mathrm{d}t}{t^2}\frac{\mathrm{d}s}{s^2}.\end{aligned}$$

Dividing by

$$\begin{aligned}\mathcal{V}(0)^2 = \frac{\sigma^2}{c_m}\int_{\mathbb{R}}\int_{\mathbb{R}}\Bigg|&\exp\left(-\frac{1}{2^{\alpha/2}}|t^2+2ts+s^2|^{\alpha/2}\right)\\ &-\exp\left(-\frac{1}{2^{\alpha/2}}\left(|t|^\alpha+|s|^\alpha\right)\right)\Bigg|^2\frac{\mathrm{d}t}{t^2}\frac{\mathrm{d}s}{s^2}\\ =\frac{\sigma^2}{c_m}\int_{\mathbb{R}}\int_{\mathbb{R}}\Bigg|&\exp\left(-\frac{1}{2^{\alpha/2}}|t+s|^\alpha\right)-\exp\left(-\frac{1}{2^{\alpha/2}}\left(|t|^\alpha+|s|^\alpha\right)\right)\Bigg|^2\frac{\mathrm{d}t}{t^2}\frac{\mathrm{d}s}{s^2},\end{aligned}$$

yields the DCF, which depends on α. This implies that we cannot consistently estimate the distance correlation by using a single realization of the sub-Gaussian field. This holds true because a single realization of a sub-Gaussian field cannot be distinguished from a realization of a Gaussian field and, therefore, any statistics depending on α cannot be consistently estimated based on a single realization of a sub-Gaussian field.

As a conclusion, the statement of Theorem 2.56 cannot be true in this case, which can be explained by the fact that a sub-Gaussian field is not q-short-range, i.e., it does not fulfill (2.30).

2.6.4 Max-stable fields

After investigating the distance variogram for $S\alpha S$ processes, it is interesting to ask if the empirical distance variogram can be used to analyze the dependence structure of max-stable processes. Unfortunately, Corollary 2.60 is not applicable in this case, as it is not clear that the theoretical distance variogram is defined for max-stable processes. This is due to the fact that max-stable processes have no finite first moments and we have no analogous result to Theorem 2.45, because the characteristic functions for increment processes of max-stable processes is not known to us. Nevertheless, we apply experimentally the empirical distance variogram to realizations of max-stable fields and then consider whether the behaviour seems plausible.

In the following we use a max-stable process given by a certain Schlather model [96] defined as follows. Let $\{Y(x) : x \in \mathbb{R}^d\}$ be a second-order stationary Gaussian RF with correlation function ρ, with $\mathbb{E}\left[\max\{0, Y(0)\}\right] = \mu > 0$. Furthermore, let Π be a Poisson process on $(0, \infty)$ with intensity measure $\mathrm{d}\Lambda(s) = \mu^{-1} s^{-2}\,\mathrm{d}s$ and let the processes $Y_s, s \in \Pi$, be independent replications of Y. Then Theorem 2 in [96] implies that

$$Z(x) = \sup_{s \in \Pi} s Y_s(x), \qquad x \in \mathbb{R}^d,$$

is a strictly stationary max-stable RF, such that the distribution of $Z(0)$ is a Fréchet distribution with parameter 1, i.e.,

$$F_{Z(0)}(t) = \exp(-t^{-1}).$$

Now we take the extreme value index $\mu = 1$ and simulate a realization on the regular grid $\Gamma \subset [0, 10]^2$ with 150×150 points. The underlying Gaussian RF Y has the isotropic damped cosine correlation function

$$\rho(h) = \exp(-\|h\|)\cos(\|h\|), \qquad h \in \mathbb{R}^2.$$

Figure 2.12 shows a realization of the max-stable RF and the corresponding empirical distance variogram. This nicely illustrates the expected dependence structure: it increases isotropically from zero to a certain limiting value. Figure 2.13 shows the empirical distance variogram as a function of the Euclidean distance. Interestingly, the empirical distance variogram displays a fluctuating behavior for distances $\|h\|$ where $\rho(h) < 0$, i.e., $1 - \rho(h) > 1$. This might be explained as follows. Without loss of generality we assume that $Y(0) > 0$. If the random variables $Y(0)$ and $Y(h)$

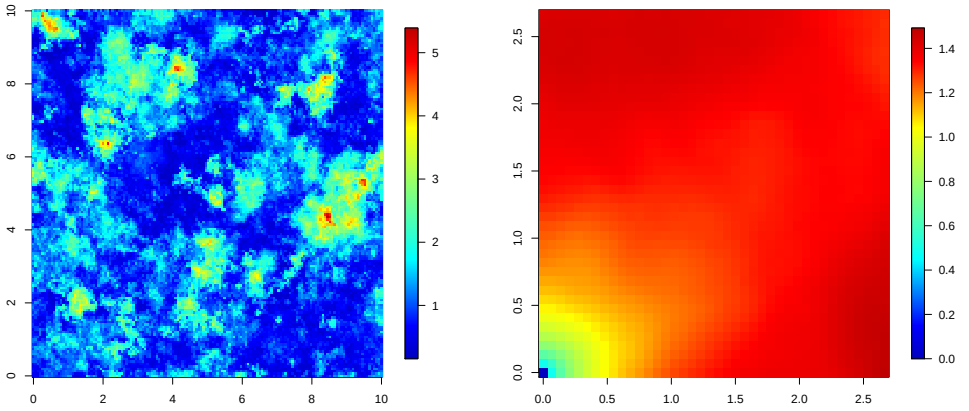


Figure 2.12: The left panel shows a realization of a max-stable random field following a Schlather model. The right panel shows the corresponding empirical distance variogram.

are negatively correlated then there is a positive probability that $Y_s(h) < 0$ for all $s \in \Pi$. If this is the case then $Z(h)$ is determined by small values of $s \in \Pi$. If many $Y_s(h)$ are positive (which has also positive probability), the behaviour of $Z(h)$ is determined by high values of $s \in \Pi$, hence, points $Z(x)$ with the same distance to the origin might behave differently, and this phenomenon is reflected in the distance variogram.

2.7 Conclusion and outlook

In this chapter, we introduced the distance correlation function and the distance variogram for spatial processes and investigated their properties. We illustrated that these dependence measures can be applied to a wide range of processes and that they yield valid information about the underlying dependence structure, even when it is highly non-linear or the underlying process possesses heavy tails. This, in combination with the consistency of the empirical estimators, makes these dependence measures attractive for conducting inference with spatial data sampled on regularly spaced grids, where the data generating processes are unknown. In such cases, these dependence measures can help detect many aspects of the dependence

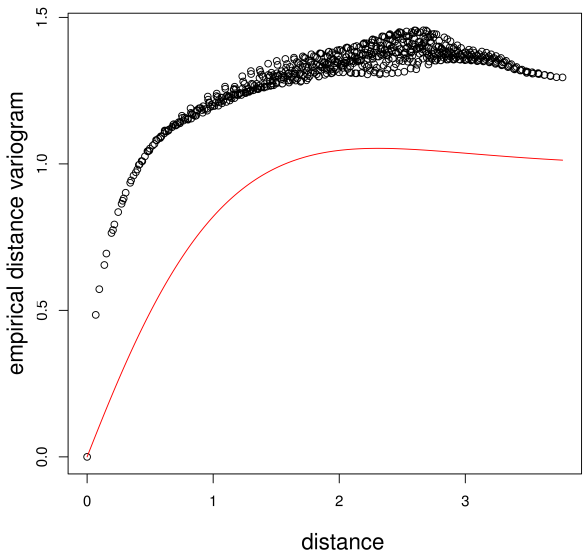


Figure 2.13: The circles show the empirical distance variogram as a function of Euclidean distance. The red curve shows $1-\rho$, where ρ is the correlation function of the underlying Gaussian field.

structure that might be overlooked by the Pearson correlation function.

In a simulation study, we showed that the empirical versions of these dependence measures work well on spatial data sets with at least $10\,000$ points, if the length scale of the spatial dependence is significantly smaller than one-half the width of the observed area. To extend the applicability of the estimators, it would be desirable to generalize them to irregularly spaced spatial data. This can be done by binning the data, analogously to the case of the usual variogram. However, it remains to be shown that the resulting estimators are consistent.

Another interesting aspect for future research might be the classification of admissible DCFs or distance variograms. If one knows classes of such functions, it would be possible to fit members of them to empirical versions and this would provide further insight into the underlying process. To this end, it would be advantageous to derive analytical expressions for the DCF or distance variogram for processes other than Gaussian processes. This is in general a difficult problem, but it might be feasible, for example, for classes of $S\alpha S$ MA processes.

A more difficult task would be the investigation of other heavy-tailed processes. For example, does the DCF or the distance variogram exist for certain max-stable processes? If they exist, are the empirical estimators consistent, and do they provide new information in comparison to established dependence measures, such as the extremogram [20] or the tail correlation function [106]?

In addition, the dependence measures might be helpful not only for inference, but also for extrapolation tasks. Spodarev and co-authors investigated the extrapolation of stable RFs in [98, Section 11.4]. It might be possible to connect aspects of their work with the concept of the DCF or distance variogram and to apply them in problems on extrapolations.

One should be aware that the distance correlation function does not provide a comparable power for modelling and extrapolation for heavy-tailed processes, as the Pearson correlation function for Gaussian processes. This lies in the nature of these processes: Gaussian processes are completely determined by the mean and the correlation function, while an analogous result does not hold true in general for heavy-tailed processes and the distance correlation function.

Most of the simulations and estimations in this chapter were performed by using `R`. As there are no packages available providing functions for the simulation of stable processes or for the calculation of the EDCF,[18] we wrote our own functions for simulation and estimation. It is planned to make this code available within an `R` package, for example the package `RandomFields`.

[18] The `energy` package provides functions for calculating the distance correlation. However, these functions are much too slow for our purposes.

3 Relating Schoenberg coefficients in Gegenbauer expansions on spheres

Every continuous real-valued function on the interval $[-1,1]$ can be represented as an infinite series in Gegenbauer polynomials C_n^λ (cf. Szegö [107, Chapter 3]). Thus any continuous function $\xi : [0,\pi] \to \mathbb{R}$ satisfying $\xi(0) = 1$ admits for every integer $d \geq 1$ the *d-Gegenbauer expansion*

$$\sum_{n=0}^{\infty} b_{n,d} \frac{C_n^{(d-1)/2}(\cos(\theta))}{C_n^{(d-1)/2}(1)}, \qquad \theta \in [0,\pi]. \tag{3.1}$$

Referring to Daley and Porcu [19] and Ziegel [118] we call the coefficients $b_{n,d}$ the *d-dimensional Schoenberg coefficients of ξ*.

For $\lambda \geq 0$, the Gegenbauer polynomials C_n^λ, also called ultraspherical polynomials, are defined as polynomials of degree n, orthogonal on the interval $[-1,1]$ with respect to the weight function $(1-x^2)^{\lambda-1/2}$, see [2, equation (22.2.3)]. In other words,

$$\int_{-1}^{1} C_n^\lambda(x) C_m^\lambda(x) (1-x^2)^{\lambda-1/2} \,\mathrm{d}x = \delta_{nm} \frac{\pi 2^{1-2\lambda} \Gamma(n+2\lambda)}{n!(n+\lambda)\Gamma(\lambda)^2},$$

where δ_{nm} is the Kronecker-delta. If $\lambda > 0$ then the Gegenbauer polynomials can also be expressed as

$$C_n^\lambda(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k \frac{\Gamma(n-k+\lambda)}{\Gamma(\lambda) k! (n-2k)!} (2x)^{n-2k},$$

cf. [2, equation (22.3.4)]. In the case $\lambda = 0$ we follow Schoenberg [100] and set $C_n^0(\cos\theta) = \cos(n\theta)$ for $\theta \in [0,\pi]$.

Our motivation for studying d-Gegenbauer expansions and their d-dimensional Schoenberg coefficients arises from the theory of isotropic positive definite functions on spheres, as we will explain in the following.

For an integer $d \geq 1$, we write $\mathbb{S}^d := \{x \in \mathbb{R}^{d+1} : \|x\| = 1\}$ to denote the unit

sphere in Euclidean space $\mathbb{R}^{d+1}$ equipped with the usual Euclidean norm. Consider a kernel $h : \mathbb{S}^d \times \mathbb{S}^d \to \mathbb{R}$. The kernel h is called

- *isotropic*, if there exists a function $\xi : [0, \pi] \to \mathbb{R}$ such that

$$h(x, y) = \xi(\theta(x, y)), \qquad x, y \in \mathbb{S}^d,$$

  where, $\theta(x, y) = \arccos(\langle x, y\rangle)$ denotes the great circle distance between x and y and $\langle \cdot, \cdot \rangle$ the standard scalar product in $\mathbb{R}^{d+1}$;

- *positive definite*, if

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j h(x_i, x_j) \geq 0, \tag{3.2}$$

  for all integers $n \geq 1$ and for every choice of constants $a_1, \ldots, a_n \in \mathbb{R}$ and every choice of pairwise distinct points $x_1, \ldots, x_n \in \mathbb{S}^d$. If the inequality (3.2) is strict, except for the case $a_1 = \cdots = a_n = 0$, the kernel is called *strictly positive definite*.

We refer to Ψ_d (respectively Ψ_d^+), $d = 1, 2, \ldots$, as the class of continuous functions $\psi : [0, \pi] \to \mathbb{R}$ with $\psi(0) = 1$ for which the associated isotropic kernel $h(x, y) = \psi(\theta(x, y))$ is positive definite (respectively strictly positive definite).

Isotropic positive definite functions on spheres have attracted interest in several areas. They occur as correlation functions for stationary and isotropic RFs on the sphere [52] and, hence, have been studied in spatial statistics [7, 50, 47]. Furthermore, they are radial basis functions for interpolating scattered data on spherical domains, see for example [115, 30, 76, 113, 11]. Recently, Gneiting [40, 41] reviewed conditions for functions to belong to Ψ_d or Ψ_d^+, and used them to study parametric families of isotropic and stationary correlation functions on spheres. In Gneiting's work, he stated several problems for future research, one of which was solved by Ziegel [118], and the solution to another is given here.

Schoenberg [100], Chen, Menegatto, and Sun [12], and Menegatto [73] showed that the members of Ψ_d and Ψ_d^+ are characterized by their d-Gegenbauer expansions. In particular, the class $\Psi_d,\ d \geq 1$, consists of the functions of the form (3.1) with $b_{n,d} \geq 0$ and $\sum_{n=0}^{\infty} b_{n,d} = 1$. For $d \geq 2$, the class Ψ_d^+ consists of those functions in Ψ_d for which $b_{n,d} > 0$ for infinitely many even and infinitely many odd integers n. Consequently, it is possible to study properties of the members of Ψ_d or Ψ_d^+ via the coefficients $b_{n,d}$ of their d-Gegenbauer expansions (3.1). As an example, for

the cases $d = 1$ and $d = 2$ it was shown by Lorentz [66] and Lang and Schwab [63], respectively, that Hölder continuity and differentiability of a function in Ψ_d are related to the decay rate of $b_{n,d}$, as $n \to \infty$.

Since $\mathbb{S}^1 \subset \mathbb{S}^2 \subset \mathbb{S}^3 \ldots$, it is evident that

$$\Psi_1 \supseteq \Psi_2 \supseteq \Psi_3 \supseteq \cdots . \tag{3.3}$$

This, in concert with $C_n^0(\cos\theta) = \cos(n\theta)$ for every integer $n \geq 0$ and $\theta \in [0, \pi]$, shows that every function in $\Psi_d,\ d \geq 1$, possesses a 1-Gegenbauer expansion, which is a well-known Fourier cosine expansion. A similar statement holds true for $d \geq 2$: because $C_n^{1/2} = P_n$ is a Legendre polynomial, every function $\psi \in \Psi_d$ satisfies an expansion in terms of Legendre polynomials. This raises the issue of how to express higher-dimensional Schoenberg coefficients in terms of Fourier or Legendre coefficients. In other words, given a particular Gegenbauer expansion (3.1), we wish to determine whether there exist so-called *connection coefficients* $a_i(n,k)$, respectively $u_i(n,k)$, for the Schoenberg coefficients $b_{n,d}$ in odd, respectively even, dimension d, such that

$$b_{n,2k+1} = \sum_{i=0}^{k} a_i(n,k) b_{n+2i,1}, \qquad \text{and} \qquad b_{n,2k+2} = \sum_{i=0}^{k} u_i(n,k) b_{n+2i,2}.$$

Connection coefficients have been investigated in the study of orthogonal polynomials, as exemplified by Maroni and Rocha [71]. Their aim was to express a monic polynomial P_n of degree n as a linear combination of monic polynomials $\widetilde{P}_m$ of degree $m \leq n$, i.e., to find connection coefficients α_{mn} such that

$$P_n(x) = \sum_{m=0}^{n} \alpha_{mn} \widetilde{P}_m(x).$$

Gneiting [40] used the following connection between Gegenbauer polynomials, stated as equation (3.42) of Askey and Fitch [5],

$$C_n^{\lambda}(x) = \frac{\Gamma(\nu)}{\Gamma(\lambda)\Gamma(\lambda-\nu)} \sum_{k=0}^{[n/2]} \frac{(n-2k+\nu)\Gamma(k+\lambda-\nu)\Gamma(n-k+\lambda)}{k!\Gamma(n-k+\nu+1)} C_{n-2k}^{\nu}(x)$$

to show that the inclusions in (3.3) are strict.

The study of connection coefficients for Schoenberg coefficients can be helpful to decide for a certain dimension d whether or not a function ψ belongs to Ψ_d or

Ψ_d^+. Given a function $\psi \in \Psi_1$ with Schoenberg coefficients $b_{n,1}$, knowing the connection coefficients $a_i(n,k)$ allows us to find higher odd-dimensional Schoenberg coefficients $b_{n,2k+1}$. A similar procedure has been used in [45] for $d = 2, 3$ to show positive definiteness of certain functions.

The problem of expressing even- and odd-dimensional Schoenberg coefficients in terms of Fourier and Legendre coefficients can be answered using the following recursive identities, stated as Corollary 3 in [40].[1] For all integers $n \geq 1$ it is true that

$$b_{0,3} = b_{0,1} - \frac{1}{2} b_{2,1} \qquad \text{and} \qquad b_{n,3} = \frac{1}{2}(n+1)(b_{n,1} - b_{n+2,1}). \tag{3.4}$$

Furthermore, if $d \geq 2$, then for all integers $n \geq 0$

$$b_{n,d+2} = \frac{(n+d-1)(n+d)}{d(2n+d-1)} b_{n,d} - \frac{(n+1)(n+2)}{d(2n+d+3)} b_{n+2,d}. \tag{3.5}$$

These recursive relationships show that it is possible to express $b_{n,2k+1}, k \geq 1$, as a linear combination of Fourier coefficients $b_{n,1}, b_{n+2,1}, \ldots, b_{n+2k,1}$. Similarly, we can express $b_{n,2k+2}, k \geq 1$, as a linear combination of Legendre coefficients $b_{n,2}, b_{n+2,2}, \ldots, b_{n+2k,2}$. If a continuous function $\xi : [0, \pi] \to \mathbb{R}$ with $\xi(0) = 1$ admits a d-Gegenbauer expansion, (3.4) and (3.5) show that ξ can also be expanded as a Fourier cosine series if $d > 1$ is odd, or a series in Legendre polynomials, if $d > 2$ is even.

The aim of this work is to provide closed form expressions for the coefficients appearing in these linear combinations, which was stated as Problem 1 in Gneiting [41].

3.1 Explicit relations between Schoenberg coefficients

In this section we provide explicit expressions for Schoenberg coefficients in terms of Fourier cosine and Legendre coefficients. The proofs are given in Section 3.3.

Here and in the following, $(x)_m = x(x+1)\cdots(x+m-1)$ denotes the *raising factorial*.

[1] Note that the proof does not require the Schoenberg coefficients to belong to the d-Gegenbauer expansion of a positive definite kernel. Hence, the recursive identities hold also true if a continuous function $\xi : [0, \pi] \to \mathbb{R}$ admits a d-Gegenbauer expansion.

Theorem 3.1. Let $k \geq 1$ be an integer and $\xi : [0, \pi] \to \mathbb{R}$ continuous with $\xi(0) = 1$. Further, assume that $\xi(\theta)$ is given by the $(2k+1)$-Gegenbauer expansion (3.1) for every $\theta \in [0, \pi]$. Then for every integer $n \geq 0$, the $(2k+1)$-Schoenberg coefficient $b_{n,2k+1}$ and the Fourier cosine coefficients $b_{n,1}, b_{n+2,1}, \ldots, b_{n+2k,1}$ are connected by

$$b_{n,2k+1} = \sum_{i=0}^{k} a_i(n,k)\, b_{n+2i,1},$$

with

$$a_i(n,k) = \frac{(-1)^i}{2^k} \binom{k}{i} \frac{(n+k)(n+2i)}{(2k-1)!!} \frac{(n+1)_{2k-1}}{(n+i)_{k+1}}, \tag{3.6}$$

for $(i,n) \neq (0,0)$, and $a_0(0,k) = 1$.

The theorem shows that $a_0(n,k)$ and $a_k(n,k)$ can be expressed in simpler forms: for general $k \geq 0$, equation (3.6) reduces for $i = 0$ and $i = k$ to

$$a_0(n,k) = \frac{1}{2^k(2k-1)!!}(n+k)_k$$

and

$$a_k(n,k) = \left(-\frac{1}{2}\right)^k \frac{1}{(2k-1)!!}(n+1)_k,$$

respectively.

It is interesting to note that the value of $\sum_{i=0}^{k} a_i(n,k)$ is either 0 or $1/2$, as follows.

Proposition 3.2. For all integers $k \geq 1$ it is true that

$$\sum_{i=0}^{k} a_i(n,k) = \begin{cases} 0, & n > 0, \\ \frac{1}{2}, & n = 0. \end{cases}$$

Now let us turn to the analogous problem of finding an expression for $b_{n,2k+2}, k \geq 1$, in terms of the Legendre coefficients $b_{n,2}, \ldots, b_{n+2k,2}$.

Theorem 3.3. Let $k \geq 1$ be an integer and $\xi : [0, \pi] \to \mathbb{R}$ continuous with $\xi(0) = 1$. Further, assume that $\xi(\theta)$ is given by the $(2k+2)$-Gegenbauer expansion (3.1) for every $\theta \in [0, \pi]$. Then for all integers $n \geq 0$ we have

$$b_{n,2k+2} = \sum_{i=0}^{k} u_i(n,k)\, b_{n+2i,2}, \tag{3.7}$$

where

$$u_i(n,k) = (-1)^i \frac{(2k-1)!!}{2^k} \binom{k}{i} \binom{2k+n}{n} \frac{1}{(n+i+1/2)_{k-i}(n+k+3/2)_i}. \tag{3.8}$$

3.2 An application of the connection results

As shown in the following example, our results can be used to decide whether a function $\psi \in \Psi_1$ is a member of Ψ_∞^+.

Example 3.4. Let

$$b_{n,1} = \frac{3}{\pi^2 n^2}, \qquad n \geq 1, \tag{3.9}$$

and $b_{0,1} = 1/2$, implying that the corresponding function ψ is in Ψ_1. Let $B(x,y)$ denote the Euler Beta function and ${}_4F_3$ be a generalized hypergeometric function [101]. As shown in Section 3.3, an application of Theorem 3.1 yields

$$b_{n,2k+1} = \sum_{i=0}^{k} a_i(n,k) b_{n+2i,1} = \frac{3k(n+k)\,[B(n/2,k)]^2}{2n\pi^2(n+2k)^2 B(n,2k)}, \tag{3.10}$$

for $n, k \geq 1$, and

$$b_{0,2k+1} = \frac{1}{2} - \frac{3k\ {}_4F_3(1,1,1,1-k;\ 2,2,2+k;\ 1)}{4(1+k)\pi^2}, \tag{3.11}$$

for $k \geq 1$. Evidently, $b_{n,2k+1} > 0$ for all $n, k \geq 1$. By the definition of the generalized

hypergeometric function [101, equation (2.1.1.1)],

$$\begin{aligned} {}_4F_3(1,1,1,1-k;\,2,2,2+k;\,1) &= \sum_{i=0}^{\infty} \frac{(1)_i(1)_i(1)_i(1-k)_i}{(2)_i(2)_i(2+k)_i}\frac{1}{i!} \\ &= \sum_{i=0}^{k-1} \frac{1}{(1+i)^2}\frac{(1-k)_i}{(2+k)_i}, \end{aligned} \tag{3.12}$$

because $(1-k)_i/(2+k)_i = 0$ for $i \geq k$. As $(1-k)_i/(2+k)_i \leq 1$ for all integers $i, k \geq 0$, we have

$$\sum_{i=0}^{k-1} \frac{1}{(1+i)^2}\frac{(1-k)_i}{(2+k)_i} \leq \sum_{i=0}^{k-1} \frac{1}{(1+i)^2} \leq \sum_{i=0}^{\infty} \frac{1}{(1+i)^2} = \frac{\pi^2}{6}.$$

Putting this together yields

$$\begin{aligned} b_{0,2k+1} &= \frac{2(k+1)\pi^2 - 3k\ {}_4F_3(1,1,1,1-k;\,2,2,2+k;\,1)}{4(1+k)\pi^2} \\ &\geq \frac{2(k+1)\pi^2 - k\pi^2/2}{4(1+k)\pi^2} = \frac{3k+4}{8(1+k)} > 0, \end{aligned}$$

for all $k \geq 1$. This in combination with $b_{0,1} > 0$ and the uniqueness of the Gegenbauer expansions finally gives $\psi \in \Psi_\infty^+$.

It is interesting to note that $b_{n,2k+1} = \mathcal{O}(n^{-2})$ for all $k \geq 1$. Thus, the Schoenberg coefficients show the same asymptotic behaviour in every odd dimension, which can be seen as follows. Stirling's formula for the Gamma function [2, eq. (6.1.37)] yields, for fixed y,

$$B(x,y) = \Gamma(y)e^y \left(1+\frac{y}{x}\right)^{1/2-y-x} x^{-y} \frac{1+\mathcal{O}(x^{-1})}{1+\mathcal{O}((x+y)^{-1})}.$$

Because $\left(1+\frac{y}{x}\right)^{1/2-y-x} \to e^{-y} > 0$ as $x \to \infty$, it follows that

$$B(x,y) = \Gamma(y)x^{-y}\mathcal{O}(1)\frac{1+\mathcal{O}(x^{-1})}{1+\mathcal{O}((x+y)^{-1})}.$$

This yields immediately

$$\frac{B(n/2,k)^2}{B(n,2k)} = \mathcal{O}(1)$$

and, hence, $b_{n,2k+1} = \mathcal{O}(n^{-2})$.

3.3 Proofs

Proof of Theorem 3.1: First, consider the case $n \geq 1$. We proceed by two-dimensional induction over $k, n \geq 1$.

Let $k = 1$. For all $n \geq 1$ we have

$$b_{n,3} = \frac{1}{2}(n+1)(b_{n,1} - b_{n+2,1}),$$

yielding $a_0(n,1) = \frac{1}{2}(n+1)$ and $a_1(n,1) = -\frac{1}{2}(n+1)$. Inserting $i = 0$ and $k = 1$ into formula (3.6) yields the same, proving the claim for $k = 1$ and every $n \geq 1$.

Assume that Equation (3.6) is true for an arbitrary $k \geq 1$ and all $n \geq 1$. It suffices to show that this implies (3.6) for $k+1$. Using (3.5) and comparing the coefficients shows

$$\begin{aligned}
b_{n,2(k+1)+1} &= \frac{(n+2k)_2}{2(2k+1)(n+k)} b_{n,2k+1} - \frac{(n+1)_2}{2(2k+1)(n+k+2)} b_{n+2,2k+1} \\
&= \frac{(n+2k)_2}{2(2k+1)(n+k)} \sum_{i=0}^{k} a_i(n,k) b_{n+2i,1} - \frac{(n+1)_2}{2(2k+1)(n+k+2)} \sum_{i=0}^{k} a_i(n+2,k) b_{n+2+2i,1} \\
&= \frac{(n+2k)_2}{2(2k+1)(n+k)} a_0(n,k) b_{n,1} - \frac{(n+1)_2}{2(2k+1)(n+k+2)} a_k(n+2,k) b_{n+2(k+1),1} \\
&\quad + \frac{1}{2(2k+1)} \sum_{i=1}^{k} \left[\frac{(n+2k)_2}{n+k} a_i(n,k) - \frac{(n+1)_2}{n+k+2} a_{i-1}(n+2,k) \right] b_{n+2i,1}.
\end{aligned}$$

The induction hypothesis together with the identity

$$(x)_k (x+k)_l = (x)_{k+l}, \tag{3.13}$$

for integers $k, l \geq 0$, implies

$$\begin{aligned}
\frac{(n+2k)_2}{2(2k+1)(n+k)} a_0(n,k) &= \frac{(n+2k)_2}{2(2k+1)(n+k)} \frac{1}{2^k} \frac{(n+k) n (n+1)_{2k-1}}{(2k-1)!! (n)_{k+1}} \\
&= \frac{1}{2^{k+1}} \frac{n (n+k+1) (n+1)_{2k+1}}{(2k+1)!! (n)_{k+2}},
\end{aligned}$$

and

$$
\begin{aligned}
&-\frac{(n+1)_2}{2(2k+1)(n+k+2)}a_k(n+2,k)\\
&\quad=-\frac{(n+1)_2}{2(2k+1)(n+k+2)}\frac{(-1)^k}{2^k}\frac{(n+2+k)(n+2+2k)(n+2+1)_{2k-1}}{(2k-1)!!(n+2+k)_{k+1}}\\
&\qquad=\frac{(-1)^{k+1}}{2^{k+1}}\frac{(n+2+2k)(n+1)_{2k+1}}{(2k+1)!!(n+2+k)_{k+1}}\\
&\qquad=\frac{(-1)^{k+1}}{2^{k+1}}\frac{(n+k+1)(n+2k+2)(n+1)_{2k+1}}{(2k+1)!!(n+1+k)_{k+2}}.
\end{aligned}
$$

These two calculations show the validity of (3.6) for $i=0$ and $i=k+1$, respectively.

It remains to show for $1\le i\le k$ that

$$
\begin{aligned}
&\frac{(-1)^i}{2^{k+1}}\binom{k+1}{i}\frac{(n+k+1)(n+2i)(n+1)_{2k+1}}{(2k+1)!!(n+i)_{k+2}}\\
&=\frac{1}{2(2k+1)}\left[\frac{(n+2k)_2}{n+k}a_i(n,k)-\frac{(n+1)_2}{n+k+2}a_{i-1}(n+2,k)\right],
\end{aligned}
\tag{3.14}
$$

where $a_i(n,k)$ and $a_{i-1}(n+2,k)$ can be expressed as in (3.6). After plugging the induction hypothesis into the right-hand side of (3.14) and reformulating it, (3.14) becomes

$$
\begin{aligned}
&\frac{(-1)^i}{2^{k+1}}\binom{k+1}{i}\frac{(n+k+1)(n+2i)(n+1)_{2k+1}}{(2k+1)!!(n+i)_{k+2}}\\
&\qquad=\frac{(-1)^i}{2^{k+1}}\frac{(n+2k)_2}{n+k}\binom{k}{i}\frac{(n+k)(n+2i)(n+1)_{2k-1}}{(2k+1)!!(n+i)_{k+1}}\\
&\qquad\quad-\frac{(-1)^{i-1}}{2^{k+1}}\frac{(n+1)_2}{n+k+2}\binom{k}{i-1}\frac{(n+2+k)(n+2i)(n+3)_{2k-1}}{(2k+1)!!(n+1+i)_{k+1}}.
\end{aligned}
$$

Simplifying the equation above with the help of (3.13) shows that we have to deduce

$$
\binom{k+1}{i}(n+k+1)=\binom{k}{i}(n+i+k+1)+\binom{k}{i-1}(n+i).
\tag{3.15}
$$

But the right-hand side of (3.15) equals

$$\begin{aligned}\frac{k!}{(k-i)!(i-1)!}&\left(\frac{n+i+k+1}{i}+\frac{n+i}{k-i+1}\right)\\ &= \frac{k!}{(k+1-i)!i!}\big[(k-i+1)(n+i+k+1)+i(n+i)\big]\\ &= \frac{k!}{(k+1-i)!i!}(k+1)(n+k+1) = \binom{k+1}{i}(n+k+1),\end{aligned}$$

which shows that (3.15) holds true.

Now we turn to the case $n = 0$. Equations (3.4) and (3.5) show that $a_i(0,k)$ is given by (3.6) for all $i > 0$. To find $a_0(0,k)$, note that equation (3.5) together with (3.4) yields for $n = 0$ and $d \geq 1$

$$b_{0,d+2} = b_{0,d} - \frac{2}{d(d+3)}b_{2,d}.$$

Using both equations recursively for $b_{2,d}$ and $b_{0,d}$, we see that

$$b_{0,d+2} = b_{0,1} - R,$$

where the remainder term R does not depend on $b_{0,1}$. This shows that $a_0(0,k) = 1$ for all $k \geq 1$. □

Proof of Proposition 3.2: Let $n \geq 1$. The identity $(n+i)_{k+1} = \binom{n+i+k}{k+1}(k+1)!$ together with Theorem 3.1 implies

$$\begin{aligned}\sum_{i=0}^{k} a_i(n,k) &= \frac{(n+k)(n+1)_{2k-1}}{2^k(2k-1)!!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}\frac{n+2i}{(n+i)_{k+1}}\\ &= \frac{(n+k)(n+1)_{2k-1}}{2^k(k+1)!(2k-1)!!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}\frac{n+2i}{\binom{n+i+k}{k+1}}.\end{aligned}$$

Hence, it suffices to prove that

$$\sum_{i=0}^{k}(-1)^i\binom{k}{i}\frac{n+2i}{\binom{n+i+k}{k+1}} = 0.$$

This statement is equivalent to

$$\sum_{i=0}^{k}(-1)^i\binom{k}{i}\frac{n}{\binom{n+i+k}{k+1}} = -2\sum_{i=0}^{k}(-1)^i\binom{k}{i}\frac{i}{\binom{n+i+k}{k+1}}, \tag{3.16}$$

where the left-hand side equals

$$\sum_{i=0}^{k}(-1)^i\binom{k}{i}\frac{n}{\binom{n+i+k}{k+1}} = n\frac{k+1}{2k+1}\frac{1}{\binom{2k+n}{n-1}} = \frac{k+1}{\binom{2k+n}{n}}.$$

Here, the first equality is a consequence of the following result of R. Frisch (cf. Note 21 in [78]), in that

$$\sum_{i=0}^{k}(-1)^i\binom{k}{i}\frac{1}{\binom{b+i}{c}} = \frac{c}{k+c}\frac{1}{\binom{k+b}{b-c}}, \tag{3.17}$$

where $b \geq c$ are positive integers.

For the right-hand side of (3.16) we deduce in a similar way

$$\begin{aligned}
-2\sum_{i=0}^{k}(-1)^i\binom{k}{i}\frac{i}{\binom{n+i+k}{k+1}} &= -2\sum_{i=0}^{k}(-1)^i\frac{k!}{(k-i)!(i-1)!}\frac{1}{\binom{n+i+k}{k+1}} \\
&= -2k\sum_{i=1}^{k}(-1)^i\binom{k-1}{i-1}\frac{1}{\binom{n+i+k}{k+1}} \\
&= 2k\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}\frac{1}{\binom{n+i+1+k}{k+1}} \\
&= 2k\frac{k+1}{2k}\frac{1}{\binom{2k+n}{n}} = \frac{k+1}{\binom{2k+n}{n}}.
\end{aligned}$$

Again, we applied (3.17) for the third equality, thereby proving (3.16).

Now consider the case $n = 0$. For $i > 0$ equation (3.6) simplifies to

$$a_i(0,k) = (-1)^i\binom{k}{i}\frac{1}{\binom{k+i}{k}},$$

which is also valid for $i = 0$, because in this case the right-hand side equals 1.

Applying (3.17) with $b = c = k$ we obtain

$$\sum_{i=0}^{k} a_i(0,k) = \sum_{i=0}^{k} (-1)^i \binom{k}{i} \frac{1}{\binom{k+i}{k}} = \frac{1}{2}.$$

□

Proof of Theorem 3.3. Similarly to the proof of Theorem 3.1, we proceed by two-dimensional induction over $k \geq 1$ and $n \geq 0$. Let $k = 1$. In this case, equation (3.5) yields for all $n \geq 0$

$$b_{n,4} = \frac{1}{2}(n+1)_2 \left[\frac{1}{2n+1} b_{n,2} - \frac{1}{2n+5} b_{n+2,2} \right],$$

implying $u_0(n,1) = \frac{1}{2(2n+1)}(n+1)_2$ and $u_1(n,1) = -\frac{1}{2(2n+5)}(n+1)_2$. Inserting $k = 1$ and $i = 0$ – respectively $i = 1$ – in equation (3.8) proves the claim for $k = 1$ and all $n \geq 0$.

Suppose we have proven (3.8) for an arbitrary but fixed $k \geq 1$ and all $n \geq 0$. Again, it suffices to show the validity of (3.8) for $k+1$. With equations (3.4) and (3.5) and the induction hypothesis we find that

$$\begin{aligned}
b_{n,2(k+1)+2} &= \frac{(n+2k+1)_2}{2(k+1)(2n+2k+1)} b_{n,2k+2} - \frac{(n+1)_2}{2(k+1)(2n+2k+5)} b_{n+2,2k+2} \\
&= \frac{(n+2k+1)_2}{2(k+1)(2n+2k+1)} \sum_{i=0}^{k} u_i(n,k) b_{n+2i,2} - \frac{(n+1)_2}{2(k+1)(2n+2k+5)} \sum_{i=0}^{k} u_i(n+2,k) b_{n+2+2i,2} \\
&= \frac{(n+2k+1)_2}{2(k+1)(2n+2k+1)} u_0(n,k) b_{n,2} - \frac{(n+1)_2}{2(k+1)(2n+2k+5)} u_k(n+2,k) b_{n+2(k+1),2} \\
&+ \sum_{i=1}^{k} \left[\frac{(n+2k+1)_2}{2(k+1)(2n+2k+1)} u_i(n,k) - \frac{(n+1)_2}{2(k+1)(2n+2k+5)} u_{i-1}(n+2,k) \right] b_{n+2i,2}.
\end{aligned}$$

Equation (3.13) together with the identity

$$(2k-1)!! = \frac{(2k)!}{2^k k!}$$

yields

$$\frac{(n+2k+1)_2}{2(k+1)(2n+2k+1)} u_0(n,k) = \frac{(n+2k+1)_2 (2k-1)!!}{2^{k+1}(k+1)(2n+2k+1)} \binom{2k+n}{n} \frac{1}{(n+1/2)_k}$$

where the right-hand side is equal to

$$\frac{(2k+1)!!}{2^{k+1}}\binom{2k+2+n}{n}\frac{1}{(n+1/2)_{k+1}},$$

proving the claim for $i=0$. Analogously, we confirm the claim for $i=k+1$ by computing

$$\begin{aligned}
&-\frac{(n+1)_2}{2(k+1)(2n+2k+5)}u_k(n+2,k)\\
&\quad=(-1)^{k+1}\frac{(n+1)_2(2k-1)!!}{2^{k+1}(k+1)(2n+2k+5)}\binom{2k+n+2}{n+2}\frac{1}{(n+k+7/2)_k}\\
&\quad=(-1)^{k+1}\frac{1}{2^{k+1}2^{k+1}}\frac{(2k+2+n)!}{(k+1)!n!}\frac{1}{(n+k+5/2)_{k+1}}\\
&\quad=(-1)^{k+1}\frac{(2k+1)!!}{2^{k+1}}\binom{2k+2+n}{n}\frac{1}{(n+k+5/2)_{k+1}}.
\end{aligned}$$

Now let $1\le i\le k$. We need to show that

$$\begin{aligned}
u_i(n,k+1)&=(-1)^i\frac{(2k+1)!!}{2^{k+1}}\binom{k+1}{i}\binom{2k+2+n}{n}\frac{1}{(n+i+1/2)_{k+1-i}(n+k+1+3/2)_i}\\
&=\frac{(n+2k+1)_2}{2(k+1)(2n+2k+1)}u_i(n,k)-\frac{(n+1)_2}{2(k+1)(2n+2k+5)}u_{i-1}(n+2,k),
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&(-1)^i\frac{(2k+1)!!}{2^{k+1}}\binom{k+1}{i}\binom{2k+2+n}{n}\frac{1}{(n+i+1/2)_{k+1-i}(n+k+1+3/2)_i}\\
&\quad=\frac{(-1)^i(n+2k+1)_2(2k-1)!!}{2^{k+1}(k+1)(2n+2k+1)}\binom{k}{i}\binom{2k+n}{n}\frac{1}{(n+i+1/2)_{k-i}(n+k+3/2)_i}\\
&\qquad-\frac{(-1)^{i-1}(n+1)_2(2k-1)!!}{2^{k+1}(k+1)(2n+2k+5)}\binom{k}{i-1}\\
&\qquad\quad\times\binom{2k+n+2}{n+2}\frac{1}{(n+i+1+1/2)_{k-i+1}(n+2+k+3/2)_{i-1}}.
\end{aligned}\tag{3.18}$$

With (3.13) and similar arguments as in the cases $i=0$ and $i=k+1$ we see that

(3.18) is equivalent to

$$\begin{aligned}\binom{k+1}{i}\frac{1}{(n+i+1/2)_{k+1-i}(n+k+5/2)_i} \\ = \binom{k}{i}\frac{1}{(n+i+1/2)_{k+1-i}(n+k+3/2)_i} \\ + \binom{k}{i-1}\frac{1}{(n+i+3/2)_{k+1-i}(n+k+5/2)_i}.\end{aligned}$$

Multiplying by $(n+i+1/2)_{k+1-i}(n+k+5/2)_i$ illustrates that we need to show

$$\binom{k+1}{i} = \binom{k}{i}\frac{n+k+i+3/2}{n+k+3/2} + \binom{k}{i-1}\frac{n+i+1/2}{n+k+3/2}.$$

Eventually, a simplification of the right-hand side yields

$$\begin{aligned}\binom{k}{i}\frac{(n+k+i+3/2)(k-i+1)+i(n+i+1/2)}{(n+k+3/2)(k-i+1)} \\ = \binom{k}{i}\frac{(k+1)(n+k+3/2)}{(n+k+3/2)(k-i+1)} = \binom{k+1}{i},\end{aligned}$$

and the proof is complete. □

Proof for Example 3.4. First, we prove equation (3.10). As in the preceding proofs, we use two-dimensional induction over $k, n \geq 1$. Let $k = 1$ and $n \geq 1$. Equation (3.4) leads to

$$\begin{aligned}b_{n,3} &= \frac{1}{2}(n+1)\left(b_{n,1} - b_{n+2,1}\right) \\ &= \frac{3(n+1)}{2\pi^2}\left(\frac{1}{n^2} - \frac{1}{(n+2)^2}\right) \\ &= \frac{6(n+1)^2}{\pi^2 n^2 (n+2)^2}.\end{aligned}$$

On the other hand, inserting $k = 1$ in the right-hand side of (3.10) and using the functional equation for the Gamma function yields

$$\frac{3(n+1)\Gamma(n/2)^2\Gamma(n+2)}{2n\pi^2(n+2)^2\Gamma(n)\Gamma(n/2+1)^2} = \frac{3(n+1)^2 n}{2n\pi^2(n+2)^2(n/2)^2} = \frac{6(n+1)^2}{\pi^2 n^2 (n+2)^2}.$$

Now assume we have proven (3.10) for a certain $k \geq 1$ and all $n \geq 1$. It suffices

to show that this implies (3.10) for $k+1$. Equation (3.5) in combination with the induction hypothesis proves the claim, as

$$\begin{aligned}
b_{n,2k+3} &= \frac{(n+2k)(n+2k+1)}{(2k+1)(2n+2k)}\frac{3k(n+k)B(n/2,k)^2}{2n\pi^2(n+2k)^2B(n,2k)} \\
&\quad - \frac{(n+1)(n+2)}{(2k+1)(2n+2k+4)}\frac{3k(n+2+k)B(n/2+1,k)^2}{2(n+2)\pi^2(n+2+2k)^2B(n+2,2k)} \\
&= \frac{3k\Gamma(k)^2}{4\pi^2(2k+1)\Gamma(2k)}\left[\frac{(n+2k+1)\Gamma(n/2)^2\Gamma(n+2k)}{n(n+2k)\Gamma(n)\Gamma(n/2+k)^2}\right. \\
&\quad \left. - \frac{(n+1)\Gamma(n/2+1)^2\Gamma(n+2+2k)}{(n+2+2k)^2\Gamma(n+2)\Gamma(n/2+1+k)^2}\right] \\
&= \frac{3k\Gamma(k)^2\Gamma(n/2+1)^2\Gamma(n+2k+2)}{4\pi^2(2k+1)\Gamma(2k)\Gamma(n+1)\Gamma(n/2+1+k)^2}\left[\frac{(n+2k+1)(n/2+k)^2}{(n+2k)(n/2)^2(n+2k)(n+2k+1)}\right. \\
&\quad \left. - \frac{1}{(n+2+2k)^2}\right] \\
&= \frac{3k\Gamma(k)^2\Gamma(n/2+1)^2\Gamma(n+2k+2)}{4\pi^2(2k+1)\Gamma(2k)\Gamma(n+1)\Gamma(n/2+1+k)^2}\left[\frac{1}{n^2} - \frac{1}{(n+2+2k)^2}\right] \\
&= \frac{24k^2\Gamma(k)^2\Gamma(n/2)^2(n/2)^2\Gamma(n+2k+2)(k+1)(k+1+n)}{4\pi^2(2k+1)(2k)\Gamma(2k)\Gamma(n)n\Gamma(n/2+1+k)^2n^2(n+2+2k)^2},
\end{aligned}$$

which equals

$$\frac{3(k+1)(k+1+n)B(n/2,k+1)^2}{2\pi^2n(n+2+2k)^2B(n,2k+2)}.$$

For the proof of equation (3.11), assume $n=0$. Theorem 3.1 implies

$$\begin{aligned}
b_{n,2k+1} = \sum_{i=0}^{k} a_i(0,k)b_{2i,1} &= \frac{1}{2} + \sum_{i=1}^{k}\frac{(-1)^i}{2^k}\binom{k}{i}\frac{2ki(2k-1)!}{(2k-1)!!(i)_{k+1}}\frac{3}{4\pi^2i^2} \\
&= \frac{1}{2} + \frac{6k(2k-1)!(k!)^2 2^k}{4\pi^2 2^k(2k)!}\sum_{i=1}^{k}(-1)^i\frac{1}{i!(k-i)!}\frac{1}{i(i)_{k+1}} \\
&= \frac{1}{2} - \frac{3(k!)^2}{4\pi^2}\sum_{i=0}^{k-1}(-1)^i\frac{1}{(i+1)^2(k-i-1)!(i+k+1)!}. \qquad (3.19)
\end{aligned}$$

But this equals

$$\frac{1}{2}-\frac{3k}{4\pi^2(k+1)}\sum_{i=0}^{k-1}\frac{1}{(i+1)^2}\frac{(1-k)_i}{(2+k)_i}=\frac{1}{2}-\frac{3k}{4\pi^2(k+1)}\,{}_4F_3(1,1,1,1-k;2,2,2+k;\,1),$$

which is a consequence of the identities

$$\frac{(k+1)!}{(k-i-1)!}=\frac{1}{(2+k)_i}\qquad\text{and}\qquad(-1)^i\frac{(k-1)!}{(k-i-1)!}=(1-k)_i$$

together with the definition of the generalized hypergeometric function (3.12). This shows our claim. □

4 Curl- and divergence-free stochastic vector processes on the sphere

Many phenomena in nature can be described by *vector fields*, where a vector-valued physical quantity is assigned to each point in space-time. Prominent examples are vector fields describing the behaviour of fluids or wind; other examples are electric or magnetic fields, cf. [49, 32]. For some of these fields we have knowledge about certain physical characteristics, such as the *curl* or the *divergence*. E.g., for vector fields describing fluids, these characteristics can be interpreted in the following way [32]. The curl is a measure of the vorticity of the field describing its eddies. The divergence is a local mass balance describing how much flows into or out of a given point. If the vector field is divergence-free, it possesses no sources or sinks – like static magnetic fields – and if it is curl-free, it is irrotational – like static electric fields.

In what follows we focus on the case of purely spatial processes, since this means no severe restriction. We will discuss this point at the end of this chapter.

Narcowich and Ward [75] and Fuselier [34] analyzed interpolation of divergence-free vector fields based on radial basis functions. They gave conditions on the radial basis functions, ensuring that the interpolated vector fields are still divergence-free.

Recently, Scheuerer and Schlather [94] generalized their approach to *random vector fields* on subsets $U \subset \mathbb{R}^d$ with dimension $d = 2$ or $d = 3$. Technically, they considered d-variate Gaussian random vector fields on U of the form

$$Z(x) = Z_1(x)\, e_1 + \cdots + Z_d(x)\, e_d, \qquad x \in U,$$

where e_i is the i-th unit vector in $\mathbb{R}^d$. Specifically, they characterized random vector fields with almost surely (a.s.) curl- and divergence-free sample paths in terms of the matrix-valued cross-covariance function $C : U \times U \to \mathbb{R}^{d\times d}$,

$$C_{ij}(x, y) = \mathrm{Cov}(Z_i(x), Z_j(y)), \qquad x, y \in U, \quad i, j = 1, \ldots, d.$$

Based on this characterization, they gave explicit principles for constructing cross-

covariance functions corresponding to random vector fields with a.s. curl- and divergence-free sample paths based on *univariate* covariance functions.

The natural domain of vector fields is not always a plane. In particular, most vector fields arising in meteorology are defined on the Earth's surface, which is approximately a sphere. Vector fields on spheres are also considered in astrophysics, see for example Mandelbaum, et al. [69], who examine cosmological shear fields. These shear fields consist almost exclusively of so-called E-mode signals, which are curl-free. Since curl- or divergence-free vector fields in $\mathbb{R}^3$ might lose this property when restricted to a sphere, it is natural to ask how to interpolate and model on curved surfaces. Narcowich, Ward, and Wright [77] considered this problem for divergence-free fields on surfaces in $\mathbb{R}^3$ by studying divergence-free radial basis functions on them.

The goal of the present chapter is to develop cross-covariance models for random vector fields with a.s. divergence- and curl-free sample paths. We are interested primarily in cross-covariance models in spherical coordinates, as we have applications in mind where the sphere is considered to be the Earth. In such applications, coordinates are mostly given in longitude and latitude. We show that characterizations similar to those of Scheuerer and Schlather [94] can be applied to random vector fields with domain $\mathbb{S}^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$, and that the methods of constructing random vector fields with a.s. curl- and divergence-free sample paths can be transferred from Euclidean spaces to the sphere. Further, we give principles for constructing the cross-covariance functions of such fields in terms of univariate covariance functions on the sphere.

4.1 Random vector fields on the sphere

We are interested in random vector fields (RVFs) on the sphere $\mathbb{S}^2$ with a spherical coordinate system with fixed radius one. Thus, every $x \in \mathbb{S}^2$ can be parameterized as

$$x = (\cos\varphi_x \cos\theta_x,\, \sin\varphi_x \cos\theta_x,\, \sin\theta_x)^t, \tag{4.1}$$

with $(\varphi_x, \theta_x) \in [0, 2\pi) \times [-\pi/2, \pi/2]$. This representation is unique for every $x \in \widetilde{\mathbb{S}}^2 = \mathbb{S}^2 \setminus \{N, S\}$, where $N = (0, 0, 1)^t$ is the North and $S = (0, 0, -1)^t$ is the South Pole. We make this representation unique on the whole sphere by setting $\varphi_x = 0$

for $x \in \{N, S\}$. Hence, for a random field (RF) $\{Z(x) : x \in \mathbb{S}^2\}$ we identify

$$Z(x) = Z(\varphi_x, \theta_x), \qquad (\varphi_x, \theta_x) \in [0, 2\pi) \times [-\pi/2, \pi/2].$$

for every $x \in \mathbb{S}^2$. For the sake of readability, we drop the subscript and write (φ, θ) if it is clear to which point we refer.

The tangent space $T_x\mathbb{S}^2$ attached to a point $x \in \mathbb{S}^2$ is isomorphic to $\mathbb{R}^2$: there are two unit vectors in $\mathbb{R}^3$ such that $T_x\mathbb{S}^2$ is the linear span of those two vectors. Obviously, for every x there are infinitely many possibilities for choosing these two *tangent vectors* but for the given coordinate system (4.1) there is a canonical choice: if (φ, θ) are the spherical coordinates of $x \in \mathbb{S}^2$ then $e_\varphi \in T_x\mathbb{S}^2$ denotes the unit tangent vector in the longitudinal direction and $e_\theta \in T_x\mathbb{S}^2$ denotes the unit tangent vector in the latitudinal direction. Given $x \in \widetilde{\mathbb{S}}^2$ with corresponding (φ, θ), these tangent vectors can be expressed in Cartesian coordinates as

$$e_\varphi = (-\sin\varphi, \cos\varphi, 0)^t \qquad \text{and} \qquad e_\theta = (-\cos\varphi\sin\theta, -\sin\varphi\sin\theta, \cos\theta)^t,$$

which are orthogonal (see, for example, Chapter 3, Example 14 in Agricola and Friedrich [4]). This is unique only for $x \in \widetilde{\mathbb{S}}^2$, because in the poles there is no canonical longitudinal or latitudinal direction. For the poles we set $e_\varphi = (0, 1, 0)^t$ and $e_\theta = (1, 0, 0)^t$ for $x = N$ and $e_\varphi = (0, -1, 0)^t$ and $e_\theta = (-1, 0, 0)^t$ for $x = S$, which yields a right-handed orthonormal frame for $T_N\mathbb{S}^2$ and $T_S\mathbb{S}^2$.

Definition 4.1. A *random vector field (RVF)* on $\mathbb{S}^2$ is a bivariate stochastic process Z indexed by $x \in \mathbb{S}^2$ with values in the tangent space $T_x\mathbb{S}^2$. Using the above parameterization we may write

$$\mathsf{Z}(x) \equiv \mathsf{Z}(\varphi, \theta) = \mathsf{Z}_1(\varphi, \theta)\, e_\varphi + \mathsf{Z}_2(\varphi, \theta)\, e_\theta,$$

with real-valued RFs $\mathsf{Z}_1, \mathsf{Z}_2$.

The concept of mean-square differentiability for RFs from Scheuerer and Schlather [94] allows us to define the notion of a derivative for a RF on $\mathbb{S}^2$.

Definition 4.2. Let $\{Z(x) : x \in \mathbb{S}^2\}$ be a univariate RF. The RF is mean-square partial differentiable in the direction e_φ in the point $x \in \mathbb{S}^2$ (corresponding to φ_x, θ_x) if there is a random variable $Z^{(\varphi)}(\varphi_x, \theta_x)$ with finite second moments such

that

$$\mathbb{E}\left(\frac{Z(\varphi_x+h,\theta_x)-Z(\varphi_x,\theta_x)}{h}-Z^{(\varphi)}(\varphi_x,\theta_x)\right)^2\longrightarrow 0\qquad\text{for}\qquad h\longrightarrow 0.$$

The partial derivative $Z^{(\theta)}$ is defined analogously. Furthermore, $Z^{(\theta,\varphi)}$, resp. $Z^{(\varphi,\theta)}$, denotes the partial derivative of $Z^{(\varphi)}$ in the direction θ, resp. the partial derivative of $Z^{(\theta)}$ in the direction φ.

As the sphere – treated as a manifold – needs two charts to be covered, we consider RVFs on $\widetilde{\mathbb{S}}^2$, as this can be bijectively mapped to $[0,2\pi)\times(-\pi/2,\pi/2)\subset\mathbb{R}^2$ via (4.1). This choice of coordinates implies that the usual formulas for divergence or curl also take different forms.

Definition 4.3. Let $\{\mathsf{Z}(\varphi,\theta):(\varphi,\theta)\in(0,2\pi]\times(-\pi/2,\pi/2)\}$ be a RVF on $\widetilde{\mathbb{S}}^2$ with a.s. differentiable sample paths. We define the *divergence* and the *curl* of Z as

$$\operatorname{div}(\mathsf{Z}(\varphi,\theta)):=\frac{1}{\cos(\theta)}\,\mathsf{Z}_1^{(\varphi)}(\varphi,\theta)+\,\mathsf{Z}_2^{(\theta)}(\varphi,\theta)-\tan(\theta)\,\mathsf{Z}_2(\varphi,\theta),\tag{4.2}$$

$$\operatorname{curl}(\mathsf{Z}(\varphi,\theta)):=\left(\mathsf{Z}_1^{(\theta)}(\varphi,\theta)-\tan(\theta)\,\mathsf{Z}_1(\varphi,\theta)-\frac{1}{\cos(\theta)}\,\mathsf{Z}_2^{(\varphi)}(\varphi,\theta)\right)e_r.\tag{4.3}$$

Here, e_r is the outward-pointing unit normal vector. In practice, we are only interested in the coefficient of the vector $\operatorname{curl}(\mathsf{Z}(\varphi,\theta))$. Hence, we will omit e_r in the following and identify $\operatorname{curl}(\mathsf{Z}(\varphi,\theta))$ with its real-valued vector component in direction e_r.

Although the curl operator can be defined for vector fields on $\mathbb{R}^2$, we are interested in applying it only to vector fields in $\mathbb{R}^3$ as it loses its physical interpretation in the two-dimensional case. Hence, we consider the RVF Z on $\mathbb{S}^2$ as a RVF $\widetilde{\mathsf{Z}}$ on $\mathbb{R}^3$ by writing

$$\mathsf{Z}(\varphi,\theta)=\widetilde{\mathsf{Z}}(\varphi,\theta,r=1)=\widetilde{\mathsf{Z}}_1(\varphi,\theta,r=1)e_\varphi+\,\widetilde{\mathsf{Z}}_2(\varphi,\theta,r=1)\,e_\theta+0\,e_{r=1},$$

where we set $r=1$ to one to eliminate the dependency of r; specifically, this yields $e_{r=1}=(\cos\varphi\cos\theta,\,\sin\varphi\cos\theta,\,\sin\theta)^t$.

Remark 4.4. Assume that all sample paths of Z are a.s. continuously differentiable.

It follows from the proof of Theorem 1 in Scheuerer and Schlather [94] that if $\mathbb{E}(\mathrm{div}(\mathsf{Z})^2) = 0$, resp. $\mathbb{E}(\mathrm{curl}(\mathsf{Z})^2) = 0$, then almost every sample path of Z is divergence-free, resp. curl-free, which can be seen as follows. Let $\mathbb{E}(\mathrm{div}(\mathsf{Z})^2) = 0$ and denote by Ω the sample space of Z. By Fubini's Theorem,

$$\mathbb{E}\left(\int_{\mathbb{S}^2} \left[\mathrm{div}(\mathsf{Z}(x))\right]^2 \,\mathrm{d}x\right) = \int_{\mathbb{S}^2} \mathbb{E}\left[\mathrm{div}(\mathsf{Z}(x))^2\right] \,\mathrm{d}x = 0,$$

which implies the existence of a set $N \subset \Omega$ with measure zero, such that $I(\omega) := \int_{\mathbb{S}^2} \left[\mathrm{div}(\mathsf{Z}(x))\right]^2 \,\mathrm{d}x = 0$ for all $\omega \in \Omega \backslash N$. In other words, for those ω the divergence of $\mathsf{Z}(x)(\omega), x \in \mathbb{S}^2$, can be non-zero only on a set with measure zero. Then, by the continuity of the partial derivatives of Z, the divergence vanishes everywhere on $\mathbb{S}^2$ for all $\omega \in \Omega \backslash N$. In the case where $\mathbb{E}(\mathrm{curl}(\mathsf{Z})^2) = 0$, the proof is analogous.

Consequently, we can transfer the property of being curl- or divergence-free in the mean square sense to the corresponding property of the sample paths.

4.2 Main results

In this section we present the main results for covariance functions corresponding to curl- and divergence-free RVFs on $\widetilde{\mathbb{S}}^2$. In the following, $\mathcal{C}^{(n,m)}(\mathbb{S}^2, \mathbb{S}^2)$ denotes the space of functions $C : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$ that are n-times continuously differentiable in the first argument and m-times continuously differentiable in the second argument. The Nabla operator in spherical coordinates equals

$$\nabla = \left(\frac{1}{\cos\theta}\,\partial_\varphi, \partial_\theta\right)^t.$$

Given a function on $\mathbb{S}^2 \times \mathbb{S}^2$, the operators ∇_1 and ∇_2 denote the Nabla operators operating on the first and second argument of the function, respectively.

The following result states a construction principle of cross-covariance functions corresponding to curl- and divergence-free RVFs. The result is similar to Theorem 2 in Scheuerer and Schlather [94]. There and in the following, we use matrix notation for the composition of operators. For example, $\nabla_2 \nabla_1^t$ denotes the operator

$$\begin{pmatrix} \frac{1}{\cos\theta_2 \cos\theta_1}\partial_{\varphi_2}\partial_{\varphi_1} & \frac{1}{\cos\theta_2}\partial_{\varphi_2}\partial_{\theta_1} \\ \frac{1}{\cos\theta_1}\partial_{\theta_2}\partial_{\varphi_1} & \partial_{\theta_2}\partial_{\theta_1} \end{pmatrix},$$

acting on a function $C : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$, with $(\theta_i, \varphi_i), i = 1, 2$, parameterizing the i-th

argument of C. Analogously, with $\mathbf{1}_2$ being the 2×2 identity matrix, the operator $(\nabla_2^t \nabla_1)\mathbf{1}_2$ stands for

$$\begin{pmatrix} \frac{1}{\cos\theta_2 \cos\theta_1}\partial_{\varphi_2}\partial_{\varphi_1} + \partial_{\theta_2}\partial_{\theta_1} & 0 \\ 0 & \frac{1}{\cos\theta_2 \cos\theta_1}\partial_{\varphi_2}\partial_{\varphi_1} + \partial_{\theta_2}\partial_{\theta_1} \end{pmatrix}.$$

Theorem 4.5. Let $C \in \mathcal{C}^{(1,1)}(\widetilde{\mathbb{S}}^2, \widetilde{\mathbb{S}}^2)$ be a covariance function. Consider the matrix-valued functions $\mathsf{C}_{div}, \mathsf{C}_{curl} : \widetilde{\mathbb{S}}^2 \times \widetilde{\mathbb{S}}^2 \to \mathbb{R}^{2\times 2}$ defined by

$$\mathsf{C}_{div} := \left[-\nabla_2\nabla_1^t + (\nabla_2^t\nabla_1)\mathbf{1}_2\right] C \qquad \text{and} \qquad \mathsf{C}_{curl} := (\nabla_1\nabla_2^t)\, C. \tag{4.4}$$

(a) Then, there exist bivariate processes on $\widetilde{\mathbb{S}}^2$ with cross-covariance functions C_{div} and C_{curl}, which we denote by Z_{div} and Z_{curl}.

(b) If a sample path of the process Z_{div}, resp. Z_{curl}, is a.s. continuously differentiable, then it also is divergence-free, resp. curl-free.

Part (a) of the theorem together with (b) provide a construction method for bivariate processes with almost surely divergence- or curl-free sample paths. An analogous construction principle for processes on the sphere was given in [29, 77] in terms of Cartesian coordinates. The usage of Cartesian coordinates has the advantage that it is feasible on the whole sphere. However, the construction given in the above theorem has the advantage that it can be used directly for data given in spherical coordinates. Further, the so constructed models might admit a better interpretation of their properties, as the usage of Cartesian coordinates does not respect the geometry of the sphere, cf. Banerjee [7].

Remark 4.6. (1) The restriction to processes without poles is no limitation for applications as the probability for curls or divergences occurring at a single prespecified point is zero.

(2) We show in Section 4.3 that the constructions provided in the above theorem are unique. In other words, every curl- and divergence-free RVF on $\widetilde{\mathbb{S}}^2$ has matrix-valued covariance functions of the forms (4.4).

Proof of Theorem 4.5. (a) By Kolmogorov's Existence Theorem, there exists a univariate centred Gaussian process Z on $\widetilde{\mathbb{S}}^2$ with covariance function C. The mean-square partial derivatives $Z^{(\varphi)}(\varphi, \theta)$ and $Z^{(\theta)}(\varphi, \theta)$ exist at every point of $\widetilde{\mathbb{S}}^2$, by the

assumption on C. Define the bivariate process Z *via*

$$\mathsf{Z}_{curl}(\varphi,\theta) := \frac{1}{\cos(\theta)} Z^{(\varphi)}(\varphi,\theta) e_\varphi + Z^{(\theta)}(\varphi,\theta) e_\theta.$$

Straightforward calculations show that Z_{curl} has cross-covariance function C_{curl}.

Similarly, consider

$$\mathsf{Z}_{div}(\varphi,\theta) := Z^{(\theta)}(\varphi,\theta) e_\varphi - \frac{1}{\cos(\theta)} Z^{(\varphi)}(\varphi,\theta) e_\theta.$$

Again, a straightforward calculation yields that C_{div} is the cross-covariance function of Z_{div}.

(b) The statement is shown by direct calculation. First, consider $\mathsf{Z}_{\mathrm{curl}}$. Starting with (4.3), we deduce that

$$\begin{aligned}&\mathrm{curl}(\mathsf{Z}_{curl}(\varphi,\theta))\\ &= \frac{\sin(\theta)}{\cos^2(\theta)} Z^{(\varphi)}(\varphi,\theta) + \frac{1}{\cos(\theta)} Z^{(\theta,\varphi)}(\varphi,\theta) - \frac{1}{\cos(\theta)} Z^{(\varphi,\theta)}(\varphi,\theta) - \frac{\tan(\theta)}{\cos(\theta)} Z^{(\varphi)}(\varphi,\theta),\end{aligned}$$

which equals zero, as $Z^{(\theta,\varphi)} = Z^{(\varphi,\theta)}$ due to the continuity of the partial derivatives of C.

The calculation for Z_{div} is analogous. We use (4.2) and get

$$\begin{aligned}&\mathrm{div}(\mathsf{Z}_{div}(\varphi,\theta))\\ &\quad= \frac{1}{\cos\theta} Z^{(\varphi,\theta)}(\varphi,\theta) - \frac{\tan\theta}{\cos\theta} Z^{(\varphi)}(\varphi,\theta) - \frac{1}{\cos\theta} Z^{(\theta,\varphi)}(\varphi,\theta) + \frac{\tan\theta}{\cos\theta} Z^{(\varphi)}(\varphi,\theta),\end{aligned}$$

which equals zero, with the same argument as above. Applying Remark 4.4 yields the result. □

The cross-covariance functions C_{curl} and C_{div} are constructed in exactly the same way as in the case for RVFs on Euclidean spaces; see Theorem 2 in Scheuerer and Schlather [94]. This is not surprising, as the connection between both cases is given by a coordinate transformation which is implicitly contained in the Nabla operators on the sphere. The idea behind this construction is the well-known fact that a divergence field is curl-free and a curl field is divergence-free.

The Helmholtz-Hodge decomposition (cf. [4, Chapter 9.3, Theorem 3]) yields

that every smooth vector field V on $\mathbb{S}^2$ can be uniquely decomposed as

$$\mathsf{V}(x) = \mathsf{V}_{curl}(x) + \mathsf{V}_{div}(x), \qquad x \in \mathbb{S}^2,$$

where V_{curl} and V_{div} are curl- and divergence-free vector fields on $\mathbb{S}^2$, respectively. Consequently, every RVF on $\widetilde{\mathbb{S}}^2$ with a.s. differentiable sample paths can be decomposed similarly as

$$\mathsf{Z}(x) = \mathsf{Z}_{curl}(x) + \mathsf{Z}_{div}(x), \qquad x \in \widetilde{\mathbb{S}}^2,$$

where Z_{curl} and Z_{div} can be generated as in the proof of the theorem above. Fan and Matsuo [29] used this composition to construct covariance functions for general RVFs on $\mathbb{S}^2$.

Remark 4.7. Note that a mean-square differentiable RF need not have a.s. differentiable sample paths and vice versa. This implies that the assumption of differentiability of Z_{div} and Z_{curl} in part (b) of the theorem is essential. Scheuerer [93] gives a concise treatment of sample path properties for second-order RFs.

Example 4.8. Consider the function $C : \widetilde{\mathbb{S}}^2 \times \widetilde{\mathbb{S}}^2 \to \mathbb{R}$, given by

$$C(\varphi_1, \theta_1, \varphi_2, \theta_2) = \cos\theta_1 \cos\theta_2 \cos(\varphi_1 - \varphi_2).$$

This is a valid covariance function, as follows. The cosine function is positive definite on $\mathbb{R}$ and $f(x)f(y)$ is a positive definite kernel on $\mathbb{R} \times \mathbb{R}$ for any function $f : \mathbb{R} \to \mathbb{R}$. Now Corollary 3.1.13 in [9] directly implies that the tensor product $\cos(x_1 - x_2)f(y_1)f(y_2)$ is positive definite on $\mathbb{R}^2 \times \mathbb{R}^2$, implying the positive definiteness of C. Applying formulas (4.4) yields

$$\begin{aligned}
\mathsf{C}_{div}(\varphi_1, \theta_1, \varphi_2, \theta_2) &= \begin{pmatrix} \partial_{\theta_1}\partial_{\theta_2} & -\frac{1}{\cos\theta_2}\partial_{\varphi_2}\partial_{\theta_1} \\ -\frac{1}{\cos\theta_1}\partial_{\theta_2}\partial_{\varphi_1} & \frac{1}{\cos\theta_1\cos\theta_2}\partial_{\varphi_2}\partial_{\varphi_1} \end{pmatrix} C(\varphi_1, \theta_1, \varphi_2, \theta_2) \\
&= \begin{pmatrix} \sin\theta_1 \sin\theta_2 \cos(\varphi_1 - \varphi_2) & \sin\theta_1 \sin(\varphi_1 - \varphi_2) \\ -\sin\theta_2 \sin(\varphi_1 - \varphi_2) & \cos(\varphi_1 - \varphi_2) \end{pmatrix}
\end{aligned}$$

and

$$\mathsf{C}_{curl}(\varphi_1, \theta_1, \varphi_2, \theta_2) = \begin{pmatrix} \frac{1}{\cos\theta_1\cos\theta_2}\partial_{\varphi_1}\partial_{\varphi_2} & \frac{1}{\cos\theta_1}\partial_{\varphi_1}\partial_{\theta_2} \\ \frac{1}{\cos\theta_2}\partial_{\theta_1}\partial_{\varphi_2} & \partial_{\theta_1}\partial_{\theta_2} \end{pmatrix} C(\varphi_1, \theta_1, \varphi_2, \theta_2)$$

$$= \begin{pmatrix} \cos(\varphi_1 - \varphi_2) & \sin\theta_2 \sin(\varphi_1 - \varphi_2) \\ -\sin\theta_1 \sin(\varphi_1 - \varphi_2) & \sin\theta_1 \sin\theta_2 \cos(\varphi_1 - \varphi_2) \end{pmatrix}.$$

Figures 4.1, resp. 4.2, show realizations for the RVF Z_{div} corresponding to C_{div}, resp. for the RVF Z_{curl} corresponding to C_{curl}. In these plots, it is not easy to see that the realizations have the desired properties. Therefore, we also plot the realizations depending only on longitude and latitude; see Figures 4.3 and 4.4. In Figure 4.3, we clearly see two curls, one near the point $(1.7, 0)$ and the other one near $(3, 0)$, but no divergences. Figure 4.4 shows a vector field with two divergences, one near $(0, 0)$ and the other one near $(-2.8, 0)$, but with no curls, as expected.

Example 4.9. Jun [53] introduced nonstationary, multivariate covariance models on the sphere based on partial derivatives of univariate processes. Using univariate processes Y_k on the sphere, she defined the bivariate process Z via

$$\mathsf{Z}_i(\varphi, \theta) = \sum_{k=1}^{n} \left(a_{i,k}(\theta) Y_k^{(\theta)}(\varphi, \theta) + b_{i,k}(\theta) Y_k^{(\varphi)}(\varphi, \theta) \right) + g_i(\theta) Y_0(\varphi, \theta), \qquad (4.5)$$

for $i = 1, 2$, where $a_{i,k}, b_{i,k}$, and g_i are functions of the latitude, and Y_k are independent, mean-square differentiable, univariate processes. We can obtain Z_{div} and Z_{curl} as special cases of (4.5). Let $n \geq 1$ and $g_i \equiv 0$ for $i = 1, 2$. Then setting

$$a_{1,k}(\theta) \equiv 1, \qquad b_{2,k}(\theta) = -\frac{1}{\cos(\theta)},$$

and

$$b_{1,k}(\theta) \equiv 0 \equiv a_{2,k}(\theta),$$

for all $1 \leq k \leq n$ yields immediately Z_{div}. On the other hand, setting

$$b_{1,k}(\theta) = \frac{1}{\cos(\theta)}, \qquad a_{2,k}(\theta) \equiv 1,$$

and

$$b_{2,k}(\theta) \equiv 0 \equiv a_{1,k}(\theta)$$

for $1 \leq k \leq n$ yields Z_{curl}, as differentials are linear operators.

In the following we consider a specific example taken from [53]. Let $n = 1$ and

Y_1 be a stochastic process on $\mathbb{S}^2$ with Matérn covariance function

$$C(\varphi, \theta) = \mathcal{M}_\nu(d) = \left(\frac{d}{a}\right)^\nu K_\nu\left(\frac{d}{a}\right),$$

where

$$d = 2\left[\sin^2\left(\frac{\theta_1 - \theta_2}{2}\right) + \cos(\theta_1)\cos(\theta_2)\sin^2\left(\frac{\varphi_1 - \varphi_2}{2}\right)\right]^{1/2}$$

denotes the *chordal distance* between two points on the sphere; $\nu > 0$, resp. $a > 0$, denote the smoothness, resp. scale, of the Matérn covariance function; and K_ν is the modified Bessel function of the second kind with parameter ν. Following the calculations in Jun [53] and Jun and Stein [54] we obtain

$$\begin{aligned}
\partial_{\varphi_1}\partial_{\varphi_2} C &= -\frac{1}{4}h_3^2\mathcal{M}_{\nu-2}\left(\frac{d}{a}\right) + \frac{1}{2}h_{33}\mathcal{M}_{\nu-1}\left(\frac{d}{a}\right),\\
\partial_{\varphi_1}\partial_{\theta_2} C &= \frac{1}{4}h_2 h_3\mathcal{M}_{\nu-2}\left(\frac{d}{a}\right) - \frac{1}{2}h_{23}\mathcal{M}_{\nu-1}\left(\frac{d}{a}\right),\\
\partial_{\theta_1}\partial_{\varphi_2} C &= -\frac{1}{4}h_1 h_3\mathcal{M}_{\nu-2}\left(\frac{d}{a}\right) + \frac{1}{2}h_{13}\mathcal{M}_{\nu-1}\left(\frac{d}{a}\right),\\
\partial_{\theta_1}\partial_{\theta_2} C &= \frac{1}{4}h_1 h_2\mathcal{M}_{\nu-2}\left(\frac{d}{a}\right) - \frac{1}{2}h_{12}\mathcal{M}_{\nu-1}\left(\frac{d}{a}\right),
\end{aligned}$$

where the functions $h_i, h_{ij}, 1 \leq i, j \leq 3$, are given explicitly in [54, Appendix A]. These identities in combination with formulas (4.4) yield matrix-valued covariance functions for Z_{curl} and Z_{div}. Note that for the application of Theorem 4.5 we have to choose $\nu > 2$, otherwise C would not be in $\mathcal{C}^{(1,1)}$.

Figure 4.5 shows a realization of a RVF corresponding to a curl-free Matérn covariance model C_{curl} with smoothness $\nu = 5$ and scale $a = 1$. We see a very smooth RVF with several divergences but no curls, as expected.

By contrast, we see in Figure 4.6 a realization of a divergence-free Matérn covariance model C_{div} with smoothness $\nu = 3$ and scale $a = 0.5$. The RVF looks less smooth and shows some curls but no divergences.

This illustrates that the bivariate Matérn model provides a flexible way of modelling curl- or divergence-free RVFs on the sphere as the smoothness and scale parameters govern the size and smoothness of the structures in the realization.

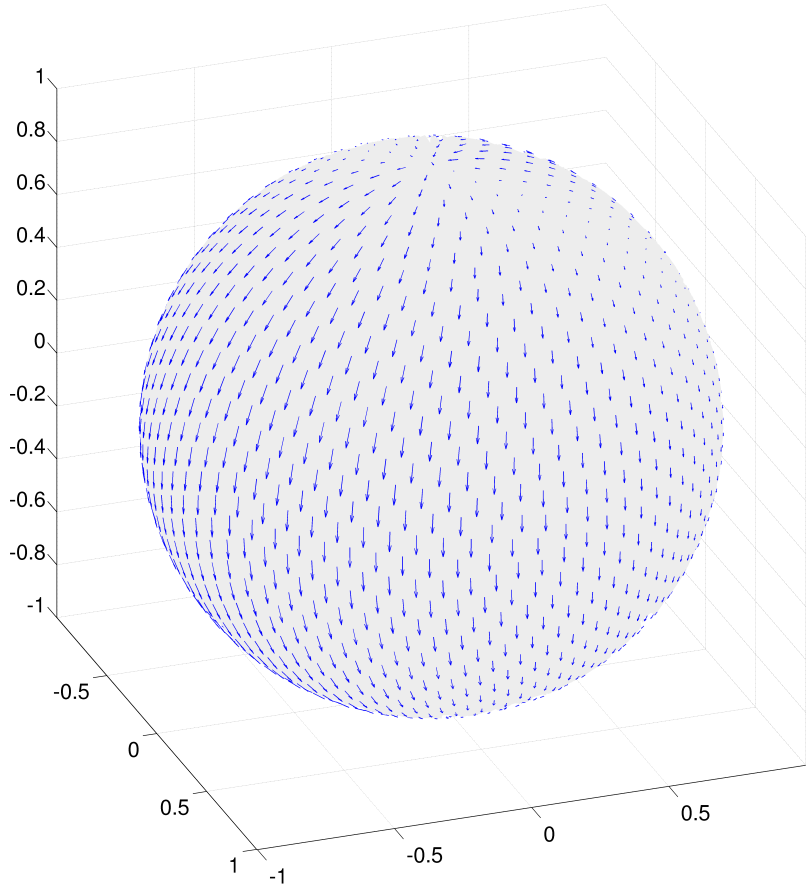


Figure 4.1: A realization of the divergence-free RVF Z_{div} given by C_{div} from Example 4.8.

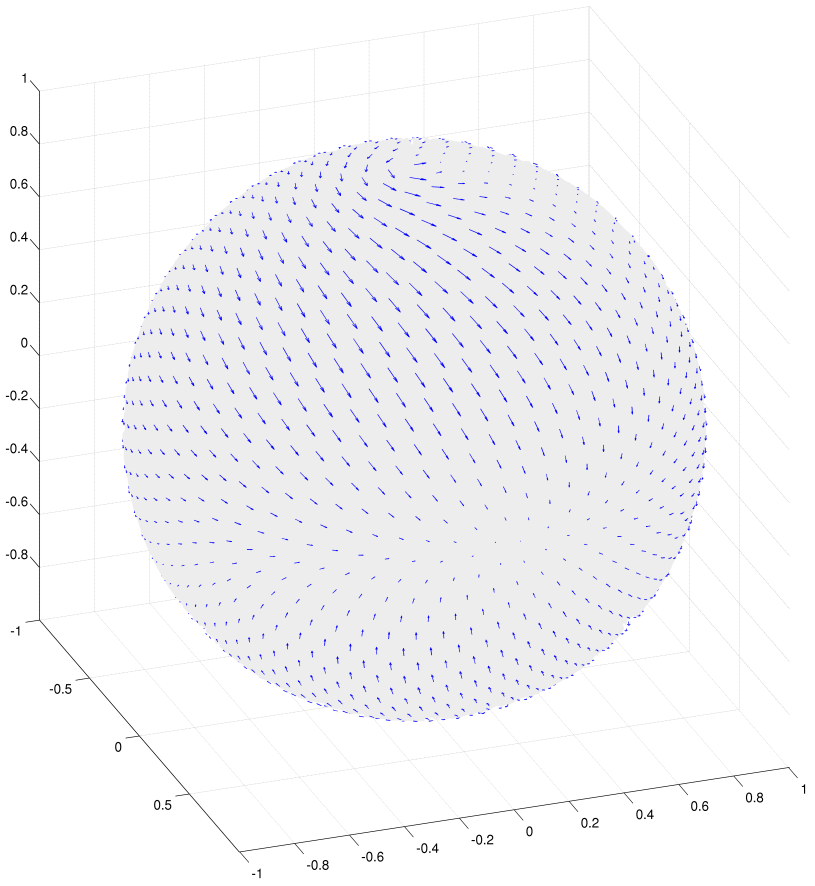


Figure 4.2: A realization of the curl-free RVF Z_{div} given by C_{div} from Example 4.8.

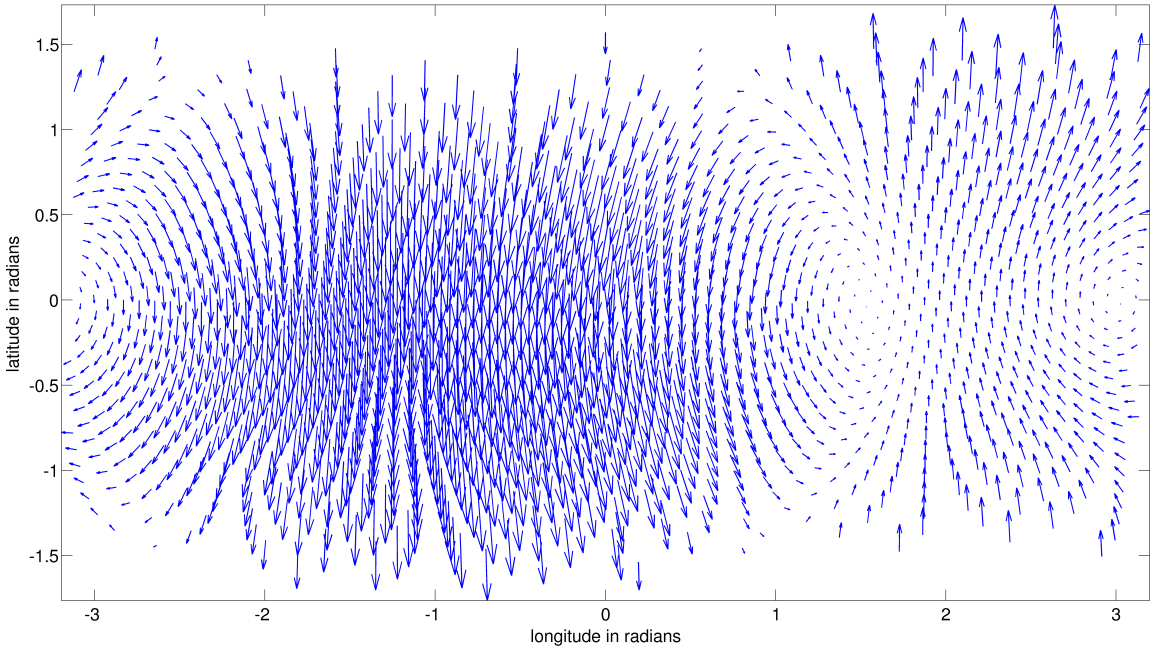


Figure 4.3: The realization of a divergence-free RVF from Figure 4.1 plotted in dependence of longitude and latitude.

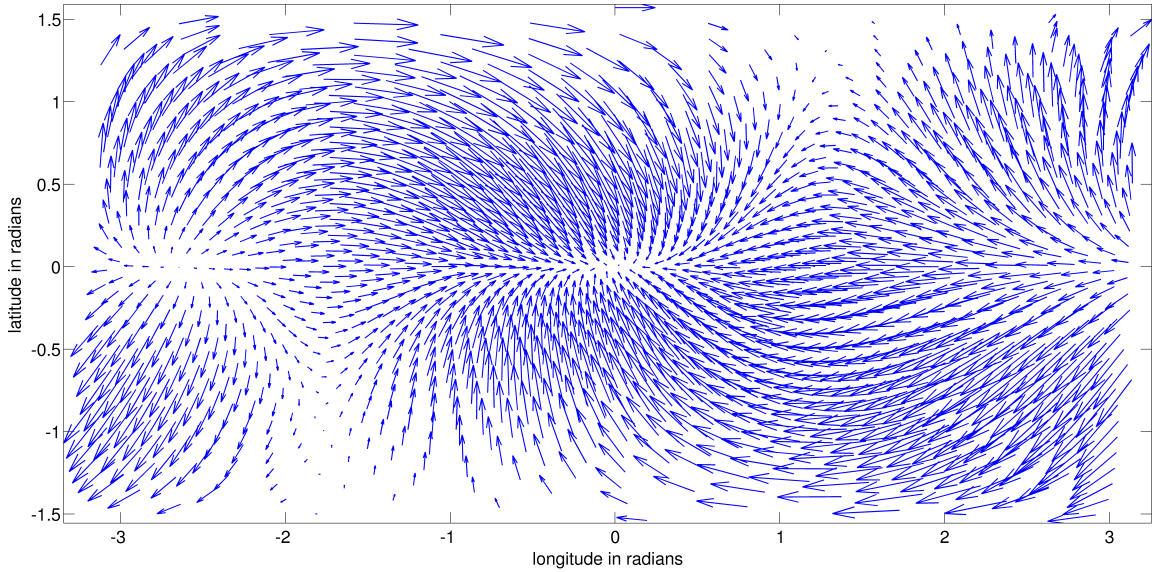


Figure 4.4: The realization of a curl-free RVF from Figure 4.2 plotted in dependence of longitude and latitude.

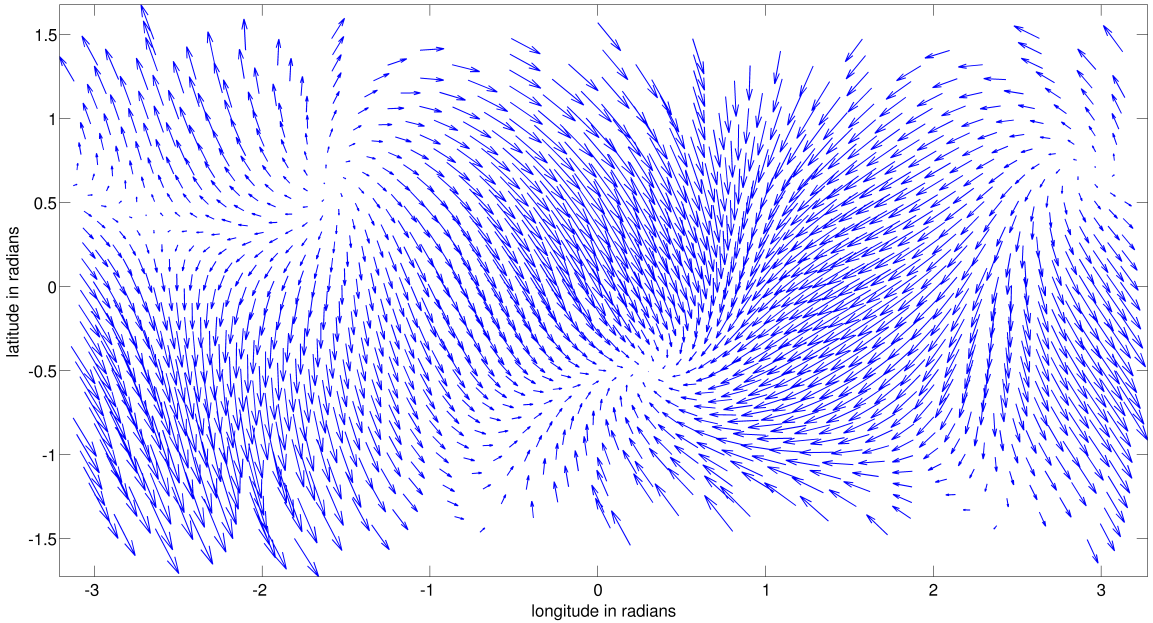


Figure 4.5: A realization of a curl-free Jun RVF with smoothness $\nu = 5$ and scale $a = 1$.

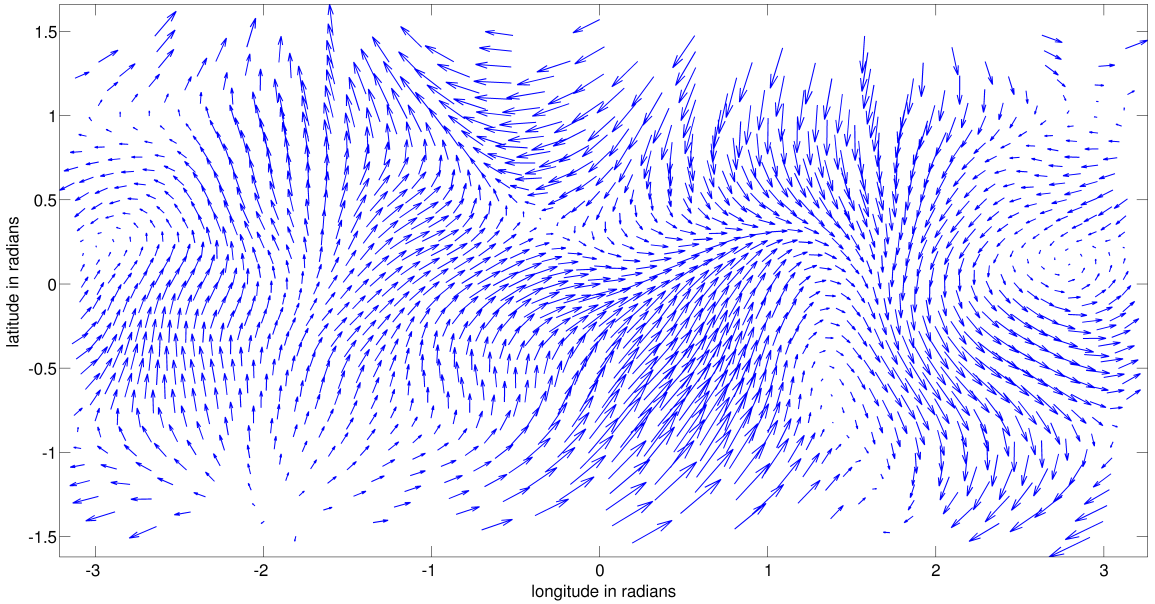


Figure 4.6: A realization of a divergence-free Jun RVF with smoothness $\nu = 3$ and scale $a = 0.5$.

4.3 Stochastic differential forms on embedded submanifolds of $\mathbb{R}^3$

We have shown in the foregoing section how to construct cross-covariance functions of curl- and divergence-free RVFs based on univariate covariance functions. Now the issue arises as to whether this construction principle is unique under certain circumstances, or whether there are further possibilities for constructing such cross-covariance functions. Theorem 4 of Scheuerer and Schlather [94] establishes the uniqueness for C_{curl} of RVFs on $\mathbb{R}^2$ and $\mathbb{R}^3$ and we show a similar result for C_{curl} of RVFs on the sphere.

To state and prove this result we introduce the concept of *stochastic differential forms*. The advantage of this approach lies in its generality: stochastic differential forms can be defined on any two- or three-dimensional submanifold of $\mathbb{R}^3$. Hence, they can be used whenever random flows have to be modelled on certain surfaces in $\mathbb{R}^3$. Another advantage is that many results of the rich theory of differential forms can be used, such as the Poincaré Lemma.

Throughout this section, M denotes a smooth submanifold of $\mathbb{R}^3$ with dimension $d = 2$ or 3 that can be covered by a single chart. By T_pM we denote the tangent space of M of point $p \in M$. All RFs and RVFs are assumed to be centered.

First, we define mean-square derivatives for RFs on M.

Definition 4.10. The RF $\{Z(p) : p \in M\}$ is mean-square differentiable in p along $v \in T_pM$ if there exists a RF $Z^{(v)}$ on M and, for some $\varepsilon > 0$, a C^1-curve $\gamma : [-\varepsilon, \varepsilon] \to M$ with $\gamma(0) = p$ and $\gamma'(0) = v$, such that

$$\mathbb{E}\left(\frac{Z(\gamma(\delta)) - Z(p)}{\delta} - Z^{(v)}(p)\right)^2 \longrightarrow 0, \qquad \delta \longrightarrow 0.$$

Example 4.11. We illustrate how this definition generalizes the mean-square differentiability for RFs on spheres, given in Definition 4.2. As mentioned before, for each $p \in M$ the tangent space T_pM is isomorphic to $\mathbb{R}^2$ for every $p \in M$, which we denote by $T_pM \cong \mathbb{R}^2$. Defining the curve γ_φ via its coordinate representation $\widetilde{\gamma}_\varphi(t) = (\varphi + t, \theta)$ we obtain $e_\varphi = \gamma'_\varphi(0)/\|\gamma'_\varphi(0)\|$ and correspondingly $Z^{(e_\varphi)}$, which coincides with $Z^{(\varphi)}$ from Definition 4.2. In a similar manner we obtain $Z^{(e_\theta)} = Z^{(\theta)}$.

Remark 4.12. If the univariate RF Z on M is mean-square differentiable in $p \in M$

in the direction $v \in T_pM$ we deduce for the covariance function

$$C : M \times M \to \mathbb{R}, \qquad (p, q) \mapsto \mathrm{Cov}(Z(p), Z(q)),$$

that

$$\mathrm{Cov}(Z^{(v)}(p), Z(q)) = D_v^1 C(p, q) \quad \text{and} \quad \mathrm{Cov}(Z(p), Z^{(v)}(q)) = D_v^2 C(p, q),$$

where D_v^i denotes the derivative of C with respect to the i-th variable in direction $v \in T_pM$.

In the remainder of this section we analyze how the sample path properties of RVFs on M are related to properties of the corresponding covariance function. In our analysis, we use the language of *differential forms* (Lee [64, Chapter 12]), as the study of vector fields on manifolds is closely connected to the study of differential forms on manifolds and there are many well-known results on differential forms on manifolds. Hence, we define stochastic differential forms on M.

Definition 4.13. (a) Let (M, h^{-1}) be the smooth chart on the n-dimensional manifold M, where $h^{-1} : M \to h^{-1}(M)$ is a diffeomorphism. A *stochastic k-form* is defined as

$$\zeta = \sum_I \zeta_I \, \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k},$$

where $I = (i_1 < \cdots < i_k)$ are ordered index-tuples with $1 \leq i_1, \ldots, i_k \leq n$; $x_1, \ldots, x_k$ are the corresponding coordinates and $\wedge$ denotes the outer product (for details we refer to [64, Chapter 12]). The coefficients ζ_I are univariate RFs on M, with a common probability space $(\Omega, \mathcal{A}, P)$ and with a.s. continuous sample paths.

The set of stochastic k-forms with coefficients ζ_I whose sample paths are almost surely l-times differentiable will be denoted by $\mathcal{A}_l^k(M)$.

(b) Let $\zeta \in \mathcal{A}_l^k(M), l \geq 1$, be given as above. The *exterior derivative* of ζ is defined as

$$\mathrm{d}\zeta = \sum_I \sum_{\alpha=1}^{n} \zeta_I^{(x_\alpha)} \, \mathrm{d}x_\alpha \wedge \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}.$$

Further, the k-form ζ is called *closed* if $\mathrm{d}\zeta(\omega) = 0$ for almost all $\omega \in \Omega$. A stochastic k-form $\zeta \in \mathcal{A}_0^k, k \geq 1$, is called *exact* if there exists $\mu \in \mathcal{A}_1^{k-1}(M)$

with $\mathrm{d}\mu(\omega) = \zeta(\omega)$ for almost all $\omega \in \Omega$.

Remark 4.14. For every vector field V on M there exists a dual differential form ζ_V, which is defined via

$$*\zeta_V = V \lrcorner \mathrm{d}M,$$

where $*$ denotes the *Hodge operator*, $\mathrm{d}M$ the *volume form* on M, and $\lrcorner$ denotes the *inner product*, see Agricola and Friedrich [4] for a detailed definition.

Given a RVF on M, the dual stochastic k-form can be found analogously.

Example 4.15. (1) Consider a RVF Z on $M = \mathbb{R}^3$, given as

$$\mathsf{Z}(x) = \mathsf{Z}_1(x)\, e_1 + \mathsf{Z}_2(x)\, e_2 + \mathsf{Z}_3(x)\, e_3.$$

The volume form equals $\mathrm{d}\mathbb{R}^3 = \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{d}x_3$ (see Example 25 in [4, Chapter 3]). Constructing the inner product yields

$$\mathsf{Z}(x) \lrcorner \mathrm{d}\mathbb{R}^3 = \mathsf{Z}_1(x)\, \mathrm{d}x_2 \wedge \mathrm{d}x_3 - \mathsf{Z}_2(x)\, \mathrm{d}x_1 \wedge \mathrm{d}x_3 + \mathsf{Z}_3(x)\, \mathrm{d}x_1 \wedge \mathrm{d}x_2.$$

Using the fact that the Hodge operator is self-inverse in this case (Chapter 1, Theorem 5 (1) in [4]) we obtain

$$\zeta_{\mathsf{Z}} = \mathsf{Z}_1\, \mathrm{d}x_1 + \mathsf{Z}_2\, \mathrm{d}x_2 + \mathsf{Z}_3\, \mathrm{d}x_3$$

as the dual stochastic 1-form of Z.

(2) Now let $\mathsf{Z} = \mathsf{Z}_1 e_\varphi + \mathsf{Z}_2 e_\theta$ be a RVF on $\widetilde{\mathbb{S}}^2$. The corresponding volume form is $\mathrm{d}\widetilde{\mathbb{S}}^2 = \cos\theta\, \mathrm{d}\varphi \wedge \mathrm{d}\theta$ (Example 27, Chapter 3 in [4]). Taking the inner product and applying the inverse of the Hodge operator yields

$$\mathsf{Z}(x) \lrcorner \mathrm{d}\widetilde{\mathbb{S}}^2 = \mathsf{Z}_1(x)\, \mathrm{d}\varphi + \cos\theta\, \mathsf{Z}_2(x)\, \mathrm{d}\theta.$$

(3) Consider spherical coordinates in $\mathbb{R}^3$ given by the transformation

$$x = r\cos\theta\cos\varphi, \quad y = r\cos\theta\sin\varphi, \quad z = r\sin\theta,$$

with $\geq 0, \varphi \in [0, 2\pi)$ and $\theta \in [-\pi/2, \pi/2]$. A RVF Z in these coordinates is

given as

$$\mathsf{Z}(r,\varphi,\theta) = \mathsf{Z}_1(r,\varphi,\theta)e_r + \mathsf{Z}_2(r,\varphi,\theta)e_\varphi + \mathsf{Z}_3(r,\varphi,\theta)e_\theta,$$

where the unit vectors are

$$\begin{aligned} e_r &= (\cos\theta, \cos\varphi, \sin\theta)^t, \\ e_\varphi &= (-\sin\varphi, \cos\varphi, 0)^t, \\ e_\theta &= (-\sin\theta\cos\varphi, -\sin\theta\cos\varphi, \cos\theta)^t. \end{aligned}$$

A similar procedure as in the cases (1) and (2) above yields the dual stochastic 1-form

$$\zeta_{\mathsf{Z}} = \mathsf{Z}_1\,\mathrm{d}r + r\cos\theta\,\mathsf{Z}_2\,\mathrm{d}\varphi + r\,\mathsf{Z}_3\,\mathrm{d}\theta.$$

Remark 4.16. We can express elegantly the curl and the divergence of a RVF Z on M in terms of its dual stochastic 1-form. With Definition 13 and Theorem 17 as found in [4, Chapters 2 and 3] we obtain

$$\mathrm{d}\zeta_{\mathsf{Z}} = *\zeta_{\mathrm{curl}(\mathsf{Z})}$$

and

$$\mathrm{d}(*\zeta_{\mathsf{Z}}) = \mathrm{div}(\mathsf{Z})\;\mathrm{d}M.$$

The remark shows that a RVF Z on M has a.s. curl-free sample paths if and only if its dual stochastic 1-form is closed. This shows that characterizing RVFs with curl-free sample paths on M is equivalent to characterizing closed stochastic 1-forms on M. This is done in the following result, which is a generalization of Theorem 4 in Scheuerer and Schlather [94].

Theorem 4.17. Let M be simply connected with dimension $d = 2$ or 3 and Z be a RVF on M. Further, assume ζ to be the corresponding stochastic 1-form and assume it is closed. Then ζ is exact, meaning that there is a univariate RF Y on M with $\mathrm{d}Y = \zeta$. Furthermore, given $v \in T_pM \cong \mathbb{R}^d$, the RF Y is mean-square differentiable along v with $Y^{(v)} = v \lrcorner \zeta$, i.e., $v \lrcorner \zeta = \zeta(v)$.

Proof. Let $(\Omega, \mathcal{A}, P)$ be the probability space belonging to ζ. We define Ω_0 to be the set of all $\omega \in \Omega$ such that $\mathrm{d}\zeta$ is not defined or the paths of the index RFs ζ_I are

not differentiable. If $\omega \in \Omega_0$ we set $Y(\cdot,\omega) \equiv 0$, and for all other $\omega \in \widetilde{\Omega} := \Omega \backslash \Omega_0$ we define

$$Y(p,\omega) = \int_0^1 \gamma^* \zeta(\omega), \qquad p \in M.$$

Here, $\gamma : [0,1] \to M$ is a piecewise smooth curve segment with $\gamma(1) = p$ and $\gamma(0) = p_0$, with $p_0 \in M$ an arbitrary point and $f^* g = g \circ f$ denotes the pullback of the function g by the function f. Now Y is again a RF on M and it does not depend on the particular choice of γ since ζ is closed. Furthermore, for each $\omega \in \widetilde{\Omega}$ we have

$$\mathrm{d}Y(\cdot,\omega) = \zeta(\omega).$$

Now we show that Y is mean-square partially differentiable along $v \in \mathbb{R}^d$. First note that for any piecewise differentiable curve $\widehat{\gamma} : [a,b] \to M$ joining $p_1, p_2 \in M$ we have

$$Y(p_1) - Y(p_2) = \int_a^b \widehat{\gamma}^* \zeta,$$

since we can reparametrize curves and the integral over closed curves is 0 and ζ is a closed stochastic form. For any fixed $p \in M$ we choose a chart (M, h^{-1}), with coordinate functions $x_1, \ldots, x_d$ corresponding to the standard basis in $\mathbb{R}^d$, and consider the differentiable curve $\gamma : [-\varepsilon, \varepsilon] \to M$, which we define via its coordinate representation $\widetilde{\gamma}$ as

$$\widetilde{\gamma}(t) = (\widetilde{\gamma}_1(t), \ldots, \widetilde{\gamma}_d(t)) = (h^{-1}(p)_1 + tv_1, \ldots, h^{-1}(p)_d + tv_d),$$

where we have used the fact that $h^{-1}(M) \subset \mathbb{R}^d \cong T_p M$. This can be done if we choose $\varepsilon > 0$ sufficiently small. Note that $\gamma(0) = p$ and $\gamma'(0) = v$.

The stochastic 1-form h^*Y on $V = h^{-1}(M)$ is given by $(h^*Y)(x) = \int_0^1 (h \circ \widetilde{\gamma})^* \zeta$, where $x = h^{-1}(p)$. Now $\widetilde{\gamma}$ is a curve joining x and $x + \varepsilon v$. Hence, we have for $\varepsilon \geq \delta > 0$

$$(h^*Y)^{(v,\delta)}(x) := \frac{1}{\delta} \int_0^\delta (h \circ \widetilde{\gamma})^* \zeta = \frac{1}{\delta} \int_0^\delta \widetilde{\gamma}^* (h^* \zeta),$$

and $(h^*\zeta)(x) = \sum_{j=1}^d \zeta_j(x)\, \mathrm{d}x_j$ on V. As can be seen in [64], Proposition 6.19, we can write

$$\int_0^\delta \widetilde{\gamma}^* (h^* \zeta) = \int_0^\delta \sum_{j=1}^d \zeta_j(\widetilde{\gamma}(t), \omega)(\widetilde{\gamma}_j)'(t)\, \mathrm{d}t$$

$$= \int_0^\delta \sum_{j=1}^d \zeta_j(\widetilde{\gamma}(t),\omega)v_j \,\mathrm{d}t = \int_0^\delta h^*(v \lrcorner \zeta)(\widetilde{\gamma}(t))\,\mathrm{d}t.$$

This results in

$$\mathbb{E}\left((h^*Y)^{(v,\delta)}(x) - h^*(v \lrcorner \zeta)(x)\right)^2 = \mathbb{E}\left(\frac{1}{\delta}\int_0^\delta h^*(v \lrcorner \zeta)(x+tv) - h^*(v \lrcorner \zeta)(x)\,\mathrm{d}t\right)^2$$

and this integral can be approximated using the same arguments as in the proof of Theorem 4 in [94]. This yields the claim. □

Remark 4.18. (1) Let $M = \widetilde{\mathbb{S}}^2$, Z be a RVF on M with a.s. curl-free sample paths, and ζ_{Z} be its dual stochastic 1-form. Then $\mathrm{d}\zeta_{\mathsf{Z}} = 0$ and the theorem above yields the existence of a RF Y on $\widetilde{\mathbb{S}}^2$ with $Y^{(v)} = v \lrcorner \zeta$ for every $v \in T_pM$ and covariance function C.

For the usual coordinates on $\widetilde{\mathbb{S}}^2$ we have $\mathsf{Z}(\varphi,\theta) = \frac{1}{\cos\theta}Y^{(e_\varphi)}e_\varphi + Y^{(e_\theta)}e_\theta$. Then it follows that

$$\mathrm{Cov}(\mathsf{Z}(\varphi_1,\theta_1),\mathsf{Z}(\varphi_2,\theta_2)) = \nabla_1\nabla_2^t C(\varphi_1,\theta_1,\varphi_2,\theta_2),$$

where ∇_i is the Nabla operator on the sphere, as defined in the previous section. This is exactly the form of C_{curl} of Theorem 4.5, and proves the uniqueness of the construction.

(b) Following [94], we can show in an analogous way that every divergence-free RVF on $\widetilde{\mathbb{S}}^2$ has a matrix-valued covariance function C_{div} of the form (4.4).

Let $\mathsf{Z} = \mathsf{Z}_1 e_\varphi + \mathsf{Z}e_\theta$ be a divergence-free RVF on $\widetilde{\mathbb{S}}^2$. Then, by (4.2) and (4.3), it follows that the curl of the RVF

$$\mathsf{X} := \mathsf{Z}_2 e_\varphi - \mathsf{Z}_1 e_\theta$$

equals zero. Now Theorem 4.17 implies that the matrix-valued covariance function of X is of the form $(\nabla_1\nabla_2^t)C$ for some univariate covariance function C. From this it follows readily that the matrix-valued covariance function of Z equals

$$\left[-\nabla_2\nabla_1^t + (\nabla_2^t\nabla_1)\mathbf{1}_2\right] C.$$

Consequently, the constructions given in Theorem 4.5 are unique.

4.4 Discussion

In this chapter, we provided criteria for random vector fields on the sphere to have almost surely curl- or divergence-free sample paths. We gave construction principles for matrix-valued covariance functions corresponding to the random vector fields possessing the respective sample path properties. These construction principles use univariate covariance functions in spherical coordinates, which are widely used in many applications, like meteorology. This allows a researcher to directly use our results to model processes with a desired physical property.

In practice, many processes are not static. Scheuerer and Schlather [94] pointed out that it is possible to use their results also for space-time processes: as the construction principle of the corresponding matrix-valued covariance functions is based on differential operators acting only on the spatial coordinates, an additional temporal variable can be safely ignored. Analogously, this also holds true in our case. Let $C(x, t, y, \tau)$ be a univariate, spatio-temporal covariance function on $(\widetilde{\mathbb{S}}^2 \times \mathbb{R}) \times (\widetilde{\mathbb{S}}^2 \times \mathbb{R})$ that is sufficiently smooth. Plugging C into (4.4) yields matrix-valued, spatio-temporal covariance functions corresponding to space-time random vector fields with almost surely curl- or divergence-free sample paths. This holds true, as the proof of Theorem 4.5 can be applied similarly in the spatio-temporal case: replacing the purely spatial processes by spatio-temporal processes and the purely spatial covariance functions by spatio-temporal covariance functions yields the claim.

5 On the dimple in spatio-temporal covariance functions

In recent years, there has been considerable interest in the construction of flexible spatio-temporal covariance models, as exemplified by the work of Cressie and Huang [17], Gneiting [37], Ma [67], Stein [105], Porcu, Mateu, and Christakos [82], Gneiting, Genton, and Guttorp [42], and Schlather [97]. Spatio-temporal covariance functions play key roles in the characterization of random processes $\{Z(s,t) : (s,t) \in \mathbb{R}^d \times \mathbb{R}\}$, where s represents a location in d-dimensional Euclidean space and t represents time. We generally assume second-order stationarity, i.e., that the covariance between $Z(s_1,t_1)$ and $Z(s_2,t_2)$ depends only on the lag vector $(s_1 - s_2, t_1 - t_2)$. Hence, the corresponding covariance function can be denoted by

$$C(h,u) = \mathrm{Cov}\{Z(s,t), Z(s+h,t+u)\}, \qquad (h,u) \in \mathbb{R}^d \times \mathbb{R},$$

where $h \in \mathbb{R}^d$ and $u \in \mathbb{R}$ are spatial and temporal lags, respectively.

The space-time covariance function is said to be *spatially isotropic* if $C(h_1,u) = C(h_2,u)$ whenever $\|h_1\| = \|h_2\|$, and it is said to be *fully symmetric* if

$$C(h,u) = C(-h,u) = C(h,-u) = C(-h,-u)$$

for all $h \in \mathbb{R}^d$ and $u \in \mathbb{R}$. Kent, Mohammadzadeh, and Mosammam [58] showed that the family of fully symmetric covariance models introduced by Gneiting [37] includes members with a *dimple*, in the sense that the correlation between the random variables $Z(s_{\text{here}}, t_{\text{now}})$ and $Z(s_{\text{there}}, t_{\text{then}})$ might be higher than the correlation between $Z(s_{\text{here}}, t_{\text{now}})$ and $Z(s_{\text{there}}, t_{\text{now}})$. Moreover, Kent, et al. argued that this property is counterintuitive, since it prevents $C(h,u)$ from being a monotonically decreasing function in u for every fixed h.

However, the dimple property arises naturally in environmental and geophysical applications, when quantities such as precipitation, wind speed, or air pollutants are affected by transport effects. Specifically, as observed by Gneiting, Genton, and Guttorp [42], in the presence of prevailing winds transport fields show spatio-

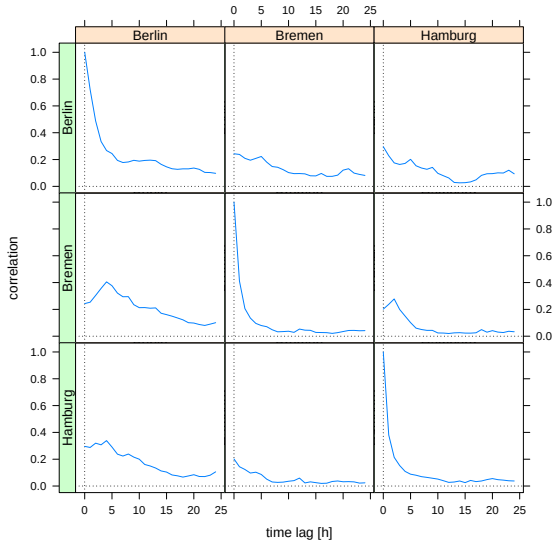


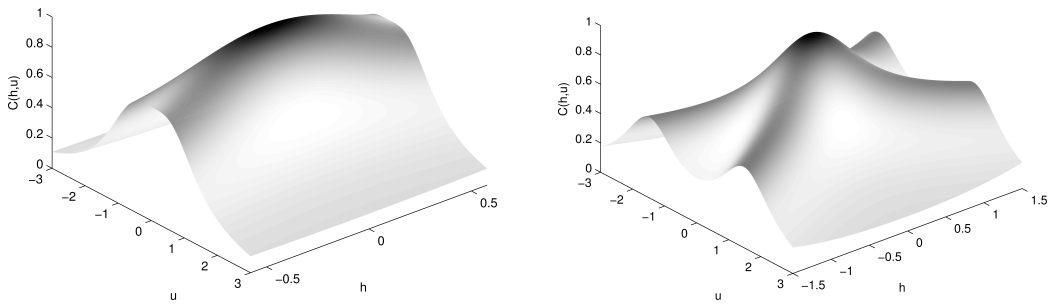
Figure 5.1: Empirical pairwise correlation for precipitation in three German cities.

temporal covariances with non-monotone behavior in certain directions. An example is shown in Figure 5.1, where hourly precipitation data are displayed for three German cities for the period December 1, 2003 through December 31, 2012. The empirical cross-correlation functions exhibit dimple-like behavior as a result of transport effects caused by predominantly westerly winds. Perhaps surprisingly, similar effects occur in situations in which the wind is uniform in all directions and the corresponding covariance function is spatially isotropic. Recent work of Huser and Davison [51, Fig. 3] suggests that dimples can also occur in empirical extremal coefficient functions.

To model transport effects we consider *transport fields* that give rise to covariance functions of the form

$$C(h,u) = \mathbb{E}\left\{C_0(h - Vu)\right\}, \qquad (h,u) \in \mathbb{R}^d \times \mathbb{R}, \tag{5.1}$$

where the *initial covariance* C_0 is purely spatial and stationary on $\mathbb{R}^d$, and the random vector $V \in \mathbb{R}^d$ represents the velocity. Uniform transport fields occur when C_0 is isotropic and V has a uniform distribution on a sphere with center at the origin, and these transport fields can give rise to spatially isotropic, fully symmetric covariance functions with the dimple property. Transport covariance functions of



(a) (b)

Figure 5.2: The transport covariance function (5.2) with initial covariance $C_0(h) = (1+h^2)^{-1}$. The dimple starts forming at $|h| = 1/\sqrt{3}$, the inflection point of C_0.

this and related forms have been studied by Cox and Isham [15], Ma [67], Gneiting, Genton, and Guttorp [42], and Schlather [97], among others. In Figure 5.2, we illustrate the dimple generated in one spatial dimension under a Cauchy initial covariance C_0.

We next introduce our key definition.

Definition 5.1. Suppose that $C(h,u)$, $(h,u) \in \mathbb{R}^d \times \mathbb{R}$ is a spatially isotropic covariance function. Then $C(h,u)$ *generates a dimple with onset at* $\kappa \geq 0$ if there is an $\epsilon > 0$ such that $C(h,u)$ has a strict local maximum at $u = 0$ for $|h| \leq \kappa$ and a strict local minimum at $u = 0$ for $\kappa < \|h\| < \kappa + \epsilon$. The dimple is *strong* if the onset is at $\kappa = 0$ and *regular* otherwise.

Remark 5.2. When introducing the concept of dimples, Kent, Mohammadzadeh, and Mosammam [58] used the following, slightly different definition of a dimple. They define a spatially isotropic and fully symmetric spatio-temporal covariance function C to have a dimple if there exists $\kappa > 0$ such that

(i) $C(h,u)$ is strictly decreasing in $u \geq 0$ for fixed $\|h\| \leq \kappa$, and

(ii) there exists u^* depending on $\|h\|$ with $C(h,u)$ strictly increasing on $(0,u^*)$ and strictly decreasing on (u^*,∞) for fixed $\|h\| > \kappa$.

Consequently, Definition 5.1 generalizes their definition, as spatially isotropic spatio-temporal covariance functions are also fully symmetric and (i) implies a

strict local maximum in $u = 0$ for $\|h\| \leq \kappa$, while (ii) implies a strict local minimum in $u = 0$ for $\|h\| > \kappa$.

In the next section, we investigate the formation of dimples in one spatial dimension and analyze them in statistical applications with the help of a simulation study in Section 5.2. Afterwards, we generalize the concept of a dimple to any spatial dimension in Section 5.3. We identify classes of initial covariance functions C_0 for which the transport covariance function (5.1) generates a dimple, characterise its onset, and illustrate our results using the powered exponential, Cauchy, and Matèrn initial covariances.

5.1 Dimples in one spatial dimension

In one spatial dimension, the transport covariance function (5.1) takes the form

$$C(h,u) = \frac{1}{2}\left\{C_0(h-u) + C_0(h+u)\right\}, \qquad (h,u) \in \mathbb{R}\times\mathbb{R}, \tag{5.2}$$

where C_0 is a stationary covariance function on $\mathbb{R}$. This can be associated with the superposition of two independent processes with covariance $C_0/2$ moving in opposite directions. Such processes are usually referred to as *frozen fields* [42, 18].

For now, we restrict attention to initial covariance functions C_0 with bell-shaped geometry, which we formalize as follows.

Definition 5.3. An even function $f : \mathbb{R} \to \mathbb{R}$ is weakly bell-shaped if

- f is continuous on $\mathbb{R}$;
- f decays to zero on $(0,\infty)$;
- f is twice differentiable on $(0,\infty)$;
- there is $c \geq 0$ such that f is strictly concave on $(0,c)$ and strictly convex on (c,∞).

We call the constant c in the above definition the *inflection point* of f, which is a slight abuse of convention when $c = 0$. In the latter case f has a cusp at the origin. See panel (a) of Figure 5.2 for an illustration of this effect on the shape of the dimple. Also note that the case $c = 0$ allows f to be convex on $(0,\infty)$.

Theorem 5.4. If C_0 is weakly bell-shaped, the transport covariance function (5.2) generates a dimple with onset at its inflection point.

Proof. Given $h \in \mathbb{R}$, consider $f_h(u) = C(h,u)$ for $u \in \mathbb{R}$, and let c denote the inflection point of C_0. If $\|h\| < c$ then f_h has a strict local maximum, and if $\|h\| > c$ a strict local minimum, at $u = 0$, because $f_h'(0) = 0$ and $f_h''(0) = C_0''(h)$ for $h \neq 0$. □

As the conditions of our results are invariant under transformations of the form $f(h) \mapsto \sigma\, f(ah)$ for $a, \sigma > 0$, there is no need to include scale and variance parameters in examples.

Example 5.5. Consider the *Cauchy family*

$$C_0(h) = (1 + |h|^\alpha)^{-\beta}, \qquad h \in \mathbb{R},$$

with smoothness parameter $0 < \alpha \leq 2$ and long-memory parameter $\beta > 0$ [44]. For all parameter values, the conditions of Theorem 5.4 are satisfied. When $0 < \alpha \leq 1$, C_0 is convex and the space-time covariance function (5.2) generates a strong dimple. When $1 < \alpha \leq 2$, the dimple is regular with onset at the inflection point

$$c = \left(\frac{\alpha - 1}{\alpha\beta + 1}\right)^{1/\alpha},$$

as illustrated in Figure 5.2 for $\alpha = 2$ and $\beta = 1$.

If the random vector V in (5.1) is not symmetrically distributed there is a prevailing transport direction, which might cause a moving dimple. We illustrate this in the following with a simple example. Let V be uniformly distributed on $\{-3/2, 1/2\}$. We can decompose $V = V_u + v_p$, where V_u is a random velocity uniformly distributed on $\{-1, 1\}$, and $v_p = -1/2$ is a deterministic term. Evidently, V_u corresponds to local random transport effects while v_p yields a prevailing transport effect. This explains the term moving dimple: if we move our inertial system with velocity v_p, i.e., if we make a Galilean transformation $\widetilde{h} = h + v_p u$, we have a uniform transport field in this transformed system, which causes a dimple there.

In Figure 5.3 we illustrate a moving dimple with the above transport effects,

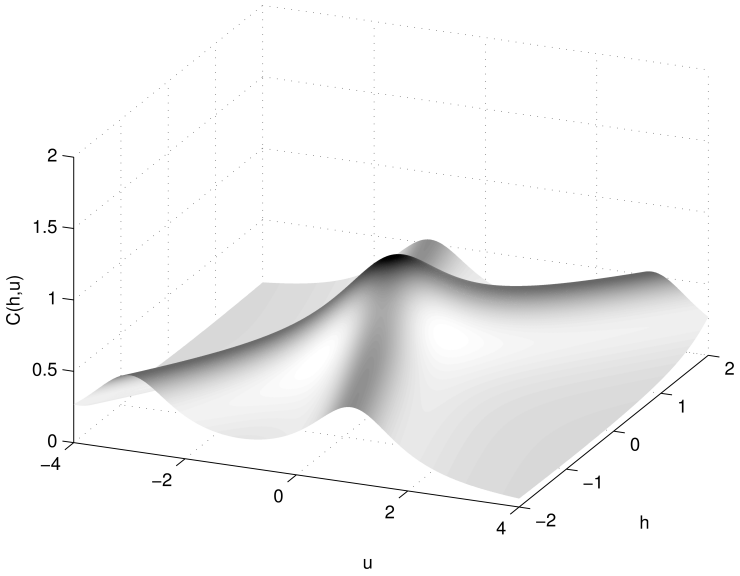


Figure 5.3: A moving dimple generated by model (5.3)

where we choose $C_0 = (1 + h^2)^{-1}$, which results in the covariance function

$$C(h,u) = \frac{2}{4 + (3u - 2h)^2} + \frac{2}{4 + (u + 2h)^2}. \tag{5.3}$$

This demonstrates that covariance models that admit the dimple property can also be applied when a prevailing transport effect occurs.

5.2 Simulation study

Before considering dimples in higher spatial dimensions, we investigate the relevance of the dimple phenomenon in statistical practice. To this end, we perform a simulation study considering the three spatio-temporal centered Gaussian RFs Z_{dimple}, Z_{transp}, and Z_{sep} on $\mathbb{R} \times \mathbb{R}$. These fields are determined by the following covariance functions:

$$C_{dimple}(h,u) = \frac{1}{2}\big[\exp(-a|h - vu|) + \exp(-a|h + vu|)\big], \qquad (h,u) \in \mathbb{R} \times \mathbb{R},$$

Simulated	Fitted	$\hat{a}$	$\hat{v}$
dimple	dimple	0.47	1.00
dimple	transport	0.53	0.91
dimple	separable	0.20	0.45
transport	dimple	0.57	0.99
transport	transport	0.56	1.00
transport	separable	0.26	0.56
separable	dimple	0.49	−10.0
separable	transport	0.51	−12.5
separable	separable	0.52	1.02

Table 5.1: The first and second column specify the type of process that was simulated and fitted, respectively. The third and the fourth column contain the corresponding estimates of the parameters a and v. Note that the true values of these parameters are $a = 1/2$ and $v = 1$.

which admits a dimple,

$$C_{transp}(h,u) = \exp(-a|h - vu|), \qquad (h,u) \in \mathbb{R} \times \mathbb{R},$$

admitting a unidirectional transport effect, and the separable covariance function

$$C_{sep}(h,u) = \exp(-a|h| - v|u|), \qquad (h,u) \in \mathbb{R} \times \mathbb{R},$$

respectively, where $a, v > 0$ are scale parameters. After setting $v = 1$ and $a = 1/2$, each RF is simulated once on the spatio-temporal grid $\{0,1,3,6\} \times \{1,\dots,3650\}$. These simulations are then separated into a training set consisting of the simulated values on $\{0,1,3,6\} \times \{1,\dots,1825\}$, and a test set consisting of the simulated values on $\{0,1,3,6\} \times \{1829,\dots,3650\}$. For every combination $i,j \in I = \{dimple, transp, sep\}$ the covariance function C_j is fitted to the empirical covariance function calculated from the training set of the realization Z_i. The fitting is done by non-weighted least square estimation, where we consider the temporal lags $-6,-5,\dots,5,6$. The thus-obtained estimates of the parameters a and v are denoted by $\hat{a}$ and $\hat{v}$, respectively, and are shown in Table 5.2.

Afterwards, these estimates are used for making one-time-step-ahead simple

kriging predictions[1] at all four sites and each pair of covariance model and realization, where we use the realizations on the last three time points as predictors. Specifically, on each point (s_0, t_0) of the corresponding test set the *simple kriging* point predictor for $Z_i(s_0, t_0), i \in I$, is given by

$$\hat{Z}_i(s_0, t_0) = \boldsymbol{c}_{i,0}' \boldsymbol{C}_i^{-1} \boldsymbol{z}_{i,0}.$$

Here the covariance function $C_i, i \in I$, with the estimated parameters is used to calculate the covariance matrix

$$\boldsymbol{C}_i = \left[C_i(x_k - x_l)\right]_{1 \le k,l \le 12},$$

and the vector

$$c_{i,0} = \left(C_i(x_0 - x_1), \ldots, C_i(x_0 - x_{12})\right)^t,$$

where $x_k \in \Gamma_p = \{0, 1, 3, 6\} \times \{t_0 - 3, t_0 - 2, t_0 - 1\}$ for all $1 \le k \le 12$. Further,

$$\boldsymbol{z}_{i,0} = \left(Z_i(x_1), \ldots, Z_i(x_{12})\right)^t, \qquad x_1, \ldots, x_{12} \in \Gamma_p,$$

is the vector consisting of the predictor variables. To judge the goodness of these predictions, we evaluate the predictive performance with the help of *Murphy diagrams*, which were very recently introduced by Ehm, et al. [28]. In particular, we use Murphy diagrams for the mean functional, which are defined as follows. Consider the point forecasts $x_1, \ldots, x_n$ and their corresponding realizations $y_1, \ldots, y_n$. Then a Murphy diagram for the mean functional is a plot of the function

$$s(\theta) = \frac{1}{n} \sum_{i=1}^{n} S_\theta(x_i, y_i), \tag{5.4}$$

where

$$S_\theta(x, y) = \begin{cases} |y - \theta|, & \min(x, y) \le \theta < \max(x, y) \\ 0, & \text{otherwise,} \end{cases}$$

and $\theta \in \mathbb{R}$. These diagrams can be used to compare pairs of forecasters graphically by their individual point forecasts, as follows. Let s_1, s_2 be calculated by (5.4) for each forecaster. If $s_1(\theta) \le s_2(\theta)$ for all $\theta \in \mathbb{R}$, then forecaster 1 *dominates* forecaster 2, in a theoretically well-defined sense [28].

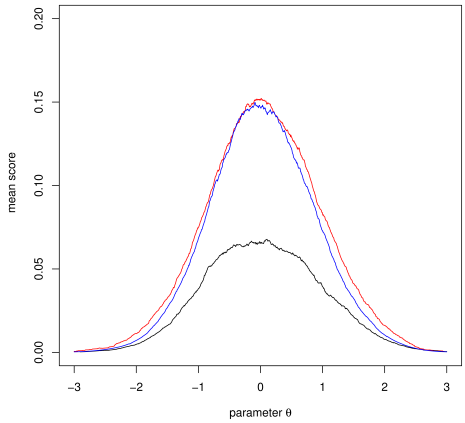
Figure 5.4 provides Murphy diagrams for the realizations for the dimple, unidi-

[1] See [13, Section 3.3] for an introduction into kriging.

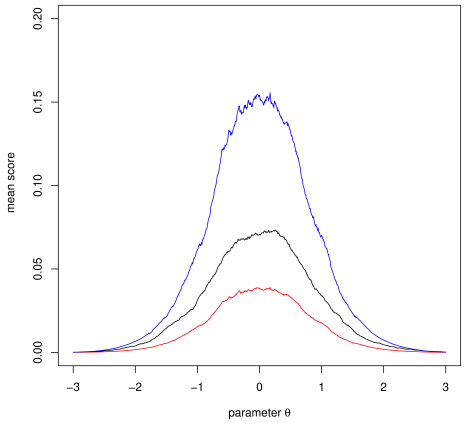
Simulated	Fitted	MSE
dimple	dimple	0.32
dimple	transport	0.62
dimple	separable	0.67
transport	dimple	0.30
transport	transport	0.15
transport	separable	0.60
separable	dimple	1.01
separable	transport	1.01
separable	separable	0.88

Table 5.2: The first two columns specify the type of process that was simulated and fitted, respectively. The third column contains the mean squared error of the corresponding fit.

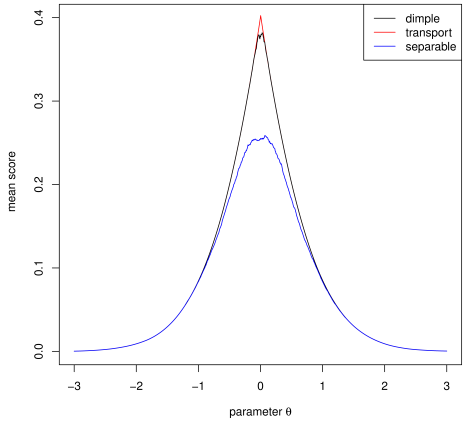
rectional transport, and separable process. Each forecaster uses one of the three different covariance functions for his predictions. We see that the forecaster using the correct model always outperforms the other forecasters, as her forecasts dominate the other forecasts. Panel (b) further shows that the dimple model performs significantly better than the separable model when the simulated data possesses a transport effect. In panels (a) and (c), the forecasters using the incorrect covariance models perform similarly. On the one hand, this indicates that using a model with dimples is more reasonable than using a model without any transport effect if the data were generated by a transport field. On the other hand, if the underlying field is not a transport field then models with transport effects do not appear to be good choices. These findings are confirmed if we compare the forecasters on the basis of the corresponding *mean squared errors (MSE)*, which are shown in Table 5.2. The actual values of the MSE are in very good accordance with the Murphy diagrams for the mean functional. Ehm, et al. [28] proved that the area under the Murphy diagram for the mean functional equals the corresponding MSE of the forecast, which can be confirmed by numerical integration.



(a) Dimple field



(b) Transport field



(c) Separable field

Figure 5.4: Murphy diagrams for the mean functional. The caption of each panel gives the type of the simulated field, and the color of the plot corresponds to the type of the fitted covariance function.

5.3 Dimples in higher dimensions

Clearly, spatio-temporal covariance functions in one spatial dimension are of limited interest, as most applications take place in two or three spatial dimensions. Therefore, this section addresses the analysis of dimples in higher spatial dimensions.

Let

$$\nu = \begin{pmatrix} \cos\varphi_1 \\ \sin\varphi_1 \cos\varphi_2 \\ \vdots \\ \sin\varphi_1 \cdots \sin\varphi_{d-2} \cos\varphi_{d-1} \\ \sin\varphi_1 \cdots \sin\varphi_{d-2} \sin\varphi_{d-1} \end{pmatrix}$$

be the unit vector normal to the surface of $\mathbb{S}^{d-1}$;

$$\mathrm{d}\sigma_{d-1} = \sin\varphi_{d-2} \sin^2\varphi_{d-3} \cdots \sin^{d-2}\varphi_1 \, \mathrm{d}\varphi_{d-1} \ldots \, \mathrm{d}\varphi_1$$

be the surface element of $\mathbb{S}^{d-1}$; and $\varphi_1, \ldots, \varphi_{d-1}$ be spherical coordinates, with φ_{d-1} ranging over $[0, 2\pi)$, while all other angles range over $[0, \pi]$. Further, let γ_{d-1} denote the surface area of $\mathbb{S}^{d-1}$. Then, for uniform transport fields in $d \geq 1$ spatial dimensions, equation (5.1) can be written as

$$C(h,u) = \frac{1}{\gamma_{d-1}} \int_{\mathbb{S}^{d-1}} C_0(h - u\nu) \, \mathrm{d}\sigma_{d-1}, \qquad (h,u) \in \mathbb{R}^d \times \mathbb{R}. \tag{5.5}$$

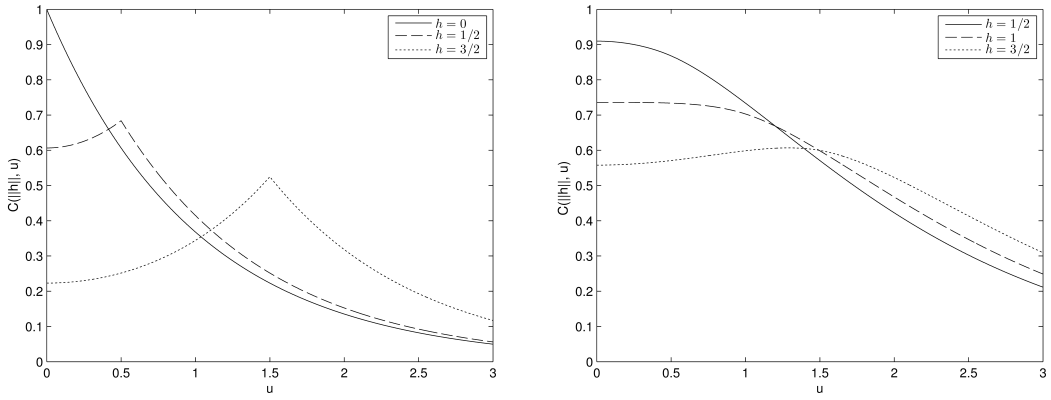
In the following, we study some parametric classes of initial covariance functions and the behaviour of the induced transport covariances.

Example 5.6. Consider the *powered exponential* family

$$C_0(h) = \exp(-\|h\|^\alpha), \quad h \in \mathbb{R}^d$$

with smoothness parameter $0 < \alpha \leq 2$. When $d = 1$ and $0 < \alpha \leq 1$ Theorem 5.4 implies the existence of a strong dimple, which is illustrated in panel (a) of Figure 5.5. Panel (b) shows a regular dimple with onset $\kappa = 1$ in the case $d = 2$.

Example 5.7. In the case of the Cauchy initial covariance $C_0(h) = (1 + \|h\|^2)^{-1}$,



(a) (b)

Figure 5.5: The transport covariance function (5.5) with initial covariance function $C_0(h) = \exp(-\|h\|)$ in dimension (a) $d = 1$, and (b) $d = 2$.

the resulting transport covariance (5.1) admits a closed-form expression in spatial dimension $d = 2$. Specifically,

$$\begin{aligned} C(h,u) =& \frac{1}{2\pi} \int_0^{2\pi} \left(1 + \left\|h - u\nu_\varphi\right\|^2\right)^{-1} \mathrm{d}\varphi \\ =& \frac{1}{2\pi} \int_0^{2\pi} \left(1 + u^2 + \|h\|^2 - 2u\|h\| \sin\varphi\right)^{-1} \mathrm{d}\varphi \\ =& \left(u^4 - 2u^2\|h\|^2 + 2u^2 + \|h\|^4 + 2\|h\|^2 + 1\right)^{-1/2}, \quad (h,u) \in \mathbb{R}^2 \times \mathbb{R}, \end{aligned}$$

where ν_φ is the normal to the unit circle, and where we have used identity (4.3.131) in Abramowitz and Stegun [2]. The zeroes of

$$\frac{\partial C(h,u)}{\partial u} = -2u \frac{u^2 - \|h\|^2 + 1}{(u^4 - 2u^2\|h\|^2 + 2u^2 + \|h\|^4 + 2\|h\|^2 + 1)^{3/2}}$$

are at $u_1 = 0$ and $u_{2,3} = \pm(\|h\|^2 - 1)^{1/2}$. Hence, $C(h,u)$ has a strict local maximum at $u = 0$ for $\|h\| < 1$ and a strict local minimum at $u = 0$ for $\|h\| > 1$. This implies that the dimple is regular with onset at $\kappa = 1$.

Acknowledgements

I would like to take the opportunity to thank all the people and organizations who helped me during this whole project.

First, and foremost, I want to express my sincerest gratitude to my supervisor, Prof. Dr. Tilmann Gneiting. With his wide and deep knowledge, his inspiration and motivation he gave me invaluable support throughout my Ph.D. study and research. He also provided me several opportunities to present my research and to get in touch with many great scientists. I would also like to say a huge 'thank you' to Prof. Dr. Donald Richards. His precise comments, questions and advices helped me immensely in finalizing my work. Further, I am also very grateful to Dr. Werner Ehm for his thoughtful counsel, which he provided me in numerous comments and long and very fruitful discussions regarding wide parts of this thesis.

I gratefully acknowledge the financial support by the German Research Foundation (DFG) within the program 'Spatio-/Temporal Graphical Models and Applications in Image Analysis' grant GRK 1653. I would also like to thank the Faculty of Mathematics and Computer Science of Heidelberg University for its financial and non-financial support.

Special thanks goes to Dr. Matthias Maier and Dr. Michael Scheuerer for many helpful discussions and counsel, not only concerning the content and the layout of this thesis.

Zu guter Letzt möchte ich allen meinen Freunden und ganz besonders meiner Familie danken. Ihre durchgängige seelische Unterstützung hat mir sehr bei der Fertigstellung dieses Projektes geholfen. Ein sehr großes Dankeschön geht dabei an meine Partnerin Jessica, deren Beistand mir über die gesamte Zeit meiner Promotion hinweg sehr viel Halt gegeben hat.

Bibliography

[1] P. Abrahamsen. *A review of Gaussian random fields and correlation functions*. Tech. rep. Oslo: Norwegian Computing Center, 1997. URL: http://publications.nr.no/917_Rapport.pdf.

[2] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. New York: Dover, 1972.

[3] R. J. Adler, R. E. Feldmann, and M. Taqqu, eds. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston: Birkhäuser, 1998.

[4] I. Agricola and T. Friedrich. *Globale Analysis*. Braunschweig: Vieweg, 2001.

[5] R. Askey and J. Fitch. "Integral representations for Jacobi polynomials and some applications". In: *Journal of Mathematical Analysis and Applications* 26 (1969), pp. 411–437.

[6] A. Astrauskas, J. B. Lévy, and M. Taqqu. "The asymptotic dependence structure of the linear fractional Lévy motion". In: *Lithuanian Mathematical Journal* 31 (1991), pp. 1–19.

[7] S. Banerjee. "On geodetic distance computations in spatial modeling". In: *Biometrics* 61 (2005), pp. 617–625.

[8] D. A. Benson, S. W Wheatcraft, and M. M. Meerschaert. "Application of a fractional advection-dispersion equation". In: *Water Resources Research* 36 (2000), pp. 1403–1412.

[9] C. Berg, J. P. R. Christensen, and P. Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*. New York: Springer, 1984.

[10] A. Blumen, G. Zumofen, and J. Klafter. "Transport aspects in anomalous diffusion: Lévy walks". In: *Physical Review A* 40 (1989), pp. 3964–3973.

[11] R. Cavoretto and A. De Rossi. "Fast and accurate interpolation of large scattered data sets on the sphere". In: *Journal of Computational and Applied Mathematics* 234 (2010), pp. 1505–1521.

[12] D. Chen, V. A. Menegatto, and X. Sun. “A necessary and sufficient condition for strictly positive definite functions on spheres”. In: *Proceedings of the American Mathematical Society* 131 (2003), pp. 2733–2740.

[13] J.-P. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. 2nd ed. New York: Wiley, 2012.

[14] J. Cohen, S. I. Resnick, and G. Samorodnitsky. “Sample correlations of infinite variance time series models: an empirical and theoretical study”. In: *Journal of Applied Mathematics and Stochastic Analysis* 11 (1998), pp. 255–282.

[15] D. R. Cox and V. S. Isham. “A simple spatial-temporal model for rainfall”. In: *Proceedings of the Royal Society of London Series A, Mathematical and Physical Sciences* 415 (1988), pp. 317–328.

[16] N. Cressie. *Statistics for Spatial Data*. Revised ed. New York: Wiley, 1993.

[17] N. Cressie and H.-C. Huang. “Classes of nonseparable, spatio-temporal stationary covariance functions”. In: *Journal of the American Statistical Association* 94 (1999), pp. 1330–1340.

[18] N. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. Hoboken, New Jersey: Wiley, 2011.

[19] D. J. Daley and Emilio Porcu. “Dimension walks and Schoenberg spectral measures”. In: *Proceedings of the American Mathematical Society* 142 (2014), pp. 1813–1824.

[20] R. A. Davis and T. Mikosch. “The extremogram: A correlogram for extreme events”. In: *Bernoulli* 15 (2009), pp. 977–1009.

[21] R. A. Davis and S. I. Resnick. “Limit theory for bilinear processes with heavy-tailed noise”. In: *The Annals of Applied Probability* 6 (1996), pp. 1191–1210.

[22] R. A. Davis and S. I. Resnick. “Limit theory for moving averages of random variables with regularly varying tail probabilities”. In: *The Annals of Probability* 13 (1985), pp. 179–195.

[23] L. de Haan, S. I. Resnick, H. Rootzén, and C. G. de Vries. “Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes”. In: *Stochastic Processes and their Applications* 32 (1989), pp. 213–224.

[24] P. J. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. 3rd ed. Boca Raton: Chapman & Hall/CRC, 2014.

[25] P. Doukhan and L. Truquet. “A fixed point approach to model random fields”. In: *Latin American Journal of Probability and Mathematical Statistics* 3 (2007), pp. 111–132.

[26] J. Dueck, D. Edelmann, and D. St. P. Richards. “Distance correlation coefficients for Lancaster distributions”. In: *Submitted* (2015). arXiv:1502.0141.

[27] J. Dueck, D. Edelmann, T. Gneiting, and D. St. P. Richards. “The affinely invariant distance correlation”. In: *Bernoulli* 20 (2014), pp. 2305–2330.

[28] W. Ehm, T. Gneiting, A. Jordan, and F. Krüger. “Of quantiles and expectiles: consistent scoring functions, Choquet representations, and forecast rankings”. In: *Journal of the Royal Statistical Society: Series B* 78 (2016), pp. 505–562.

[29] M. Fan and T. Matsuo. *Cross-covariance functions for tangent vector fields on the sphere*. Talk at ICSA/Graybill Conference, Fort Collins, 2015. URL: http://anson.ucdavis.edu/$\sim$minjay/Session_C03_Minjie_Fan.pdf.

[30] G. E. Fasshauer and L. L. Schumaker. “Scattered Data Fitting on the Sphere”. In: *Mathematical Methods for Curves and Surfaces: Volume II*. Ed. by M. Dæhlen, T. Lyche, and L. L. Schumaker. Nashville: Vanderbilt University Press, 1998, pp. 117–166.

[31] T. S. Ferguson. *A Course in Large Sample Theory*. London: Chapman & Hall/CRC, 1996.

[32] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics: Volume 2*. Reading, Massachusetts: Addison-Wesley, 1964.

[33] B. Finkenstadt, L. Held, and V. S. Isham, eds. *Statistics Methods for Spatio-Temporal Systems*. Boca Raton: Chapman & Hall/CRC, 2007.

[34] E. J. Fuselier. “Improved stability estimates and a characterization of the native space for matrix-valued RBFs”. In: *Advances in Computational Mathematics* 29 (2008), pp. 269–290.

[35] B. Garel and B. Kodia. “Signed symmetric covariation coefficient for α-stable dependence modeling”. In: *Comptes Rendus Mathematique* 347 (2009), pp. 315–320.

[36] A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, eds. *Handbook of Spatial Statistics*. Boca Raton: Chapman & Hall/CRC, 2010. ISBN: 9781420072877.

[37] T. Gneiting. “Compactly supported correlation functions”. In: *Journal of Multivariate Analysis* 83 (2002), pp. 493–508.

[38] T. Gneiting. “Criteria of Pólya type for radial positive definite functions”. In: *Proceedings of the American Mathematical Society* 129 (2001), pp. 2309–2318.

[39] T. Gneiting. “Nonseparable, stationary covariance functions for space-time data”. In: *Journal of the American Statistical Association* 97 (2002), pp. 590–600.

[40] T. Gneiting. “Strictly and non-strictly positive definite functions on spheres”. In: *Bernoulli* 19 (2013), pp. 1327–1349.

[41] T. Gneiting. *Supplement to* “Strictly and non-strictly positive definite functions on spheres". 2013. DOI: 10.3150/12-BEJSP06SUPP.

[42] T. Gneiting, M. G. Genton, and P. Guttorp. “Geostatistical space-time models, stationarity, separability and full symmetry”. In: *Statistical Methods for Spatio-Temporal Systems*. Ed. by B. Finkenstädt, L. Held, and V. S. Isham. Boca Raton: Chapman & Hall/CRC, 2006, pp. 151–175.

[43] T. Gneiting, W. Kleiber, and M. Schlather. “Matérn cross-covariance functions for multivariate random fields”. In: *Journal of the American Statistical Association* 105 (2010), pp. 1167–1177.

[44] T. Gneiting and M. Schlather. “Stochastic models that separate fractal dimension and Hurst effect”. In: *SIAM Review* 46 (2004), pp. 269–282.

[45] J. Guinness and M. Fuentes. “Isotropic covariance functions on spheres: Some properties and modeling considerations”. In: *Journal of Multivariate Analysis* 143 (2016), pp. 143–152.

[46] M. S. Handcock and J. R. Wallis. “An approach to statistical spatial-temporal modeling of meteorological fields”. In: *Journal of the American Statistical Association* 89 (1994), pp. 368–378.

[47] L. V. Hansen, T. L. Thorarinsdottir, E. Ovcharov, T. Gneiting, and D. St. P. Richards. “Lévy particles: Modelling and simulating star-shaped random sets”. In: *Advances in Applied Probability* 47 (2015), pp. 307–327.

[48] C. D. Hardin. “On the spectral representation of symmetric stable processes”. In: *Journal of Multivariate Analysis* 12 (1982), pp. 385–401.

[49] J. R. Holton and G. J. Hakim. *An Introduction to Dynamic Meteorology*. 5th ed. San Diego: Academic Press, 2013.

[50] C. Huang, H. Zhang, and S. Robeson. “On the validity of commonly used covariance and variogram functions on the sphere”. In: *Mathematical Geosciences* 43 (2011), pp. 721–733.

[51] R. Huser and A. C. Davison. “Space-time modelling of extreme events”. In: *Journal of the Royal Statistical Society: Series B* 76 (2014), pp. 439–461.

[52] R. H. Jones. “Stochastic processes on a sphere”. In: *Annals of Mathematical Statistics* 34 (1963), pp. 213–218.

[53] M. Jun. “Nonstationary cross-covariance models for multivariate processes on a globe”. In: *Scandinavian Journal of Statistics* 38 (2011), pp. 726–747.

[54] M. Jun and M. L. Stein. “An approach to producing space-time covariance functions on spheres”. In: *Technometrics* 49 (2007), pp. 468–479.

[55] M. Kanter. “Linear sample spaces and stable processes”. In: *Journal of Functional Analysis* 9 (1972), pp. 441–459.

[56] W. Karcher, H.-P. Scheffler, and E. Spodarev. “Efficient simulation of stable random fields and its applications”. In: *Stereology and Image Analysis. Ecs10 - Proceedings of the 10th European Congress of ISS*. Bologna, 2009, pp. 1–10.

[57] M. G. Kendall. “A new measure of rank correlation”. In: *Biometrika* 30 (1938), pp. 81–93.

[58] J. T. Kent, M. Mohammadzadeh, and A. M. Mosammam. “The dimple in Gneiting’s spatial-temporal covariance model”. In: *Biometrika* 98 (2011), pp. 489–494.

[59] S. Khan. “Nonlinear Dependence and Extremes in Hydrology and Climate”. PhD thesis. University of South Florida, 2007.

[60] A. Klenke. *Wahrscheinlichkeitstheorie*. 3rd ed. Berlin: Springer, 2013.

[61] D. G. Krige. “A statistical approach to some basic mine valuation problems on the Witwatersrand”. In: *Journal of the Chemical, Metallurgical and Mining Society of South Africa* 52 (1951), pp. 119–139.

[62] J. Kuelbs. “A representation theorem for symmetric stable processes and stable measures on H”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 26 (1973), pp. 259–271.

[63] A. Lang and C. Schwab. “Isotropic Gaussian random fields on the sphere: Regularity, fast simulation, and stochastic partial differential equations”. In: *Annals of Applied Probability* 6 (2015), pp. 3047–3094.

[64] J. M. Lee. *Introduction to Smooth Manifolds*. New York: Springer, 2003.

[65] B. Li and H. Zhang. "An approach to modeling asymmetric multivariate spatial covariance structures". In: *Journal of Multivariate Analysis* 102 (2011), pp. 1445–1453.

[66] G. G. Lorentz. "Fourier-Koeffizienten und Funktionenklassen". In: *Mathematische Zeitschrift* 51 (1948), pp. 135–149.

[67] C. Ma. "Long-memory continuous-time correlation models". In: *Journal of Applied Probability* 40 (2003), pp. 1133–1146.

[68] M. E. Machkouri, D. Volný, and W. B. Wu. "A central limit theorem for stationary random fields". In: *Stochastic Processes and their Applications* 123 (2013), pp. 1–14.

[69] R. Mandelbaum, B. Rowe, J. Bosch, C. Chang, F. Courbin, M. Gill, M. Jarvis, A. Kannawadi, T. Kacprzak, C. Lackner, A. Leauthaud, H. Miyatake, R. Nakajima, J. Rhodes, M. Simet, J. Zuntz, B. Armstrong, S. Bridle, M. Gentile, C. Heymans, A. S. Jurling, S. M. Kent, D. Kirkby, D. Margala, R. Massey, P. Melchior, J. Peterson, A. Roodman, and T. Schrabback. "The third gravitational lensing accuracy testing (GREAT3) challenge handbook". In: *The Astrophysical Journal Supplement Series* 212 (2014), pp. 1–28.

[70] B. Mandelbrot. "The variation of certain speculative prices". In: *The Journal of Business* 36 (1963), pp. 394–419.

[71] P. Maroni and Z. da Rocha. "Connection coefficients between orthogonal polynomials and the canonical sequence: an approach based on symbolic computation". In: *Numerical Algorithms* 47 (2008), pp. 291–314.

[72] E. Martínez-Gómez, M. T. Richards, and D. St. P. Richards. "Distance correlation methods for discovering associations in large astrophysical databases". In: *The Astrophysical Journal* 781 (2014), pp. 1–11.

[73] V. A. Menegatto. "Strictly positive definite kernels on the circle". In: *Rocky Mountain Journal of Mathematics* 25 (1995), pp. 1149–1163.

[74] G. Miller. "Properties of certain symmetric stable distributions". In: *Journal of Multivariate Analysis* 8 (1978), pp. 346–360.

[75] F. J. Narcowich and J. D. Ward. "Generalized hermite interpolation via matrix-valued conditionally positive definite functions". In: *Mathematics of Computation* 63 (1994), pp. 661–687.

[76] F. J. Narcowich and J. D. Ward. “Scattered data interpolation on spheres: error estimates and locally supported basis functions”. In: *SIAM Journal of Mathematical Analysis* 33 (2002), pp. 1393–1410.

[77] F. J. Narcowich, J. D. Ward, and G. B. Wright. “Divergence-free RBFs on surfaces”. In: *Journal of Fourier Analysis and Applications* 13 (2007), pp. 643–663.

[78] E. Netto. *Lehrbuch der Combinatorik*. 2nd ed. New York: Chelsea Publishing Company, 1927.

[79] F. W. J. Olver. *NIST Handbook of Mathematical Functions*. New York: Cambridge University Press, 2010.

[80] V. J. Paulauskas. “Some remarks on multivariate stable distributions”. In: *Journal of Multivariate Analysis* 6 (1976), pp. 356–368.

[81] K. Pearson. “Notes on regression and inheritance in the case of two parents”. In: *Proceedings of the Royal Society of London* 58 (1895), pp. 240–242.

[82] E. Porcu, J. Mateu, and G. Christakos. “Quasi-arithmetic means of covariance functions with potential applications to space-time data”. In: *Journal of Multivariate Analysis* 100 (2009), pp. 1830–1844.

[83] S. J. Press. “Multivariate stable distributions”. In: *Journal of Multivariate Analysis* 2 (1972), pp. 444–462.

[84] S. I. Resnick. “Heavy tail modeling and teletraffic data”. In: *The Annals of Statistics* 25 (1997), pp. 1805–1849.

[85] S. I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer, 2007.

[86] S. I. Resnick, G. Samorodnitsky, and F. Xue. “Growth rates of sample covariances of stationary symmetric α-stable processes associated with null recurrent Markov chains”. In: *Stochastic Processes and their Applications* 85 (2000), pp. 321–339.

[87] S. I. Resnick, G. Samorodnitsky, and F. Xue. “How misleading can sample acf’s of stable MA’s be? (Very!)” In: *Annals of Applied Probability* 9 (1999), pp. 797–817.

[88] M. T. Richards, D. St. P. Richards, and E. Martínez-Gómez. “Interpreting the distance correlation results for the COMBO-17 survey”. In: *The Astrophysical Journal Letters* 784 (2014), pp. 1–5.

[89] E. Rio. “Moment inequalities for sums of dependent random variables under projective conditions”. In: *Journal of Theoretical Probability* 22 (2009), pp. 146–163.

[90] J. Rosiński. “On stochastic integral representation of stable processes with sample paths in Banach spaces”. In: *Journal of Multivariate Analysis* 20 (1986), pp. 277–302.

[91] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Boca Raton: Chapman & Hall/CRC, 1994.

[92] Z. Sasvári. *Positive Definite and Definitizable Functions*. Berlin: Akademie Verlag, 1994.

[93] M. Scheuerer. “Regularity of the sample paths of a general second order random field”. In: *Stochastic Processes and their Applications* 120 (2010), pp. 1879–1897.

[94] M. Scheuerer and M. Schlather. “Covariance models for divergence-free and curl-free random vector fields”. In: *Stochastic Models* 28 (2012), pp. 433–451.

[95] M. Schilder. “Some structure theorems for the symmetric stable laws”. In: *The Annals of Mathematical Statistics* 41 (1970), pp. 412–421.

[96] M. Schlather. “Models for stationary max-stable random fields”. In: *Extremes* 5 (2002), pp. 33–44.

[97] M. Schlather. “Some covariance models based on normal scale mixtures”. In: *Bernoulli* 16 (2010), pp. 780–797.

[98] V. Schmidt, ed. *Stochastic Geometry, Spatial Statistics and Random Fields: Models and Algorithms*. Heidelberg: Springer, Lecture Notes in Mathematics, vol. 2100, 2015.

[99] I. J. Schoenberg. “Metric spaces and positive definite functions”. In: *Transactions of the American Mathematical Society* 44 (1938), pp. 522–536.

[100] I. J. Schoenberg. “Positive definite functions on spheres”. In: *Duke Mathematical Journal* 9 (1942), pp. 96–108.

[101] L. J. Slater. *Generalized Hypergeometric Functions*. Cambridge: Cambridge University Press, 1966.

[102] C. Spearman. “The proof and measurement of association between two things”. In: *The American Journal of Psychology* 15 (1904), pp. 72–101.

[103] E. Spodarev and W. Karcher. “Spatial risk analysis and modelling in insurance”. In: *Gumbel lecture in the context of Statistische Woche* (2007). URL: https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/mitarbeiter/spodarev/publications/GumbelLecture.pdf.

[104] E. Spodarev, E. Shmileva, and S. Roth. *Extrapolation of stationary random fields*. Tech. rep. 2013. arXiv:1306.6205v1.

[105] M. L. Stein. “Space-time covariance functions”. In: *Journal of the American Statistical Association* 100 (2005), pp. 310–321.

[106] Kirstin Strokorb, Felix Ballani, and Martin Schlather. “Tail correlation functions of max-stable processes”. In: *Extremes* 18 (2015), pp. 241–271.

[107] Gábor Szegö. *Orthogonal Polynomials*. Providence: American Mathematical Society, 1959.

[108] G. J. Székely, M. L. Rizzo, and N. K. Bakirov. “Measuring and testing dependence by correlation of distances”. In: *The Annals of Statistics* 35 (2007), pp. 2769–2794.

[109] V. V. Uchaikin and V. M. Zolotarev. *Chance and Stability: Stable Distributions and Their Applications*. Utrecht: VSP, 1999.

[110] H. Wackernagel. *Multivariate Geostatistics*. 3rd ed. Berlin: Springer, 2003.

[111] L. A. Waller and C. A. Gotway. *Applied Spatial Statistics for Public Health Data*. Hoboken: Wiley-Interscience, 2004.

[112] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Berlin: Springer, 2003.

[113] H. Wendland. *Scattered Data Approximation*. New York: Cambridge University Press, 2005.

[114] W. B. Wu. “Nonlinear system theory: Another look at dependence.” In: *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005), pp. 14150–14154.

[115] Y. Xu and E. W. Cheney. “Strictly positive definite functions on spheres”. In: *Proceedings of the American Mathematical Society* 116 (1992), pp. 977–981.

[116] A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions. Volume I: Basic Results*. New York: Springer, 1987.

[117] Z. Zhou. “Measuring nonlinear dependence in time-series, a distance correlation approach”. In: *Journal of Time Series Analysis* 33 (2012), pp. 438–457.

[118] J. Ziegel. “Convolution roots and differentiability of isotropic positive definite functions on spheres”. In: *Proceedings of the American Mathematical Society* 142 (2014), pp. 2063–2077.

Nomenclature

C^{ds} distance covariance function, page 27

$\mathrm{C}^{\mathrm{ds}}_{P,Q}$ cross-distance covariance function, page 27

γ^{ds} normalized distance variogram, page 42

R^{ds} distance correlation function, page 27

$\mathrm{R}^{\mathrm{ds}}_{P,Q}$ cross-distance correlation function, page 27

$|||X|||_\alpha$ scale parameter of the univariate, α-stable random variable X, page 12

$L^\alpha(E,m)$ space of α-integrable functions, page 14

$S\alpha S$ symmetric α-stable, page 9

$X \sim \mathcal{S}_\alpha(\Lambda,\mu)$ X is α-stable with spectral representation (Λ,μ), page 10

$X \overset{d}{\sim} Y$ the random vectors X and Y have the same distribution, page 8

DCF distance correlation function, page 27

EDCF empirical distance correlation function, page 28

i.i.d. independent and identically distributed, page 6

RF random field, page 5

RVF random vector field, page 95