Extrapolation of Stationary Random Fields

Evgeny Spodarev, Elena Shmileva and Stefan Roth

Abstract We introduce basic statistical methods for the extrapolation of stationary random fields. For square integrable fields, we set out basics of the kriging extrapolation techniques. For (non–Gaussian) stable fields, which are known to be heavy tailed, we describe further extrapolation methods and discuss their properties. Two of them can be seen as direct generalizations of kriging.

1 Introduction

In this chapter, we consider the problem of extrapolation (prediction) of random fields arising mainly in geosciences, mining, oil exploration, hydrosciences, insurance, etc. It is one of the fundamental tools in geostatistics that provides statistical inference for spatially referenced variables of interest. Examples of such quantities are the amount of rainfall, concentration of minerals and vegetation, soil texture, population density, economic wealth, storm insurance claim amounts, etc.

The origins of geostatistics as a mathematical science can be traced back to the works by B. Mathérn (1960) [26], L. Gandin (1963) [9], G. Matheron (1962-63) [27, 28]. However, the mathematical foundations were already laid in the paper by A.N.Kolmogorov (1941) [13] as well as in the book by N.Wiener (1949) [45], where the extrapolation of stationary time series was studied, whereas their practical ap-
Evgeny Spodarev, Elena Shmileva and Stefan Roth

Application is known since 1951 due to mining engineer D. G. Krige [20]. Typical practical problems to solve are e.g. plotting the contour concentration map of minerals (interpolation), inference of the the mean areal precipitation and evaluation of accuracy of the estimate from spatial measurements (averaging or generalization), selection of locations of new monitoring points so that the concentration can be evaluated with sufficient accuracy (monitoring network design).

The remainder of this chapter is divided into three sections. Section 2 contains preliminaries about distributional invariance properties and dependence structure of random fields. In Section 3, we concentrate on kriging which is a widely used probabilistic extrapolation technique for the fields with the finite second moment. Section 4 contains recent results on the extrapolation of heavy tailed random fields with infinite variance, namely of stable random fields.

In Sections 2 and 3 we mainly follow the books [4, 5, 38, 42]. Section 4 is based on the paper [16], it also contains some new results for stable fields with the infinite first moment, see Section 4.4.

2 Basics of Random Fields

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

**Definition 1.** A random field $X$ is a random function on $(\Omega, \mathcal{F}, P)$ indexed by points of $\mathbb{R}^d$, $d \in \mathbb{N}$, i.e. $X$ is a measurable mapping $X : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}$.

For an introduction into the theory of random functions see e.g. [38, Chap. 9].

2.1 Random Fields with Invariance Properties

A random field with the finite-dimensional distributions that are invariant with respect to the action of a group $G$ of transformations of $\mathbb{R}^d$ is called $G$-invariant in strict sense. In case if this invariance is given only for the first two moments of the field which are assumed to be finite we speak about the $G$–invariance in wide sense. Thus, if $G$ is the group of all translations of $\mathbb{R}^d$ then one calls such random fields stationary (in respective sense). For $G$ being the group of rotations $SO_d$ one claims the random field to be isotropic. If $G$ is the group of all rigid motions then such field is called motion invariant. The same notions of invariance can be transferred to the increments of random fields. In this case, the stationarity is often called intrinsic. The intrinsic stationarity in wide sense is called intrinsic stationarity of order two. For more details on invariance properties confer [38, Sect. 9.5].

**Exercise 1.** Show that the expectation (if it exists) of any process $(d = 1)$ with stationary increments is a linear function, i.e., $EX(t) = a \cdot t + c$ for all $t \in \mathbb{R}$, $a \in \mathbb{R}$, $c \in \mathbb{R}$. 
A popular class of random fields are Gaussian fields.

**Definition 2.** A random field $X = \{X(t), \ t \in \mathbb{R}^d\}$ is Gaussian if all its finite dimensional distributions are Gaussian.

Their use for modelling purposes in applications is explained mainly by the simplicity of their construction and analytic tractability combined with the normal distribution of marginals which describes many real phenomena due to the Central Limit Theorem.

By Kolmogorov’s theorem, the probability law of a Gaussian random field is defined uniquely by its mean value and covariance function; see [38, Sect. 9.2.2] for more details. If the mean value function $E X(t), \ t \in \mathbb{R}^d$ is identically zero we call $X$ to be centered. Without loss of generality we tacitly assume all random fields of this chapter to be centered.

**Exercise 2.** Show that for Gaussian random fields stationarity (isotropy, motion invariance) in strict sense and stationarity (isotropy, motion invariance) in wide sense are equivalent. In this case we call a Gaussian field just stationary (isotropic, motion invariant).

**Examples of Gaussian Random Fields**

1. **Ornstein-Uhlenbeck Process**

A centered Gaussian process $X = \{X(t), \ t \in \mathbb{R}\}$ with the covariance function $E (X(s)X(t)) = e^{-|s-t|/2}, \ s, t \in \mathbb{R}$ is called Ornstein-Uhlenbeck Process. Breiman (1968) [2, p. 350] has shown that $X$ is the only stochastically continuous stationary Markov Gaussian process. Additionally, it has short memory, i.e.,

$$X(t) \overset{d}{=} e^{-t/2}X(0) + V(t), \ t > 0,$$

where $V(t)$ does not depend on the past $\{X(s), s \leq 0\}$, cf. [23] Example 2.6, p.11]. Defined on $\mathbb{R}_+$, $X$ is the strong solution of the Langevin stochastic differential equation

$$dX(t) = -1/2X(t)dt + dW(t)$$

with initial value $X(0) \sim N(0, 1)$, where $W = \{W(t), t \geq 0\}$ is the standard Wiener process, see e.g. [3] Chapt. 8, Theorem 7]. It holds also $X \overset{d}{=} \{e^{-t/2}W(e^t), \ t \in \mathbb{R}\}$, cf. [3] Chapt. 3, p.107].

2. **Gaussian Linear Random Function**

A Gaussian linear random function $X = \{X(t), \ t \in l_2\}$ is defined by $X(t) = \langle N, t \rangle_2$, $t \in l_2$, where $N = \{N_i\}_{i=1}^\infty$ is an i.i.d. sequence of $N(0, 1)$-random variables, and $l_2$ is
the Hilbert space of sequences $t = \{t_i\}_{i=1}^{\infty}$ such that $\|t\|_2^2 := \sum_{i=1}^{\infty} t_i^2 < \infty$ with scalar product $(s,t)_2 = \sum_{i=1}^{\infty} s_i t_i$, $s,t \in l_2$. Since $N$ is not an element of $l_2$ a.s., the expression $\langle N,t \rangle_2$ is understood formally as the series $\sum_{i=1}^{\infty} N_i t_i$ which converges in the mean square sense:

$$\mathbb{E} \left| \sum_{i=n}^{m} N_i t_i \right|^2 = \sum_{i=n}^{m} t_i^2 \to 0, \quad n,m \to \infty.$$ 

It holds

$$X(t) \sim N(0,\|t\|_2^2), \quad X(t) - X(s) = X(t-s), \quad \mathbb{E} (X(s)X(t)) = \langle s,t \rangle_2, \quad s,t \in l_2.$$ 

Its variogram $\gamma(h) := 1/2 \cdot \mathbb{E} (X(t+h) - X(t))^2$ can be computed as

$$\gamma(h) = \frac{1}{2} \mathbb{E} |X(h)|^2 = \frac{\|h\|^2}{2}, \quad h \in l_2,$$

see more about variograms in Sect. 2.2.2. Here we have $\gamma(h) \to \infty$ as $\|h\|_2 \to \infty$. Transferring the notions of stationarity from the index space $\mathbb{R}$ to $l_2$, it is clear that $X$ is intrinsic stationary of order two but not wide sense stationary. Confer [12] for the general theory of Gaussian random functions on Hilbert index spaces.

3. Fractional Brownian Field

A fractional Brownian field $X = \{X(t), \ t \in \mathbb{R}^d\}$ is a centered Gaussian field with covariance (see more about covariance in Sect. 2.2.1)

$$\mathbb{E} (X(s)X(t)) = \frac{1}{2} \left( \|s\|^{2H} + \|t\|^{2H} - \|s-t\|^{2H} \right), \quad s,t \in \mathbb{R}^d$$

for some $H \in (0,1]$ where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$. Parameter $H$ (often called Hurst index) is responsible for the regularity of the paths of $X$. The greater $H$, the smoother are the paths. For $d = 1$, $X$ is called the fractional Brownian motion, including the two-sided Wiener process (defined on the whole $\mathbb{R}$) if $H = 1/2$. In the case $d > 1$, $H = 1/2$ it is called the Brownian Lévy field (see, e.g., [23 Sect. 2]).

It is easy to check that $X$ is intrinsically stationary of order two and isotropic. Its variogram $\gamma(h) = 1/2 \cdot \|h\|^{2H}$ is clearly motion invariant. This field is not wide sense stationary as its variance is not constant.

Exercise 3. Show that $X$

1. has stationary increments which are positively correlated for $H \in (1/2,1)$ and negatively correlated for $H \in (0,1/2)$.
2. is $H$–self similar, i.e., $X(\lambda t) \overset{d}{=} |\lambda|^H X(t)$ for all $\lambda \in \mathbb{R}$ and $t \in \mathbb{R}^d$.
3. has a version with a.s. Hölder continuous paths of any order $\beta \in (0,H)$.
4. has nowhere differentiable paths for any $H \in (0,1)$. 
5. is a linear process for $d = H = 1$, i.e., $X(t) \overset{d}{=} tX_0$, $t \in \mathbb{R}$ for a random variable $X_0 \sim N(0, 1)$.

Examples of Non-Gaussian Random Fields

1. Lévy Process with Finite Second Moments

Let $X = \{X(t), t \geq 0\}$ be a Lévy process with finite second moments. It is usually defined via the Lévy–Khinchin triplet coding its jump structure, see e.g. [33]. It is clear that $X$ is intrinsic stationary of order two, but not wide sense stationary. For each of these processes one can calculate the variance of increments and the variogram, for example,

$$
\gamma(h) = \frac{1}{2} \cdot \mathbb{E}(X(t+h) - X(t))^2 = \lambda |h|/2, \quad h, t \geq 0
$$

for the stationary Poisson point process with intensity $\lambda > 0$.

2. Poisson Shot Noise Field

A Poisson shot noise field $X = \{X(t), t \in \mathbb{R}^d\}$ is defined by

$$
X(t) = \sum_{x_i \in \Phi} f(t - x_i) = \int_{\mathbb{R}^d} f(t - x) \Phi(dx), \quad t \in \mathbb{R}^d,
$$

where $\Phi$ is a stationary Poisson point process on $\mathbb{R}^d$ with intensity $\lambda$, $f \in L^1(\mathbb{R}^d)$. It follows from [38, Exercise 9.10] that $X$ is strictly stationary.

It can be shown that

$$
\mathbb{E}X(t) = \lambda \int_{\mathbb{R}^d} f(x) dx,
$$

and if additionally $f \in L^2(\mathbb{R}^d)$ then

$$
\text{cov}(X(s), X(t)) = \lambda \int_{\mathbb{R}^d} f(t - s + x) f(x) dx,
$$

i.e., the Poisson shot noise field is also wide sense stationary (cf. [38, Exercise 9.29]). If $f$ is rotation invariant then $X$ is isotropic of order two. See Figure 1(b) for a realization of $X$.

3. Boolean Random Function

Let $\{Z(x), x \in \mathbb{R}^d\}_{t \in \mathbb{R}}$ be a family of independent lower semi-continuous random functions with subgraphs having almost surely compact sections and $\Pi = \{(x_i, t_i)\}_{i=1}^m$ be a Poisson point process in $\mathbb{R}^d \times \mathbb{R}$ with intensity measure $\nu_d \otimes \theta$, where
6 Evgeny Spodarev, Elena Shmileva and Stefan Roth

(a) Gaussian random field with Whittle-Matérn–type covariance function (see Sect. 2.2.1, Example 6), \(a = 2, b = \nu = 1\)

(b) Poisson shot noise field with \(\lambda = 1\) and

\[
f(x) = \frac{1}{2\pi} \left(1 - \frac{1}{4} \|x\|^2 \right) \mathbf{1}(\|x\| \leq 2)
\]

Fig. 1 Simulated realizations of (strictly and wide sense) motion invariant random fields.

where \(\nu_d\) denotes the Lebesgue measure on \(\mathbb{R}^d\) and \(\theta\) is a \(\sigma\)-finite measure on \(\mathbb{R}\).

The random function

\[
Z(x) = \sup_{(x_k, x_k) \in \Pi} Z_t(x - x_k), \quad x \in \mathbb{R}^d
\]

is called a Boolean random function. The functions \(Z_t\) are referred to as primary functions. Boolean random functions have been introduced by D. Jeulin for modelling rough morphologies ([13]), see for example [4, Sect. 7.8.1] and references therein.

2.2 Elements of Correlation Theory for Square Integrable Random Fields

Let us recall the following basic concepts.

**Definition 3.** A symmetric function \(f: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}\) is called positive semi–definite if for any \(n \in \mathbb{N}\), \(w_1, \ldots, w_n \in \mathbb{C}\) and any \(t_1, \ldots, t_n \in \mathbb{R}^d\) it holds

\[
\sum_{i,j=1}^{n} w_i \bar{w}_j f(t_i, t_j) \geq 0.
\]

**Definition 4.** A symmetric function \(f: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}\) is called positive definite if for any \(n \in \mathbb{N}\), \(w_1, \ldots, w_n \in \mathbb{C}\) such that \((w_1, \ldots, w_n)^\top \neq 0 \in \mathbb{C}^n\) and any \(t_1, \ldots, t_n \in \mathbb{R}^d\) it holds

\[
\sum_{i,j=1}^{n} w_i \bar{w}_j f(t_i, t_j) > 0.
\]
**Definition 5.** A symmetric function \( f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) is called **conditionally negative semi–definite** if for any \( n \in \mathbb{N} \), \( w_1, \ldots, w_n \in \mathbb{C} \) such that \( \sum_{i=1}^n w_i = 0 \) and any \( t_1, \ldots, t_n \in \mathbb{R}^d \) it holds
\[
\sum_{i,j=1}^n w_iw_j f(t_i, t_j) \leq 0.
\]

**Exercise 4.** Prove that functions \( \cos(a \cdot x), \ a \in \mathbb{R}, \ e^{-|x|^p}, \ p \in (0, 2] \) are positive semi–definite, whereas \( e^{-|x|^p}, \ p > 2, \ |\cos x|, a^2 + \cos^2 x, \ a \in \mathbb{R} \) are not.

**Exercise 5.** Find a positive semi–definite function with discrete support.

### 2.2.1 Covariance function

**Definition 6.** For a real-valued random field \( X = \{X(t), \ t \in \mathbb{R}^d\} \) with \( \mathbb{E}(X(t))^2 < \infty \), \( t \in \mathbb{R}^d \), the function \( C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) given by
\[
C(s, t) = \text{cov}(X(s), X(t)) = \mathbb{E}(X(s) - \mathbb{E}(X(s))(X(t) - \mathbb{E}(X(t))), \ s, t \in \mathbb{R}^d
\]
is called the **covariance function**.

If \( X \) is wide sense stationary (motion invariant), then \( C(s, t) \) depends only on \( s - t \) (\( ||s - t|| \), respectively), \( s, t \in \mathbb{R}^d \). For the properties of the covariance function see [38, Sect. 9.4-9.6]. We mention just a few:

1. **Generic property.** A function \( f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) is a covariance function of some square integrable random field iff it is positive semi–definite.

**Exercise 6.** Prove this fact. **Hint:** Calculate the variance of linear combinations \( \sum_{i=1}^n x_i X(t_i) \) for arbitrary \( n \in \mathbb{N}, t_i \in \mathbb{R}^d, x_i \in \mathbb{R} \).

2. **Spectral representation.** By Bochner-Kinchin theorem (see, e.g., [11] or [38, Theorem 9.6]), any continuous at the origin positive semi–definite function \( f : \mathbb{R}^d \to \mathbb{R} \) is a Fourier transform of some symmetric finite measure \( \mu_f \) on \( \mathbb{R}^d \). Thus for a wide sense stationary mean square continuous field \( X \) with covariance function \( C \) we have
\[
\text{cov}(X(s), X(t)) = C(s - t) = \int_{\mathbb{R}^d} e^{i(s \cdot x - t)} \mu_C(dx).
\]

Here \( \langle \cdot, \cdot \rangle \) is the Euclidean scalar product in \( \mathbb{R}^d \). Measure \( \mu_C \) is called a **spectral measure** of \( X \). If \( \mu_C \) is absolutely continuous with respect to the Lebesgue measure, then its density is called a **spectral density**. The above field \( X \) has itself the **spectral representation**
\[
X(t) = \int_{\mathbb{R}^d} e^{i(s \cdot x)} \Lambda(dx),
\]
(1)
where \( \Lambda(\cdot) \) is a complex-valued orthogonal random measure with \( \mathbb{E} \Lambda(A) = 0 \) and \( \mathbb{E} \left( \Lambda(A) \overline{\Lambda(B)} \right) = \mu_C(A \cap B) \) for any Borel sets \( A, B \subset \mathbb{R}^d \). The integral in
is understood in the mean square sense, i.e. its integral sums converge in $L^2(\Omega, \mathcal{F}, P)$. For more details on the spectral representation of stationary processes see [3, Sect. 7, §9, §10], [23, Sect. 3.2, pp. 20-21] or [4, Sect. 2.3.3], [44, Sect. 4.2, p. 90]. The spectral representation is used e.g. to simulate stationary Gaussian random fields approximating the integral in (1) by its finite integral sums with respect to a Gaussian white noise measure $\Lambda$.

**Parametric Families of Covariance Functions**

1. **White Noise Model**

   $$ C(s,t) = \begin{cases} \sigma^2, & s = t \\ 0, & s \neq t \end{cases}, \quad s,t \in \mathbb{R}^d. $$

   It is a covariance function of a random field $X$ consisting of independent random variables $X(t), t \in \mathbb{R}^d, d \geq 1$ with variance $\sigma^2 > 0$.

2. **Normal Scale Mixture**

   $$ C(s,t) = \int_0^\infty e^{-s|s-t|^2} \mu(dx), \quad s,t \in \mathbb{R}^d $$

   for some finite measure $\mu$ on $[0, \infty)$ is the covariance function of a motion invariant random field for any $d \geq 1$ (see [33]).

3. **Bessel Family**

   $$ C(s,t) = b(a|s-t|)^{-\nu} J_\nu(a|s-t|), \quad \nu = \frac{d-2}{2}, \quad a,b > 0, \quad s,t \in \mathbb{R}^d, $$

   where

   $$ J_\nu(r) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j! \Gamma(v+j+1)} \left( \frac{r}{2} \right)^{v+2j}, \quad r \in \mathbb{R}. $$

   is the Bessel function of the 1st kind of order $\nu$ (cf. [24]) and $d \geq 1$. The positive semi–definiteness of $C$ is proven in [46, p. 367]. The spectral density of $C$ is given by

   $$ f(h) = \frac{b(a^2-h^2)^{\nu-\frac{d}{2}}}{2^\nu \pi^{\frac{\nu}{2}} a^{2\nu} \Gamma(\nu+1-\frac{d}{2})} I(h \in [0,a]). $$

   A special case of $d = 3$, i.e., $\nu = \frac{1}{2}$ yields the so-called hole effect model.
Extrapolation of Stationary Random Fields

\[ C(s, t) = b \frac{\sin(a \|s - t\|)}{a \|s - t\|}, \quad s, t \in \mathbb{R}^d. \]

This model is valid only for \( d \leq 3. \)

4. Cauchy Family

\[ C(s, t) = \frac{b}{1 + (a \|s - t\|)^2}, \quad a, b, \nu > 0, \quad s, t \in \mathbb{R}^d. \]

Up to scaling, this function is positive semi-definite as a normal scale mixture with \( \mu(dx) = c \nu^{-1}e^{-x}dx \) for some constant \( c > 0. \)

5. Stable Family

\[ C(s, t) = be^{-a\|s - t\|^\nu}, \quad \nu \in (0, 2], \quad s, t \in \mathbb{R}^d. \]

This function is positive semi-definite for all \( d \geq 1 \) since it is made by substitution \( \theta \mapsto \|s - t\| \) out of the characteristic function of a symmetric \( \nu \)-stable random variable, cf. Definition [11]. A special case \((\nu = 2)\) of the stable family is a Gaussian model: \( C(s, t) = be^{-a\|s - t\|^2} \). Its spectral density is equal to \( f(h) = \frac{b \sqrt{\pi}}{\nu} he^{-\frac{ah^2}{2}}. \)

6. Whittle-Matérn Family

\[ C(s, t) = W_\nu(\|s - t\|) = b2^{1-\nu}(a\|s - t\|)^\nu K_\nu(a\|s - t\|), \quad s, t \in \mathbb{R}^d, \quad s \neq t, \]

where \( \nu, a, b > 0, \ d \geq 1 \) and \( K_\nu \) is the modified Bessel function of third kind, also called Macdonald function:

\[ K_\nu(r) = \frac{\pi}{2 \sin(\pi \nu)} (e^{i \frac{\nu}{2}} J_{-\nu}(re^{i \frac{\nu}{2}}) - e^{-i \frac{\nu}{2}} J_{\nu}(re^{-i \frac{\nu}{2}})), \quad r \in \mathbb{R}, \quad \nu \notin \mathbb{N}. \]

For \( \nu = n \in \mathbb{N} \) the above definition of \( K_\nu \) is understood in the sense of a limit as \( \nu \to n \), see [24] p. 69]. For \( s = t \), we set \( C(t, t) = b \). The spectral density of \( C \) is given by

\[ f(h) = \frac{2 \Gamma(\nu + \frac{d}{2})}{\Gamma(\frac{d}{2} \nu) (1 + (ah)^2)^{\frac{\nu + d}{2}}} I(h > 0). \]

If \( \nu = \frac{2d + 1}{2} \) then a random field with covariance function \( C \) is \( d \) times differentiable in mean-square sense. If \( \nu = \frac{1}{2} \) then the exponential model

\[ C(s, t) = be^{-a\|s - t\|}, \quad s, t \in \mathbb{R}^d \]
is an important special case. The same exponential covariance belongs to the stable family for \( \nu = 1 \).

Figure 1(a) shows a realization of a centered Gaussian random field \( X \) with Whittle-Matérn type covariance function.

7. Spherical Model

is given for \( 1 \leq d \leq 3 \) by

\[
C(s,t) = b \left( 1 - \frac{3}{2} \frac{\|s-t\|}{a} + \frac{1}{2} \frac{\|s-t\|^3}{a^3} \right) I(\|s-t\| \leq a), \quad a, b > 0, \quad s, t \in \mathbb{R}^d.
\]

If \( d = 3 \) the above formula yields the volume of \( B_2(0) \cap B_2(x_0) \), where \( x_0 \in \mathbb{R}^3 \), \( \|x_0\| = \|s-t\| \). This is exactly the way how it can be generalized to higher dimensions:

\[
C(s,t) = \nu_d \left( B_2(0) \cap B_2(s-t) \right), \quad s, t \in \mathbb{R}^d,
\]

where \( \nu_d \) is the Lebesgue measure. The advantage of spherical models is that they have a compact support.

8. Geometric Anisotropy

It is easy to see that all covariance models considered above are motion invariant. An example of an anisotropic covariance structure can be provided by rotating and stretching the argument of a motion invariant covariance model. Let \( C_0(\|h\|), h \in \mathbb{R}^d \) be a covariance function of a motion invariant field where \( C_0: \mathbb{R}^+ \rightarrow \mathbb{R}^+ \). For a positive definite \((d \times d)\)-matrix \( Q \),

\[
C(h) = C_0(\sqrt{h^T Q h}), \quad h \in \mathbb{R}^d
\]

is a covariance function of some wide sense stationary anisotropic random field (see [42, Chap. 9]).

9. Cyclone Model

For \( d = 3 \), let

\[
C(x,y) = \frac{2^{3/2} \det(S_x)^{1/4} \det(S_y)^{1/4}}{\sqrt{\det(S_x + S_y)}} W_\nu \left( \sqrt{(x-y)^T S_x(S_x + S_y)^{-1} S_y(x-y)} \right),
\]

where \( x, y \in \mathbb{R}^3 \), \( S_x = Id + xx^T \), \( Id \) is a \((3 \times 3)\)-identity matrix and \( W_\nu \) is the Whittle-Matérn model. In [34, Theorem 5, Example 16], it is shown that \( C \) is a valid covari-
ance function belonging to a more general class of covariances that mimic cyclones.

**Exercise 7.** Show $C$ is a covariance function of isotropic but not wide sense stationary random field, i.e., $C(x, y) = C(Rx, Ry)$ for any $R \in SO_3$, but $C(x, y)$ does not depend on $x - y$, $x, y \in \mathbb{R}^3$.

For more sophisticated covariance models including spatio–temporal effects see e.g. [34] and references therein.

### 2.2.2 Variogram

**Definition 7.** For a random field $X = \{X(t), t \in \mathbb{R}^d\}$ the following expression

$$
\gamma(t, s) := \frac{1}{2} \mathbb{E} (X(t) - X(s))^2, \quad s, t \in \mathbb{R}^d
$$

is called a variogram of $X$ whenever it is finite for any $s, t \in \mathbb{R}^d$.

For square integrable random fields $X$, it obviously holds

$$
\gamma(s, t) = \frac{1}{2} \text{var} X(s) + \frac{1}{2} \text{var} X(t) - \text{cov}(X(t), X(s)) + \frac{1}{2} (\mathbb{E} X(s) - \mathbb{E} X(t))^2. \quad (2)
$$

If the field $X$ is intrinsic stationary of order two (motion invariant) then $\gamma(s, t)$ depends only on the difference $s - t$ ($\|s - t\|$, respectively). With slight abuse of notation in these cases, we write $\gamma(s - t)$ and $\gamma(\|s - t\|)$ for functions $\gamma : \mathbb{R}^d \to \mathbb{R}$ and $\gamma : \mathbb{R}_+ \to \mathbb{R}$, respectively. For a wide sense stationary random field $X$ with covariance function $C$ the relation (2) reads

$$
\gamma(h) = C(0) - C(h), \quad h \in \mathbb{R}^d. \quad (3)
$$

### Basic Properties of Variograms

Let $X$ be a random field with covariance function $C$ and variogram $\gamma$. The following properties hold:

1. $\gamma(t, t) = 0$, $t \in \mathbb{R}^d$.
2. Symmetry: $\gamma(t, s) = \gamma(s, t)$, $s, t \in \mathbb{R}^d$.
3. Characterization of variograms:
   
   (a) A function $\gamma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ is a variogram of some random field if $\gamma$ is conditionally negative semi–definite, see, for example, [10] Theorem 1] or [4] Sect. 2.3.3, p.61.

**Exercise 8.** Prove that the variogram of any intrinsic stationary random field $X$ is a conditionally negative semi–definite function.

**Hint:** Calculate $\text{Var} (\sum_{i=1}^n \lambda_i X(t_i))$ applying (2) with $\sum_{i=1}^n \lambda_i = 0$. 

(b) A continuous even function $\gamma : \mathbb{R}^d \to \mathbb{R}^+$ with $\gamma(0) = 0$ is a variogram of a wide sense stationary random field if $e^{-\lambda \gamma}$ is a covariance function for all $\lambda > 0$, cf. [36].

4. Stability: If $\gamma_1, \gamma_2$ are variograms then $\gamma = \gamma_1 + \gamma_2$ is a variogram as well.

**Exercise 9.** Prove this fact. Show in particular that $\gamma(h) = \gamma_1(h) + \gamma_2(h)$, where $h = (h_1, \ldots, h_d)^\top \in \mathbb{R}^d$ and $\gamma_1, \gamma_2$ are univariate variograms, is a variogram.

5. Mixture: Let $\gamma_x : \mathbb{R}^d \to \mathbb{R}^+$ be a variogram of an intrinsic stationary (of order two) random field for any $x \in \mathbb{R}$. Then the function

$$\gamma(t) = \int_{\mathbb{R}} \gamma_x(t) \mu(dx), \quad t \in \mathbb{R}^d$$

is a variogram of some random field if $\mu$ is a measure on $\mathbb{R}$ and the above integral exists for any $t \in \mathbb{R}^d$, see [4] Sect. 2.3.2, pp. 60-61).

6. If $X$ is wide sense stationary and $C(\infty) := \lim_{\|h\| \to \infty} C(h) = 0$, then it follows from (3) that there exists the so-called sill $\gamma(\infty) := \lim_{\|h\| \to \infty} \gamma(h) = C(0)$.

7. If $X$ is mean square continuous then $\gamma(h) \leq c\|h\|^2$, $h \in \mathbb{R}^d$ for a constant $c > 0$ and large $\|h\|$, see [46] pp. 397-398).

8. If $X$ is mean square differentiable then $\lim_{\|h\| \to \infty} \frac{\gamma(h)}{\|h\|^2} = 0$, see [47] pp. 136-137].

9. Let $\gamma : \mathbb{R} \to \mathbb{R}^+$ be an even twice continuously differentiable function with $\gamma(0) = 0$. Then $\gamma$ is a variogram iff $\gamma''$ is a covariation function, cf. [10] Theorem 7).

**Exercise 10.** Show that for a variogram $\gamma$ the function $e^{\lambda \gamma}$ is a variogram for any $\lambda > 0$.

**Exercise 11.** Let a bounded function $\gamma : \mathbb{R}^d \to \mathbb{R}^+$ be the variogram of some intrinsic stationary of order two real valued random field $X$. Consider $C(x, y) = \gamma(x) + \gamma(y) - \gamma(x - y)$, $x, y \in \mathbb{R}^d$. Show that $C$ is a covariance function of a random field $Z$ such that $Z(0) = 0$ a.s.

**Parametric families of variograms**

Most parametric models for variograms of stationary random fields, which are widely used in applications, can be constructed from the corresponding families of covariance functions (such as those described in Sect. 2.2.1) by applying the relation (3) as well as stability and geometric anisotropy properties. Most models of variograms inherit their names from the corresponding covariance models (e.g., exponential, spherical one). One of few exceptions is the variogram corresponding to the white noise which is called nugget effect.

Stability property can be also used to create different anisotropy effects, for instance, the so-called purely zonal anisotropy. To explain this on an example,
let $\gamma(h) = a\gamma_1(h_x) + b\gamma_2(h_y) + c\gamma_3(h_z)$, $h = (h_x, h_y, h_z) \in \mathbb{R}^3$, $a, b, c \geq 0$, where $\gamma_i \quad i = 1, 2, 3$ are variograms in dimension $d = 1$. Then $\gamma$ is a variogram in dimension $d = 3$ which allows for different dependence ranges in three different axes directions. An example of mixed anisotropy models is

$$\gamma(h) = \gamma_1(||h||) + \gamma_2\left(\sqrt{h_x^2 + h_z^2}\right) + \gamma_3(h_z), \quad h = (h_x, h_y, h_z) \in \mathbb{R}^3.$$  

This is a mixture of 3D-isotropic variogram $\gamma_1$, 2D-isotropic (in the xy-plane) variogram $\gamma_2$ and a 1D-variogram $\gamma_3$. Addition of a linear combination of $\gamma_2$ and $\gamma_3$ creates anisotropy in direction of z-axis.

See more about variograms in [4, Chap. 2].

2.2.3 Statistical Estimation of Covariances and Variograms

The numerous approaches to estimate a covariance function or a variogram are well described in the literature and therefore will not be reviewed here. An interested reader can see e.g. [4, Sect. 2.2] and [38, Sect. 9.8] and references therein.

Example 1. To illustrate the above theory, consider microscopic steel data (figure 2(a)). This data is obviously isotropic. Figure 2(b) shows estimates for the corresponding variogram. For this purpose Mathéron’s estimator (see [38, p. 325]) was calculated for different directions and $0 \leq h \leq 0.5$. The directions can be distinguished by the color of their plots. Since these data are isotropic the estimates differ not too much.

![Microscopic image of a steel surface](image1)

![Estimates for the x-direction (red), y-direction (green), all directions (black) for values $0 \leq h \leq 0.5$.](image2)

Fig. 2 Microscopic steel image (left) and its empirical variogram estimated in different directions (right)
Example 2. Let us construct an example of zonally anisotropic variogram, in which the value for the sill depends on the direction of the input vector \( h \). Consider
\[
\gamma(h) = \gamma_1(h) + \gamma_2(h)
\]
where \( \gamma_1 \) is an isotropic variogram
\[
\gamma_1(h) = 1 - e^{-|h|}, \quad h \in \mathbb{R}^2
\]
and \( \gamma_2 \) is a geometrical anisotropic variogram model
\[
\gamma_2(h) = 1 - e^{-\frac{\sqrt{h^T Q h}}{s}}, \quad h \in \mathbb{R}^2
\]
with \( Q = \sqrt{\Lambda} \cdot R \) with \( R \) being a rotation matrix with rotation angle \( \alpha = 2 \) and \( \Lambda = \text{diag}(5, 1) \) being a diagonal matrix. Figure 3(a) shows \( \gamma \) on \([-1, 1]^2\). Figure 3(b) illustrates the elliptic form of the contour lines of a zonally anisotropic variogram.

![Zonally anisotropic model with rotation angle \( \alpha = 2 \) and scaling factors \( \lambda_1 = 5, \lambda_2 = 1 \).](image)

![Contour lines of 3(a)](image)

Fig. 3 Zonally anisotropic theoretical variogram

### 2.3 Stable Random Fields

In this Section, we review the basic notions of the theory of stable distributions, random measures and fields. A very good reference which covers most of this topic is [32], see also [29], [33] Chapter 3, [48], etc.
2.3.1 Stable Distributions

Let \( n \in \mathbb{N} \). We begin with the definition of stability for random vectors.

**Stable Random Vectors**

**Definition 8.** A random vector \( X = (X_1, \ldots, X_n)^T \) in \( \mathbb{R}^n \) is called *stable* if for all \( m \geq 2 \) there exist \( c = c(m) > 0 \) and \( k = k(m) \in \mathbb{R}^n \) such that

\[
    X^{(1)} + X^{(2)} + \cdots + X^{(m)} \overset{d}{=} cX + k,
\]

where \( \{X^{(i)}\}_{i=1}^m \) are independent copies of \( X \).

It can be shown that \( c = m^{1/\alpha} \) for some \( 0 < \alpha \leq 2 \) which is called the *stability index*, see [32, Theorem 2.1.2]. There is an equivalent definition of stable vectors which is often used in mathematical practice to check stability.

**Definition 9.** Let \( \alpha \in (0, 2) \). We say that a random vector \( X = (X_1, \ldots, X_n)^T \) in \( \mathbb{R}^n \) is *\( \alpha \)-stable* if its characteristic function is given by

\[
    \phi_X(\theta) = \begin{cases} 
        e^{-\frac{1}{\alpha}|\theta|^\alpha (1 - \text{sign}(|\theta|) \tan \frac{\pi}{\alpha} \Gamma(d\theta)) + i(\theta, \mu)}, & \alpha \neq 1, \\
        e^{-\frac{1}{2}|\theta|^\alpha (\ln |\theta|) \Gamma(d\theta) + i(\theta, \mu)}, & \alpha = 1,
    \end{cases}
\]

(4)

where \( \Gamma \) is a finite measure on the unit sphere \( \mathbb{S}^{n-1} \) of \( \mathbb{R}^n \) and \( \mu \) is an arbitrary vector in \( \mathbb{R}^n \).

The pair \( (\mu, \Gamma) \) gives a unique parametrization of the distribution of \( \alpha \)-stable random vectors for \( \alpha \in (0, 2) \), and we write \( X \sim S_\alpha(\mu, \Gamma) \). This means that there is no other pair \( (\mu', \Gamma') \) yielding the same characteristic function \( \phi_X \) in (4). The measure \( \Gamma \) is called the *spectral measure of \( X \)* and contains all the information about the dependence between the vector components \( X_i \) (see also Exercise 15). The vector \( \mu \) reflects the *shift* with respect to the origin.

**Definition 10.** A random vector \( X = (X_1, \ldots, X_n)^T \) is called *singular* if \( \sum_{i=1}^n c_i X_i = 0 \) a.s. for some \( (c_1, \ldots, c_n)^T \in \mathbb{R}^n \setminus \{0\} \). Otherwise, it is called *full-dimensional*.

If \( \alpha = 2 \), then Definition 9 yields a *Gaussian* random vector which is equivalently defined via its characteristic function

\[
    \phi_X(\theta) = \exp \left\{ i(\theta, \mu) - \frac{1}{2} \theta^T \Sigma \theta \right\}.
\]

(5)

Here \( \mu \in \mathbb{R}^n \) is the *mean* of \( X \) and \( \Sigma \) is a symmetric, positive semi–definite \( (n \times n) \)–*covariance matrix* of \( X \). The matrix \( \Sigma \) has the elements \( \sigma_{ij} = \mathbb{E}(X_i - \mu_i)(X_j - \mu_j) \), where \( X_i \) and \( \mu_i \) are the components of vectors \( X \) and \( \mu \), respectively. It is easy to see that if \( \det \Sigma = 0 \) then the Gaussian random vector \( X \) is singular.
Exercise 12. Prove that Definition 8 is equivalent to Definition 9 for \( \alpha \in (0, 2) \), and it is equivalent to the definition of a Gaussian random vector via relation (5) for \( \alpha = 2 \).

Exercise 13. Show that for \( X \sim S_\alpha(\mu, \Gamma) \) the relation between the drift \( k \) in Definition 8 and the shift \( \mu \) in Definition 9 is
\[
k(m) = \mu \left( m - m^{1/\alpha} \right).
\]
Hint: First show that
\[
\Sigma_{i=1}^m X^{(i)} \sim S(m\mu, m\Gamma) \quad \text{and} \quad m^{1/\alpha}X + k(m) \sim S(m^{1/\alpha}\mu + k(m), m\Gamma).
\]

Remark 1. For \( \alpha = 2 \), the characteristic function (4) has the form
\[
\varphi(\theta) = \exp \left\{ - \int_{S^{n-1}} (\theta, s)^2 \Gamma(ds) + i(\theta, \mu) \right\}.
\] (6)

Exercise 14. Check that the following two finite measures on the unite sphere in \( \mathbb{R}^2 \)
\[
\Gamma_1(ds) = \delta_{(\sqrt{2}/2, \sqrt{2}/2)}(ds) + \delta_{(-\sqrt{2}/2, -\sqrt{2}/2)}(ds),
\]
\[
\Gamma_2(ds) = 2\delta_{(\sqrt{2}/2, \sqrt{2}/2)}(ds)
\]
and a shift \( \mu \in \mathbb{R}^2 \) yield the same expression in (4) if \( \alpha = 2, n = 2 \). Here \( \delta_x(\cdot) \) is the Dirac measure concentrated at the point \( x \in \mathbb{R}^2 \). Verify that this expression corresponds to the characteristic function of the Gaussian vector with shift \( \mu \) and covariance matrix
\[
\Sigma = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}.
\]

A random vector \( X \) in \( \mathbb{R}^n \) is called symmetric if \( \mathbb{P}(X \in A) = \mathbb{P}(-X \in A) \) for any Borel set \( A \in \mathbb{R}^n \). For symmetric \( \alpha \)-stable distributions, we use the standard abbreviation \( \text{S}_{\alpha}S \).

Lemma 1 ([32], Theorem 2.4.3). An \( \alpha \)-stable random vector \( X \) is symmetric iff its shift \( \mu = 0 \) and spectral measure \( \Gamma \) is symmetric.

Exercise 15. Let \( X = (X_1, X_2)^T \) be an \( \alpha \)-stable random vector, \( \alpha \in (0, 2) \), with the spectral measure \( \Gamma \). Let \( \text{supp}(\Gamma) \) be the support of \( \Gamma \). Show that

- \( X_1 \) is independent of \( X_2 \) iff \( \text{supp}(\Gamma) \) lies within the intersection of the sphere with the coordinate axes.
- \( X_1 = c \cdot X_2 \) a.s. for some \( c \in \mathbb{R} \) (i.e. the vector \( X \) is singular) iff \( \text{supp}(\Gamma) \) is a subset of the unite sphere intersected by a hyperplane.

Stable Random Variables

If \( n = 1 \) we deal with stable random variables whose distribution laws are defined by four parameters \( \alpha, \sigma, \beta, \) and \( \mu \).
Definition 11. The random variable $X$ is called $\alpha$-stable if its characteristic function has the form
\[
\varphi_X(\theta) = \begin{cases} 
\exp \left\{ -\sigma^\alpha |\theta|^\alpha \left( 1 - i\beta \text{sign}(\theta) \tan \frac{\pi \alpha}{2} \right) + i\mu \theta \right\}, & \alpha \in (0, 2], \alpha \neq 1, \\
\exp \left\{ -\sigma |\theta| \left( 1 + i\beta \frac{2}{\pi} \text{sign}(\theta) \ln |\theta| \right) + i\mu \theta \right\}, & \alpha = 1.
\end{cases}
\]

We write $X \sim S_\alpha(\sigma, \beta, \mu)$. Compared with representation (4), two new parameters $\sigma \geq 0$ and $\beta \in [-1, 1]$ introduced in lieu of the spectral measure $\Gamma$ are interpreted as parameters of scale and skewness, respectively.

Exercise 16. Show that the spectral measure of $X \sim S_\alpha(\sigma, \beta, \mu)$ is given by
\[
\Gamma(ds) = \frac{\sigma^\alpha}{2} (1 + \beta) \delta_1(ds) + \frac{\sigma^\alpha}{2} (1 - \beta) \delta_{-1}(ds).
\]

Hence, it holds
\[
\sigma^\alpha = \frac{\Gamma(\{1\}) - \Gamma(\{-1\})}{\Gamma(\{1\}) + \Gamma(\{-1\})}, \quad \beta = \frac{\Gamma(\{1\}) - \Gamma(\{-1\})}{\Gamma(\{1\}) + \Gamma(\{-1\})}.
\]

Remark 2. Stable distributions are absolutely continuous. Nevertheless, their densities are not known in the closed form except for the cases $\alpha = 1/2$, $\alpha = 1$ and $\alpha = 2$.

Example 3.
1. $X \sim S_2(\sigma, 0, \mu)$ is a Gaussian random variable with mean $\mu$ and variance $2\sigma^2$.
2. Random variable $X \sim S_\alpha(\sigma, \pm 1, \mu)$ is called totally skewed. Notice that if $\alpha \in [1, 2)$ then $X$ attains values in the whole $\mathbb{R}$. On the contrary, if $\alpha \in (0, 1)$ and $\mu = 0$, then $X \geq 0$, $X \leq 0$ a.s. when $\beta = 1$ ($\beta = -1$), respectively.

Exercise 17. Show that the characteristic function of $S\alpha S$ random variable $X$ is equal to $\varphi_X(\theta) = \exp \{-\sigma^\alpha |\theta|^\alpha\}$, i.e., $X \sim S_\alpha(\sigma, 0, 0)$ for some $\sigma > 0$.

Tails and Moments

The non–Gaussian stable distributions are fat tailed. This means that they belong to a subclass of heavy tailed distributions with especially slow large deviation behavior, see more details on heavy tailed distributions e.g. in [7], [25], etc. Namely, for $X \sim S_\alpha(\sigma, \beta, \mu)$ with $\alpha \in (0, 2)$ there exists $c > 0$ such that
\[
P(|X| > x) \sim cx^{-\alpha}, x \to \infty.
\]

Here and in what follows we say that $a_\alpha \sim b_\alpha$ if $\lim_{x \to \infty} \frac{a_\alpha}{b_\alpha} = 1$. As a corollary of [7], the absolute moments of $X$ behave like
They are finite if $p \in (0, \alpha)$ and infinite for any $p \in [\alpha, \infty)$.

**Exercise 18.** Show that

1. normal distribution $X \sim N(\mu, \sigma^2)$ is not heavy tailed (this is equivalent to the statement that the tails are exponentially bounded), i.e.,
   \[ P(X < -x) = P(X > x) \approx \frac{1}{\sqrt{2\pi} \sigma x} e^{-x^2/(2\sigma^2)}, \quad x \to \infty. \]

2. for $X \sim S_\alpha(\sigma, \beta, 0), \alpha \in (0, 2), \alpha \neq 1$ it holds
   \[ (E|X|^p)^{1/p} = c_{\alpha, \beta}(p) \sigma \]
   for every $p \in (0, \alpha)$. Here $c_{\alpha, \beta}(p) = (E|\xi|^p)^{1/p}$ with $\xi \sim S_\alpha(1, \beta, 0)$. If $\alpha = 1$ then equation (8) holds only for $\beta = 0$.

3. for any $\alpha$-stable random variables $X$ and $Y$ the sum $aX + bY$, $a, b \in \mathbb{R}$ is again $\alpha$-stable. Moreover, components $X_i$ of the stable vector $X = (X_1, X_2) \sim S_\alpha(\mu, \Gamma)$ are stable, and it holds $\sigma_{aX_1 + bX_2} = \int_0^\infty |as_1 + bs_2|^\alpha |\Gamma| (ds_1, ds_2)$ for any $a, b \in \mathbb{R}$.

### 2.3.2 Integration with Respect to Stable Random Measures

Let $(E, \mathcal{E}, m)$ be an arbitrary measurable space with $\sigma$-finite measure $m$ and $\mathcal{E}_0 := \{A \in \mathcal{E} : m(A) < \infty\}$. Let $\beta : E \to [-1, 1]$ be a measurable function.

**Definition 12.** A random function $M = \{M(A), A \in \mathcal{E}_0\}$ is called an **independently scattered random measure** (random noise) if

1. for any $n \in \mathbb{N}$ and pairwise disjoint sets $A_1, A_2, \ldots, A_n \in \mathcal{E}_0$ random variables $M(A_1), \ldots, M(A_n)$ are independent,
2. $M(\bigcup_{j=1}^n A_j) = \sum_{j=1}^n M(A_j)$ a.s. for a sequence of disjoint sets $A_1, A_2, \ldots \in \mathcal{E}_0$ with $\bigcup_{j=1}^n A_j \in \mathcal{E}_0$.

**Definition 13.** An independently scattered random measure $M$ on $(E, \mathcal{E}_0)$ is called **$\alpha$-stable** if for each $A \in \mathcal{E}_0$

\[ M(A) \sim S_\alpha \left( (m(A))^{1/\alpha}, \int_A \beta(x)m(dx) \frac{m(A)}{m(A)}, 0 \right). \]

Measure $m$ is called **control measure**, and $\beta$ is the **skewness function** of $M$.

Our goal is to define an integral $\int_E f(x)M(dx)$ of a deterministic function $f : E \to \mathbb{R}$ with respect to an $\alpha$-stable random measure $M$. For a simple function $f(x) = \sum_{i=1}^n c_i 1_{A_i}(x)$, where $\{A_i\}_{i=1}^n \subset \mathcal{E}_0$ are pairwise disjoint, we set
\[
\int_E f(x) M(dx) = \sum_{i=1}^n c_i M(A_i).
\]

It can be shown that, so defined, the integral \(\int_E f(x) M(dx)\) does not depend on the representation of \(f\) as a simple function, see [32, Sect.3.4]. For an arbitrary \(f : E \to \mathbb{R}\) such that \(\int_E |f(x)|^\alpha m(dx) < \infty\) consider a pointwise approximation of \(f\) by simple functions \(f^{(n)}\). Then we set

\[
\int_E f(x) M(dx) = \text{plim}_{n \to \infty} \int_E f^{(n)}(x) M(dx).
\]

Here \(\text{plim}\) denotes the limit in probability. This definition is independent of the choice of the approximating sequence \(\{f^{(n)}\}\), cf. [32, Sect. 3.4] for more details.

**Lemma 2.** Let \(X = \int_E f(x) M(dx)\), where \(M\) is an \(\alpha\)-stable random measure with control measure \(m\) and skewness function \(\beta\). Then \(X\) is an \(\alpha\)-stable random variable with zero shift, scale parameter

\[
\sigma^\alpha_X = \int_E |f(x)|^\alpha m(dx),
\]

and skewness parameter

\[
\beta_X = \frac{\int_E f(x)^{<\alpha>} \beta(x) m(dx)}{\int_E |f(x)|^\alpha m(dx)},
\]

where \(a^{<p>} = \text{sign}(a) \cdot |a|^p\).

For the proof see [32, Sect.3.4]. Notice that if \(\beta(x) = 0\) for all \(x \in E\) then the integral \(X\) is a S\&S random variable.

In case of stable vectors with an integral representation, we have the following criterion of their full–dimensionality / singularity.

**Lemma 3.** Consider a \(n\)-dimensional \(\alpha\)-stable random vector \(X = (X_1, \ldots, X_n)^T\) with \(0 < \alpha \leq 2\) and integral representation

\[
X = \left(\int_E f_1(x) M(dx), \ldots, \int_E f_n(x) M(dx)\right)^T.
\]

Then \(X\) is singular if and only if \(\sum_{i=1}^n c_i f_i(x) = 0\) \(m\)-almost everywhere for some vector \((c_1, \ldots, c_n)^T \in \mathbb{R}^n \setminus \{0\}\).

The proof of Lemma 3 follows from Definition 10 and the fact that \(\sigma^\alpha_X = \int_E |\sum_{i=1}^n c_i X_i|^{\alpha} m(dx)\), see relation (9).

**Remark 3.** A more universal criterion of singularity for stable random vectors is in terms of their spectral measure. If measure \(\Gamma(ds)\) on \(\mathbb{S}^{n-1}\) is a spectral measure of an \(\alpha\)-stable vector \(X\) in \(\mathbb{R}^n\) and is concentrated on the intersection of \(\mathbb{S}^{n-1}\) with a \((n - 1)\)-dimensional linear subspace, then the random vector \(X\) is singular. Otherwise, \(X\) is full–dimensional. For the proof see [15].
2.3.3 Stable Random Fields with an Integral Spectral Representation

Definition 14. A random field $X$ is called $\alpha$-stable if all its finite-dimensional distributions are $\alpha$-stable.

Consider random fields $X = \{X(t), t \in \mathbb{R}^d\}$ of the form

$$X(t) = \int_E f_t(x) M(dx), \quad t \in \mathbb{R}^d,$$

(10)

where $f_t : E \to \mathbb{R}$ are measurable functions such that $\int_E |f_t(x)|^\alpha m(dx) < \infty$ and in the case $\alpha = 1$ additionally $\int_E |f(x)| \beta |\ln |f(x)|| m(dx) < \infty$ for any $t \in \mathbb{R}^d$. Here $M$ is an $\alpha$-stable random measure on $(E, \mathcal{E}_0)$ with control measure $m$ and skewness function $\beta$. Obviously, the marginals of the random field $X$ in (10) are $\alpha$-stable. If $\beta(x) = 0$ for all $x \in E$ then all finite-dimensional distributions of $X$ are symmetric $\alpha$-stable, so we call $X$ to be a $S\alpha S$ random field.

A natural question is which stable fields allow for an integral representation (10). A necessary and sufficient condition for this is the condition of separability of $X$ in probability, see [32, Theorem 13.2.1].

Definition 15. A stable random field $X = \{X(t), t \in M\}, M \subseteq \mathbb{R}^d$ is separable in probability if there exists a countable subset $M_0 \subseteq M$ such that for every $t \in M$ and any sequence $\{t_k\}_{k \in \mathbb{N}} \subset M_0$ with $t_k \to t$ as $k \to \infty$ it holds $X(t) = \plim_{k \to \infty} X(t_k)$.

In particular, all stochastically continuous $\alpha$-stable random fields are separable in probability.

2.4 Dependence Measures for Stable Random Fields

The dependence of two $\alpha$-stable random variables cannot be digitized by using the covariance because of the absence of the second moments if $\alpha < 2$. We consider two different ways of measuring the degree of dependence of two stable random variables.

Covariation

Definition 16. Let $X = (X_1, X_2)^\top$ be an $\alpha$-stable random vector with $\alpha \in (1, 2]$ and spectral measure $\Gamma$. The covariation of $X_1$ on $X_2$ is the real number

$$[X_1, X_2]_\alpha = \int_{\mathbb{S}^1} s_1 s_2^{\alpha-1} \Gamma(ds_1, ds_2).$$

It has the following properties.

Theorem 1 (Properties of Covariation). Let $(X_1, X_2, X_3)^\top$ be an $\alpha$-stable random vector with $\alpha \in (1, 2]$. 
1. Linearity in the first entry: for \( a, b \in \mathbb{R} \) it holds
\[
[aX_1 + bX_2, X_3]_\alpha = a[X_1, X_3]_\alpha + b[X_2, X_3]_\alpha.
\]

2. If \( X_1 \) and \( X_2 \) are independent then \( [X_1, X_2]_\alpha = 0 \).

3. Gaussian case: for \( \alpha = 2 \), it holds \( [X_1, X_2]_\alpha = 0/2 \cdot \text{cov}(X_1, X_2) \).

4. Covariation and mixed moments: Let \( 1 < \alpha < 2 \) and \( \Gamma \) be spectral measure of \( (X_1, X_2)^\top \) with \( X_1 \sim S_\alpha(\sigma_1, \beta_1, 0) \), \( X_2 \sim S_\alpha(\sigma_2, \beta_2, 0) \). For \( 1 \leq p < \alpha \), it holds
\[
\mathbb{E}\left(X_1X_2^{p-1}\right) = \mathbb{E}[X_1, X_2]_\alpha \left( 1 - c \cdot \beta_2 \right) + c \cdot (X_1, X_2)_\alpha \sigma_2^2,
\]
where \( (X_1, X_2)_\alpha = \int_{\mathbb{R}^2} |s_1 s_2|^{\alpha-1} \Gamma(ds) \) and
\[
c = \frac{\tan(\alpha \pi/2)}{1 + \beta_2^2 \tan^2(\alpha \pi/2)} \left[ \beta_2 \tan(\alpha \pi/2) - \tan\left( \frac{p}{\alpha} \arctan(\beta_2 \tan(\alpha \pi/2)) \right) \right].
\]

**Proof.** 1. Linearity in the first argument is obvious. However, the covariation is not symmetric, so that there is no linearity in the second argument.

2. To see this, use Exercise [15]

3. The assertion (together with the useful relation \( \var(X_i) = 2 \int_{\mathbb{R}} s_i^2 \Gamma(ds_1, ds_2), i = 1, 2 \)) follows from the comparison of the characteristic function \( \phi(\theta) \) of the Gaussian random vector \( (X_1, X_2)^\top \) in representations (5) and (6).

4. See [16] for the proof.

**Remark 4.** If \( X_2 \) is symmetric, i.e. \( \beta_2 = 0 \), then \( c = 0 \) and formula (11) has the following simple form
\[
\mathbb{E}\left(X_1X_2^{p-1}\right) = \mathbb{E}[X_1, X_2]_\alpha \sigma_2^2,
\]
which allows for the estimation of \( [X_1, X_2]_\alpha \) via empirical mixed moments of \( X_1 \) and \( X_2 \).

For a stable random field \( X \) with integral representation (10), the covariation can be calculated by the formula
\[
[X(t_1), X(t_2)]_\alpha = \int_{E} f_{t_1}(x)(f_{t_2}(x))^{\alpha-1} \, dm(x).
\]
Notice that its proof given in [32, Proposition 3.5.2] for the S\( \alpha \)S case holds true for skewed random fields as well.
Codifference

Drawbacks of the covariation are the lack of symmetry and the impossibility to define it for $\alpha \in (0, 1]$. The following measure of dependence does not have these drawbacks. That is however compensated by a mathematically less convenient form.

**Definition 17.** Let $(X_1, X_2)^\top$ be an $\alpha$-stable vector. The *codifference* of $X_1$ and $X_2$ is

$$\tau(X_1, X_2) = \sigma_{X_1} + \sigma_{X_2} - \sigma_{X_1-X_2},$$

where $\sigma_Y$ is the scale parameter of a $\alpha$-stable random variable $Y$.

**Theorem 2 (Properties of Codifference).**

1. **Symmetry:** $\tau(X_1, X_2) = \tau(X_2, X_1)$.
2. If $X_1$ and $X_2$ are independent then $\tau(X_1, X_2) = 0$. The inverse statement holds only for $\alpha \in (0, 1)$.
3. **Gaussian case:** for $\alpha = 2$, it holds $\tau(X_1, X_2) = \text{cov}(X_1, X_2)$.
4. Let $(X_1, X_2)$ and $(X'_1, X'_2)$ be $\alpha$-stable vectors such that $\sigma_{X_1} = \sigma_{X_2} = \sigma_{X'_1} = \sigma_{X'_2}$. If $\tau(X_1, X_2) \leq \tau(X'_1, X'_2)$ then for any $c > 0$

$$P\{|X_1 - X_2| > c\} \geq P\{|X'_1 - X'_2| > c\},$$

i.e., the larger the codifference, the greater the dependence.

**Proof.**

1. Symmetry is obvious.
2. Use Exercise 18 to see the first part of the statement. Now let $\tau(X_1, X_2) = 0$. It holds $\sigma_{X_1} + \sigma_{X_2} = \sigma_{X_1-X_2}$ iff

$$\int_{S_1} |s_1|^\alpha \Gamma(ds) + \int_{S_1} |s_2|^\alpha \Gamma(ds) = \int_{S_1} |s_1 - s_2|^\alpha \Gamma(ds).$$

We know however that $|s_1 - s_2|^\alpha = |s_1|^\alpha + |s_2|^\alpha$ iff $\alpha < 1$ and $s_1 s_2 = 0$.
3. It holds $\tau(X_1, X_2) = 1/2(\text{var} X_1 + \text{var} X_2 - \text{var}(X_1 - X_2)) = \text{cov}(X_1, X_2)$.
4. See [32, Property 2.10.6] for the proof.

### 2.5 Examples of Stable Processes and Fields

#### 1. Stable Lévy Process

This is a process defined by $X(t) = M([0, t]), t \in \mathbb{R}_+$ where $M$ is an $\alpha$-stable measure on $\mathbb{R}_+$ with skewness function $\beta$ and Lebesgue control measure multiplied by $\sigma > 0$. $X$ has representation (10) with $f_t(x) = 1(x \in [0, t])$. It obviously holds $X(0) = 0$ a.s. Moreover, $X$ has independent and stationary increments.

Depending on $\beta$ the skewness of the process may vary. So, for $\alpha < 1$ and $\beta \equiv 1$ we obtain a stable Lévy process with non-decreasing sample paths, the so–called
stable subordinator. To see this use one-to-one correspondence between the infinitely divisible distributions and the Lévy processes, thus $X(1) \sim S_\alpha(\sigma, 1, 0)$ corresponds to a Lévy process with the triplet $(0, 0, \int_{[1-\alpha\cos(\pi\alpha/2)]}^{\infty} \frac{ds}{s^{\alpha+1}} \cdot 1(x > 0))$, which has only positive integrable jumps, see also [23] Examples 21.7 and 24.12).

2. Stable Moving Average Random Fields

A stable moving average $X = \{X(t), t \in \mathbb{R}^d\}$ is defined by the formula

$$X(t) = \int_{\mathbb{R}^d} f(t-s)M(ds), \quad t \in \mathbb{R}^d,$$

where $f \in L^\alpha(\mathbb{R}^d)$ is called a kernel function and $M$ is an $\alpha$–stable random measure with Lebesgue control measure. It can be easily seen that $X$ is strictly stationary. See Figures 4(a) and 4(b) for simulated realizations of moving averages in $d = 2$ with the bisquare and the cylindric kernels.

Fig. 4 Continuous (left) and discontinuous (right) realization of a 0.8-stable moving average random field with \(S\alphaS\) random measure $M$.

Stable Ornstein–Uhlenbeck process is a stable moving average process $X(t) = \int_{-\infty}^{t} e^{-\lambda(t-s)}M(ds), \ t \in \mathbb{R}$ where $M$ is a $\S\alphaS$ random measure on $\mathbb{R}$ with the Lebesgue control measure. The process $X = \{X(t), t \in \mathbb{R}\}$ is strictly stationary.

3. Linear Multifractional Stable Motion

is given by
\[ X(t) = \int_{\mathbb{R}} \left( (t-x)^{H(t)-1/\alpha} - (-x)^{H(t)-1/\alpha} \right) M(dx), \quad t \in \mathbb{R}, \]

where \( M \) is an \( \alpha \)-stable random measure with skewness function \( \beta \) and Lebesgue control measure, \( \alpha \in (0, 2] \). The continuous function \( H : \mathbb{R} \rightarrow (0, 1) \) is called a local scaling exponent, and \( (x)_+ = \max\{x, 0\} \). It is known that \( X \) is a locally self–similar random field, for more details see e.g. [39, 40]. In case \( \alpha = 2 \) we have a Gaussian process called multifractional Brownian motion, cf. [31]. For constant \( H \in (0, 1) \), we get the usual linear fractional stable motion which has stationary increments and is \( H \)-self–similar (see [38] Sect. 9.5)).

4. Stable Riemann–Liouville Process

It is given by \( R^H(t) = \int_{0}^{t} (t-s)^{-1/\alpha} M(ds), t \in \mathbb{R}_+ \), where \( M \) is an \( \alpha \)-stable random measure on \( \mathbb{R}_+ \) and \( H > 0 \). This is a family of \( H \)-self–similar random processes. Notice that \( R^H \) has no stationary increments, unless \( H = 1/\alpha \). For \( \alpha = 2 \) we get the Gaussian Riemann–Liouville process, see e.g. [23, Example 3.4].

5. Sub–Gaussian Random Fields

are fields \( X \) of the form

\[ X \overset{d}{=} \{ A^{1/2} G(t), t \in \mathbb{R}^d \}, \]

where \( A \sim S_{\alpha/2}((\cos(\pi \alpha/4))^{2/\alpha}, 1, 0) \) and \( G = \{ G(t), t \in \mathbb{R}^d \} \) is a zero mean Gaussian random field with a positive definite covariance function which is independent of \( A \). The following lemma (cf. [32] Proposition 3.8.1) shows that \( X \) is \( \alpha \)-stable.

**Lemma 4.** If the random variable \( A \) is as above and \( \xi \sim N(0, 2\sigma^2) \) independent of \( A \) then \( X = A^{1/2} \xi \sim S_{\alpha}(\sigma, 0, 0) \).

To prove the lemma, it suffices to calculate the characteristic function of \( X \) using the conditional expectation provided that \( A \) is fixed. If \( G \) is stationary then the resulting sub–Gaussian field \( X \) is strictly stationary as well.

A strictly stationary sub–Gaussian random field \( X \) with a mean square continuous Gaussian component \( G \) is not ergodic since it differs from \( G \) by a random scaling. A sufficient condition for ergodicity of \( G \) is that its spectral measure has no atoms, see [41] Theorem A).

3 Extrapolation of Stationary Random Fields

Let \( X = \{ X(t), t \in \mathbb{R}^d \} \) be a stationary (in the appropriate sense to be specified later) random field. We are looking for a linear predictor \( \hat{X}(t) \) of the unknown field value
Extrapolation of Stationary Random Fields 25

\( X(t) \) at location \( t \in \mathbb{R}^d \) based on observations \( X(t_1), \ldots, X(t_n) \) at locations \( t_1, \ldots, t_n, \) \( n \in \mathbb{N} \) of the form

\[
\hat{X}(t) = \sum_{i=1}^{n} \lambda_i(t)X(t_i) + \lambda_0(t).
\] (13)

The weights \( \lambda_0(\cdot), \ldots, \lambda_n(\cdot) \) are functions of \( t, t_1, \ldots, t_n \) which may depend on the distribution of \( X. \) For simplicity of notation, we omit all their arguments except for \( t. \) They have to be computed in a way (which depends on the integrability properties of \( X \)) such that the predictor is in some regard close to \( X(t). \)

**Definition 18.** A predictor \( \hat{X}(t) \) for \( X(t) \) is called

1. **exact** if \( \hat{X}(t) = X(t) \) a.s. whenever \( t = t_i \) for any \( i \in \{1, \ldots, n\} \). In this case, the predictor \( \hat{X}(\cdot) \) is an extrapolation surface for \( X(\cdot) \) with knots \( t_1, \ldots, t_n. \)
2. **unbiased** if \( \mathbb{E}|X(0)| < \infty \) and \( \mathbb{E}(|\hat{X}(t) - X(t)|) = 0. \)
3. **continuous** if weights \( \lambda_i(\cdot), \ i = 0, \ldots, n \) are continuous with respect to \( t, \) i.e., any realization of \( \hat{X} \) is continuous in \( t \in \mathbb{R}^d. \)

### 3.1 Kriging Methods for Square Integrable Random Fields

If the field \( X \) has finite second moments then the most popular prediction technique for \( X \) in geostatistics is the so–called **kriging.** It is named after D.G. Krige who first applied it (in 1951) to gold mining. Namely, he predicted the size of a gold deposit by collecting the data of gold concentration at some isolated locations. Apart from kriging, there are many other prediction techniques such as inverse distance, spline and nearest neighbor interpolation, triangulation, see for details [5, Sect. 5.9.2], [43, Chapt. 3], [37], etc. However, the latter methods ignore the correlation structure contained in the spatial data; see [5, Sect. 3.4.5, p.180; Chapt. 5.9], [8], [22] for their comparison.

The main idea of kriging is to compute prediction weights \( \lambda_i \) by minimizing the mean square error between the predictor and the field itself, i.e., solve the minimization problem

\[
\mathbb{E}(X(t) - \hat{X}(t))^2 \rightarrow \min_{\lambda_0, \ldots, \lambda_n}
\] (14)

under some additional conditions on \( \lambda_i \) for each fixed \( t \in \mathbb{R}^d. \)

Depending on the assumptions about \( X, \) numerous variants of kriging are available. We mention just few of them and refer an interested reader to the vast literature.

1. **Simple kriging:** for square integrable random fields \( X \) with known mean function \( \mathbb{E}X(t) = m(t), \ t \in \mathbb{R}^d. \) See Section 3.2
2. **Ordinary kriging:** for second order intrinsic stationary random fields \( X \) (with unknown but constant mean). See Section 3.3
3. **Kriging with drift:** \( \mathbb{E}X(t) = a + b\|t\|, \ a, b \in \mathbb{R} \) and these constants are unknown. See [4] Sect. 3.4.6 for details.
4. Universal kriging: the unknown mean $E_X(t) = m(t) \neq \text{const}$ belongs to some parametric family of functions, see \[4, 42\]. Ordinary kriging and kriging with drift are special cases of universal kriging.

3.2 Simple Kriging

Let $X$ be a square integrable random field with known mean function $m(t)$. It is easy to see that the minimum of the mean square error

$$ E(X(t) - \hat{X}(t))^2 = \text{var}(X(t) - \hat{X}(t)) + (E(X(t) - \hat{X}(t)))^2 $$

is attained exactly when the predictor $\hat{X}(t)$ is unbiased, i.e. if $E\hat{X}(t) = EX(t)$. This yields

$$ \lambda_0(t) = m(t) - \sum_{i=1}^n \lambda_i(t)m(t_i) $$

and

$$ \hat{X}(t) = \sum_{i=1}^n \lambda_i(t)(X(t_i) - m(t_i)) + m(t). $$

It follows from the above relation that the knowledge of function $m$ leads to centering the field $X$ (subtracting $m$) in the prediction.

Taking derivatives of the goal function in (14) with respect to $\lambda_i$, we obtain

$$ \sum_{i=1}^n \lambda_i(t)\text{cov}(X(t_i),X(t_j)) = \text{cov}(X(t),X(t_j)), \quad j = 1, \ldots, n. \quad (15) $$

The matrix form of this system of equations is

$$ \Sigma \cdot \lambda(t) = \sigma(t), $$

where $\Sigma = [\text{cov}(X(t_i),X(t_j))]_{i,j=1}^n$ is the covariance matrix, $\lambda(t) = (\lambda_1(t), \ldots, \lambda_n(t))^\top$,

$$ \sigma(t) = (\text{cov}(X(t),X(t_1)), \ldots, \text{cov}(X(t),X(t_n)))^\top. $$

If $\Sigma$ is non-degenerate then the solution exists and is unique. The covariance matrix is non-degenerate if the covariance function of $X$ is positive definite and all $t_i, i = 1, \ldots, n$ are distinct.

**Exercise 19.** Let the random field $X = \{X(t), \; t \in \mathbb{R}^d\}$ be as above. Show that the random vector $(X(t_1), \ldots, X(t_n))^\top$ is singular iff $\det \Sigma = 0$. *Hint:* A symmetric matrix is positive definite (positive semi–definite) if and only if all of its eigenvalues are positive (non–negative).

Finally, we have the following form of the predictor:

$$ \hat{X}(t) = \hat{X}^\top \Sigma^{-1} \sigma(t), $$

where $\hat{X} = (X(t_1), \ldots, X(t_n))^\top$. 
Let $\delta_{ij} = 1(i = j)$ be the Kronecker delta.

Properties of Simple Kriging

1. **Exactness**: to see that $\hat{X}(t_j) = X(t_j)$ for any $j$, set $t = t_j$ and check that $\lambda_i(t_j) = \delta_{ij}, i, j = 1, \ldots, n$ is the solution of system of equations (15).

2. **Continuity and smoothness**: rewrite (16) as $\hat{X}(t) = b^T \sigma(t)$ with $b = \Sigma^{-1} \bar{X}$ which means that sample path properties of the extrapolation surface such as continuity and smoothness directly depend on the properties of $\sigma(t)$. Thus if the covariance function is continuous and smooth, so is the extrapolation surface. See Figure 5(b).

3. **Shrinkage property**: The mean prediction error $E(\hat{X}(t) - X(t))^2$ can be found by direct calculations using the system (15). Thus

$$E(\hat{X}(t) - X(t))^2 = \text{var}X(t) - \text{var}\hat{X}(t).$$

Equation (17) yields the following shrinkage property: for all $t \in \mathbb{R}^d$

$$\text{var}\hat{X}(t) \leq \text{var}X(t).$$

The simple kriging predictor is less dispersed than the data. In a sense, kriging performs linear averaging (or smoothing) and does not perfectly imitate the trajectory properties of the original random field.

4. **Geometric interpretation**: The predictor $\hat{X}(t)$ for any fixed $t$ can be interpreted as a metric projection of $X(t)$ onto the linear subspace $L_n = \text{span}\{X(t_1), \ldots, X(t_n)\}$ of Hilbert space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with scalar product $\langle X, Y \rangle = E(XY)$ for $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, that is,

$$\hat{X}(t) = \text{Proj}_{L_n} X(t) = \arg\min_{\xi \in L_n} \langle X(t) - \xi, X(t) - \xi \rangle.$$

Equation (17) yields the following geometric property:

$$\text{var}\hat{X}(t) \leq \text{var}X(t).$$

It is known from the Hilbert space theory that this projection is unique if the vector $(X(t_1), \ldots, X(t_n))^T$ is not singular (cf. Definition 10).

5. **Orthogonality**: The above projection is also orthogonal, i.e., $\langle \hat{X}(t) - X(t), \xi \rangle = 0$ for all $\xi \in L_n$. In particular, it holds

$$\langle \hat{X}(t) - X(t), X(t_i) \rangle = 0 \text{ for all } i = 1, \ldots, n$$

which rewrites as a dependence relation

$$E(\hat{X}(t)X(t_i)) = E(X(t)X(t_i)) \text{ for all } i = 1, \ldots, n$$

yielding

$$\text{cov}(\hat{X}(t) - X(t), \hat{X}(s)) = 0, \quad s \in \mathbb{R}^d.$$
**Exercise 20.** Prove relation (17) via the Pythagorean theorem.

6. **Gaussian case:** Under the assumptions that $X$ is Gaussian and $\Sigma$ is non–singular it is easy to show that

$$\hat{X}(t) = \mathbb{E}(X(t)|X(t_1), \ldots, X(t_n)), \quad t \in \mathbb{R}^d. \quad (21)$$

**Exercise 21.** Prove relation (21) using the uniqueness of the kriging predictor and the following properties of the conditional expectation and of the Gaussian multivariate distribution, respectively:

1) $\mathbb{E}((\eta - \mathbb{E}(\eta|\xi))h(\xi)) = 0$ for random variables $\xi, \eta$ and any measurable function $h(\cdot)$,

2) If $\eta, \xi_1, \ldots, \xi_n$ are jointly Gaussian then there exist real numbers \(\{a_i\}_{i=1}^n\) such that $\mathbb{E}(\eta|\xi_1, \ldots, \xi_n) = \sum_{i=1}^n a_i \xi_i$.

In the Gaussian case, simple kriging has additional properties of

a. **Conditional unbiasedness:** $\mathbb{E}(X(t)|\hat{X}(t)) = \hat{X}(t)$ a.s. for any $t \in \mathbb{R}^d$, cf. [4, p. 164]. This property is important in practice for resource assessment problems and selective mining.

b. **Homoscedasticity:** The conditional mean square estimation error does not depend on the data, i.e.,

$$\mathbb{E}\left((\hat{X}(t) - X(t))^2|X(t_1), \ldots, X(t_n)\right) = \mathbb{E}(\hat{X}(t) - X(t))^2 \quad \text{a.s. for any } t \in \mathbb{R}^d.$$

### 3.3 Ordinary Kriging

When the mean $m$ of a square integrable random field $X$ is constant but unknown the ordinary kriging can be applied. We are looking for a predictor in the form (13).

For an arbitrary (but fixed) location $t \in \mathbb{R}^d$, the mean square prediction error is

$$\mathbb{E}(\hat{X}(t) - X(t))^2 = \text{var}(\hat{X}(t) - X(t)) + \left(\lambda_0 + \left(\sum_{i=1}^n \lambda_i - 1\right) m\right)^2.$$

Assuming that

$$\lambda_0 = 0, \quad \sum_{i=1}^n \lambda_i = 1, \quad (22)$$

we get the smallest possible error together with unbiasedness $\mathbb{E}\hat{X}(t) = \mathbb{E}X(t)$. The ordinary kriging predictor writes then

$$\hat{X}(t) = \sum_{i=1}^n \lambda_i X(t_i), \quad t \in \mathbb{R}^d.$$
The prediction error can be computed as

\[
E(\hat{X}(t) - X(t))^2 = \sum_{i,j=1}^{n} \lambda_i \lambda_j \text{cov}(X(t_i), X(t_j)) - 2 \sum_{i=1}^{n} \lambda_i \text{cov}(X(t_i), X(t)) + \text{var}X(t).
\]

One should minimize this error under the constraint (22).

Taking partial derivatives of the Lagrange function

\[
L(\lambda, \mu) = E(\hat{X}(t) - X(t))^2 + 2\mu \left( \sum_{i=1}^{n} \lambda_i - 1 \right)
\]

with respect to \(\lambda_i = \lambda_i(t), i = 1, \ldots, n\), and \(\mu = \mu(t)\) and putting them equal to zero we obtain the following system of \(n + 1\) linear equations

\[
\begin{align*}
\sum_{i=1}^{n} \lambda_i \text{cov}(X(t_i), X(t_j)) + \mu = \text{cov}(X(t_j), X(t)) , \quad j = 1, \ldots, n, \\
\sum_{i=1}^{n} \lambda_i = 1
\end{align*}
\]

for each \(t \in \mathbb{R}^d\) of interest. The solution \((\lambda_1, \ldots, \lambda_n, \mu)^\top\) of this system is unique iff the covariance matrix of the vector \((X(t_1), \ldots, X(t_n))^\top\) is non–singular.

The above linear system of equations can be rewritten in terms of variogram \(\gamma(\cdot, \cdot)\). By formula (2) and direct calculation we get the following ordinary kriging system of equations with respect to the weights \(\lambda_i, i = 1, \ldots, n\), and \(\mu\):

\[
\begin{align*}
\sum_{i=1}^{n} \lambda_i \gamma(t_i, t_j) + \mu = \gamma(t_j, t), \quad j = 1, \ldots, n, \\
\sum_{i=1}^{n} \lambda_i = 1
\end{align*}
\]

The corresponding mean square prediction error is

\[
\sigma_{\text{OK}}^2 = E(\hat{X}(t) - X(t))^2 = \sum_{i=1}^{n} \lambda_i \gamma(t_i, t) + \mu.
\]

**Exercise 22.** Show that \(\mu = -(1 - e^\top \Gamma^{-1} \gamma)/e^\top \Gamma^{-1} e\), where \(e\) is the unit vector, \(\gamma = (\gamma(t_1, t), \ldots, \gamma(t_n, t))^\top\) and \(\Gamma = [\gamma(t_i, t_j)]_{i,j=1,\ldots,n}\).

The main advantage of this way of posing the problem is that it is solvable even if the variance of \(X(t)\) is infinite whereas the variogram is finite, e.g., if \(X\) is intrinsic stationary of order two.

**Properties of the Ordinary Kriging**

1. **Exactness:** For \(t = t_j\), notice that \(\lambda_i(t_j) = \delta_{ij}, i, j = 1, \ldots, n, \mu(t_j) = 0\) is a solution of the ordinary kriging system.
2. **Orthogonality:** For any real weights \(a_i, i = 1, \ldots, n\) with the property \(\sum_{i=1}^{n} a_i = 1\) it holds
\[
\left\langle \hat{X}(t) - X(t), \sum_{i=1}^{n} a_i X(t_i) \right\rangle = 0.
\]

3. **Conditional unbiasedness:** The ordinary kriging predictor reduces the conditional bias \(\mathbb{E} \left[ X(t) \mid \hat{X}(t) \right] - \hat{X}(t)\). To see this, check the following formula showing that the minimum of the kriging error corresponds to the minimum of the conditional bias error:
\[
\mathbb{E} \left[ (X(t) - \hat{X}(t))^2 \right] = \mathbb{E} \left[ (\hat{X}(t) - X(t))^2 \right] - \mathbb{V} \left[ X(t) \mid \hat{X}(t) \right],
\]

as well as \(\mathbb{E}(Y \cdot \mathbb{E}(Z \mid Y)) = \mathbb{E}(YZ)\) for any random variables \(Y, Z\) defined on the same probability space.

**Example 4.** A simulated realization (see Figure 5(a)) of a centered stationary isotropic Gaussian random field \(X = \{X(t), t \in \mathbb{R}^d\}\) with Whittle–Matérn–type covariance function \(C(s,t) = 21(s = t) + 21(s \neq t) \|s-t\| K(2 \|s-t\|)\) exhibiting a nugget effect of height one is observed on a grid of locations \ \{(3i,2j), i,j \in \mathbb{N} \cap [0,3]\}\. The corresponding theoretical variogram together with the Matheron estimator (given in [38], Formula (9.67)) are shown on Figure 6. A Whittle–Matérn–type variogram model with a nugget effect \(\sigma^2\)
\[
\gamma(s,t) = 1(s \neq t) \left( \sigma^2 + b - b 2^{1-v}(a \|s-t\|)^v K_v(a \|s-t\|) \right), \quad s,t \in \mathbb{R}^d,
\]
was fitted to the estimated variogram by an ordinary least squares method yielding the parameter estimates \(\hat{\sigma}^2 = 0.932766523543886, \hat{a} = 1.9674556902269302, \hat{b} = 1.0672476194785714\). An extrapolation by ordinary kriging with the fitted variogram model \(\gamma\) is shown on Figure 5(b).

### 4 Extrapolation of Stable Random Fields

Let \(X\) be an \(\alpha\)-stable random field having integral representation
\[
X(t) = \int_{\mathbb{R}^d} f_i(x) M(dx), \quad t \in \mathbb{R}^d,
\]
confer formula (10). For \(\alpha \in (1,2]\), assume that the field \(X\) is centered. If \(\alpha \in (0,1]\), the mean value of \(X\) does not exist.

We are looking for a predictor \(\hat{X}(t)\) of the value \(X(t)\) at location \(t \in \mathbb{R}^d\) based on the random vector \((X(t_1), \ldots, X(t_n))^T\) in the form
Extrapolation of Stationary Random Fields

(a) Simulated realization of a stationary Gaussian random field with nugget effect.
(b) Extrapolation by ordinary kriging for the field in Figure 5(a)

Fig. 5 Application of ordinary kriging to simulated data from Example 4

Fig. 6 Theoretical variogram (red), estimator (green) and fitted variogram (black) for the realization in Figure 5(a), compare Example 4

\[ \hat{X}(t) = \sum_{i=1}^{n} \lambda_i X(t_i). \]  

Let \( T_j = \{t_{j,1}, \ldots, t_{j,n_j}\}, \) \( j \in \mathbb{N} \) be a sequence of observation locations such that \( \text{dist}(T_j,t) \to 0 \) as \( j \to \infty \) where \( \text{dist}(A,B) = \inf \{\|x - y\| : x \in A, y \in B\} \) is the Euclidean distance between two arbitrary sets \( A, B \subset \mathbb{R}^d \). The predictor \( \hat{X}_j(t) = \sum_{i=1}^{n_j} \lambda_i^{(j)} X(t_{ji}) \) is weakly consistent if \( \hat{X}_j(t) \xrightarrow{P} X(t) \) for any \( t \in \mathbb{R}^d \). It is stochastically continuous if \( \hat{X}_j(s) \xrightarrow{P} \hat{X}_j(t) \) for any \( j \in \mathbb{N} \) and \( t \in \mathbb{R}^d \).

Let

\[ \|f\|_\alpha = \left( \int_E |f(x)|^\alpha m(dx) \right)^{1/\alpha} \]  

(25)

denote the norm of \( f \in L^\alpha(E,m), \alpha \geq 1 \).
Theorem 3. Let the $\alpha$-stable random field $X$ in (23) be stochastically continuous, $\alpha \in (1, 2]$. Let the predictor $\hat{X}_j$ defined above exist and be unique, exact and stochastically continuous for any $j \in \mathbb{N}$. Then $\hat{X}_j$ is weakly consistent.

Proof. Fix an arbitrary $t \in \mathbb{R}^d$. By [32, Proposition 3.5.1], it is sufficient to show that $\sigma_{\hat{X}_j(t) - X(t)} \to 0$ as $j \to \infty$ to prove weak consistency. Let $s_j \in T_j$ be the point at which $\text{dist}(s_j, t) = \text{dist}(T_j, t)$ for any $j \in \mathbb{N}$. It is clear that $s_j \to t$ as $j \to \infty$. Since $\hat{X}_j$ is exact it holds $\hat{X}_j(s_j) = X(s_j)$ for any $j$. Then we have

$$\sigma_{\hat{X}_j(t) - X(t)} \leq \sum_{i=1}^{n_j} \lambda_i^{(j)} f_{t,i} - f_t \| \alpha \leq \sum_{i=1}^{n_j} \lambda_i^{(j)} f_{t,i} - f_s \| \alpha \to 0$$

as $j \to \infty$ by [32, Proposition 3.5.1], stochastic continuity of $X$ and $\hat{X}_j$ as well as exactness of $\hat{X}_j$.

4.1 Least Scale Predictor

For $\alpha \in (0, 2]$, consider the following optimization problem

$$\int_E |f_t(x) - \sum_{i=1}^n \lambda_i f_{t,i}(x)|^\alpha m(dx) \to \min_{\lambda_1, \ldots, \lambda_n}.$$ (26)

It is clear that the solution of the optimization problem (in case if it exists and is unique) will be an extrapolation. To see this, put $t = t_j$ and $\lambda_i(t_j) = \delta_i, j = 1, \ldots, n$.

The predictor $\tilde{X}(t)$ based on a solution of this minimization problem is called least scale linear (LSL) predictor. This method is reminiscent of the least mean square error property (14) of the kriging.

If $\alpha \in (1, 2]$ it is easy to see that any solution of the problem (26) is also a solution of the following system of equations

$$\int_E f_t(x) \left( f_t(x) - \sum_{i=1}^n \lambda_i f_{t,i}(x) \right)^{<\alpha-1>} m(dx) = 0, \quad j = 1, \ldots, n,$$ (27)

or equivalently

$$\left[ X(t_j), X(t) - \sum_{i=1}^n \lambda_i X(t_i) \right]_\alpha = 0, \quad j = 1, \ldots, n,$$ (28)

where $[\cdot, \cdot]_\alpha$ is the covariation, see Definition 16.

Exercise 23. Show that any solution of the problem (26) solves also the system of equations (27) or (28). Use the dominated convergence theorem.

1 The idea of this proof belongs to Adrian Zimmer.
Notice that equations in (26) are nonlinear in \( \lambda_1, \ldots, \lambda_n \) if \( \alpha < 2 \) because the covariation is not linear in the second argument (cf. Section 4.4). Thus, numerical methods have to be applied to solve problem (26).

**Properties of LSL predictor**

Assume \( 1 < \alpha \leq 2 \). For the case \( 0 < \alpha \leq 1 \) see Section 4.4.

**Theorem 4.** The LSL predictor exists. If the random vector \( \mathbf{X} = (X(t_1), \ldots, X(t_n))^{\top} \) is full–dimensional then the LSL predictor is unique.

**Proof.** We are using the properties of the best approximation in \( L^\alpha(E, m) \)-spaces for \( 1 < \alpha \leq 2 \). Let \( L = \text{span} \{ f_1, \ldots, f_n \} \). This is a finite dimensional space. Denote for simplicity \( f = f_t \) and \( E(f) = \inf_{x \in L} \| f - x \|_\alpha \). Let us show that this infimum is attained in \( L \).

Consider \( \{ x_m \}_{m \in \mathbb{N}} \) such that \( x_m \in L \ \forall m \in \mathbb{N} \) and \( \| x_m - f \|_\alpha \to E(f) \) as \( m \to \infty \).

By the triangle inequality \( \| x_m \|_\alpha \leq \| f \|_\alpha + \| x_m - f \|_\alpha \), so \( \{ x_m \}_{m \in \mathbb{N}} \) is a bounded sequence in a finite dimensional subspace. Thus, there exists a convergent subsequence \( \{ m_j \}_{j \in \mathbb{N}} \) and \( f_0 \in L \) such that \( \| x_{m_j} - f_0 \|_\alpha \to 0 \) as \( j \to \infty \). Since \( \| f - x_{m_j} \|_\alpha \to \| f - f_0 \|_\alpha \) and \( \| f - x_{m_j} \|_\alpha \to E(f) \) as \( j \to \infty \), it holds \( E(f) = \| f - f_0 \|_\alpha \). So \( f_0 \) is the best approximation.

For the proof of uniqueness, we use the strict convexity property. If \( \alpha > 1 \) the space \( L^\alpha(E, m) \) is strictly convex (see e.g. [5], p. 59]), i.e. for all \( g_1, g_2 \in L^\alpha(E, m) \) such that \( \| g_1 \|_\alpha = \| g_2 \|_\alpha = 1 \), \( g_1 \neq g_2 \) it follows \( \| \beta g_1 + (1 - \beta) g_2 \|_\alpha < 1 \) for any \( \beta \in (0, 1) \).

Take \( y_j = \sum_{i=1}^n \lambda_i^{(j)} f_i \in L \), \( j = 1, 2 \) such that \( y_1 \neq y_2 \) and \( \| f - y_1 \|_\alpha = \| f - y_2 \|_\alpha = E(f) \). Thus by strict convexity we have

\[
E(f) \leq \left\| f - \frac{1}{2} (y_1 + y_2) \right\|_\alpha = \left\| \frac{1}{2} (f - y_1) + \frac{1}{2} (f - y_2) \right\|_\alpha < E(f).
\]

So we obtain a contradiction, and \( y_1 = y_2 = f_0 \). By the full–dimensionality of the random vector \( \mathbf{X} \) and by Lemma 3 one can easily see that the set of the weights \( \lambda_i \) in the representation \( f_0 = \sum_{i=1}^n \lambda_i f_i \) is unique.

**Theorem 5 ([16]).** Let the stable random field \( X \) in (23) be stochastically continuous. If the random vector \( \mathbf{X} = (X(t_1), \ldots, X(t_n))^{\top} \) is full–dimensional then the LSL predictor is continuous.

### 4.2 Covariation Orthogonal Predictor

Throughout this Section, assume \( \alpha \in (1, 2] \). The linear predictor (24) with weights \( \lambda_1, \ldots, \lambda_n \) that are a solution of the system of equations
is called \textit{Covariation Orthogonal Linear (COL) predictor}. If the solution of (29) exists and is unique then it is an exact predictor, since we can put \( \lambda_i(t_j) = \delta_{ij}, i, j = 1, \ldots, n \). This extrapolation method is reminiscent of the generic orthogonality property of simple kriging, cf. relation (20). It is also \textit{symmetric} (in a sense) to the LSL predictor, compare the systems (28) and (29). In contrast to (28), the system (29) is linear which makes the computation of the weights \( \lambda_i \) easier.

Introduce the \textit{covariation function} \( \kappa : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \) of \( X \) by

\[
\kappa(s, t) = [X(s), X(t)]_\alpha.
\] (30)

Note that this function is not symmetric in its arguments, as opposed to the covariance function, cf. Definition 6.

By additivity of the covariation in the first argument (see Section 2.4), the system (29) rewrites as

\[
\begin{pmatrix}
\kappa(t_1, t_1) & \cdots & \kappa(t_n, t_1) \\
\vdots & \ddots & \vdots \\
\kappa(t_1, t_n) & \cdots & \kappa(t_n, t_n)
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{pmatrix} =
\begin{pmatrix}
\kappa(t, t_1) \\
\vdots \\
\kappa(t, t_n)
\end{pmatrix}.
\] (31)

If matrix \( K = [\kappa(t_i, t_j)]_{i,j=1,\ldots,n} \) is positive definite the solution of this system exists and is unique.

For moving average and for sub–Gaussian fields \( X \), sufficient conditions for the positive definiteness of \( K \) can be given.

4.2.1 The COL Predictor for Moving Averages

Consider a moving average stable random field \( X \) with representation

\[
X(t) = \int_{\mathbb{R}^d} f(t-x) M(dx), \quad t \in \mathbb{R}^d,
\]

where \( M \) is an \( \alpha \)–stable random measure with Lebesgue control measure and \( f \in L^\alpha(\mathbb{R}^d) \) (see Section 2.5 for the definition). By strict stationarity of \( X \), it holds \( [X(h), X(0)]_\alpha = [X(t+h), X(t)]_\alpha \) for all \( t, h \in \mathbb{R}^d \). With slight abuse of notation, we write \( \kappa(s-t) = [X(s-t), X(0)]_\alpha = \kappa(s, t), s, t \in \mathbb{R}^d \) and the system of equations (31) is equivalent to

\[
\begin{pmatrix}
\kappa(0) & \cdots & \kappa(t_n-t_1) \\
\vdots & \ddots & \vdots \\
\kappa(t_n-t_1) & \cdots & \kappa(0)
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{pmatrix} =
\begin{pmatrix}
\kappa(t-t_1) \\
\vdots \\
\kappa(t-t_n)
\end{pmatrix}.
\] (32)
The next theorem gives a sufficient condition for the existence and uniqueness of the COL predictor.

**Theorem 6.** If the kernel \( f : \mathbb{R}^d \to \mathbb{R}_+ \) is a positive definite function that is positive on a set of non–zero Lebesgue measure then \( \kappa \) is positive definite.

**Proof.** By formula (12), we have

\[
\kappa(h) = \int_{\mathbb{R}^d} f(h - x) f^{(\alpha - 1)}(-x) \, dx, \quad h \in \mathbb{R}^d.
\]

Thus for any \( m \in \mathbb{N}, z_1, \ldots, z_m \in \mathbb{R}, (z_1, \ldots, z_m)^\top \neq (0, \ldots, 0)^\top \) and \( s_1, \ldots, s_m \in \mathbb{R}^d \) it holds

\[
\sum_{i,j=1}^{m} \kappa(s_i - s_j) z_i z_j = \int_{\mathbb{R}^d} \sum_{i,j=1}^{m} f(s_i - s_j - x) z_i z_j f^{(\alpha - 1)}(-x) \, dx > 0.
\]

An example of a process \( X \) satisfying conditions of Theorem 6 is the \( S\alpha S \) Ornstein–Uhlenbeck process: for any fixed \( \lambda > 0 \)

\[
X(t) = \int_{\mathbb{R}} e^{-\lambda(t-x)} 1(t-x \geq 0) M(dx), \quad t \in \mathbb{R}.
\]

By [32, p. 138], we have \( \hat{X}(t) = e^{-\lambda(t-t_n)} X(t_n) \) if \( t_1 < \ldots < t_n < t \).

**Theorem 7.** If the covariation function \( \kappa \) is positive definite and continuous then the COL predictor is continuous.

**Proof.** Since \( \kappa \) is positive definite, matrix \( K \) is invertible, and we have

\[
\begin{pmatrix}
\lambda_1(t) \\
\vdots \\
\lambda_n(t)
\end{pmatrix} = \begin{pmatrix}
\kappa(0) & \cdots & \kappa(t_n - t_1) \\
\vdots & \ddots & \vdots \\
\kappa(t_n - t_1) & \cdots & \kappa(0)
\end{pmatrix}^{-1} \begin{pmatrix}
\kappa(t-t_1) \\
\vdots \\
\kappa(t-t_n)
\end{pmatrix}.
\]

Since \( \kappa \) is continuous, the weights \( \lambda_1, \ldots, \lambda_n \) are continuous in \( t \).

**Exercise 24.** Show that continuous kernel functions with compact support yield a continuous covariation function \( \kappa \). Use the dominated convergence theorem.

### 4.2.2 The COL Predictor for Gaussian and sub–Gaussian Random Fields

Let \( X \) be a sub–Gaussian random field, i.e., \( X(t) = A^{1/2} G(t), \ t \in \mathbb{R}^d \) where \( A \sim S_{\alpha/2}((\cos(\pi \alpha/4))^{2/\alpha}, 1, 0) \) and \( G \) is a zero mean stationary Gaussian field independent of \( A \). In [32, Example 2.7.4], it is shown that for sub–Gaussian random fields, the covariation function is given by

\[
\kappa(h) = 2^{-\alpha/2} C(h) C(0)^{(\alpha - 2)/2}, \quad h \in \mathbb{R}^d,
\]

(33)
where $C(\cdot)$ is the covariance function of $G$.

It is easy to see that in this case the system (31) coincides with the simple kriging system (15) for $G$:

$$
\begin{pmatrix}
C(0) & \cdots & C(t_n - t_1) \\
\vdots & \ddots & \vdots \\
C(t_n - t_1) & \cdots & C(0)
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{pmatrix}
= 
\begin{pmatrix}
C(t - t_1) \\
\vdots \\
C(t - t_n)
\end{pmatrix}.
$$

(34)

If $C$ is positive definite then the corresponding covariance matrix is invertible which ensures the existence and uniqueness of the solution of the system (34).

**Theorem 8.** If $(X(t_1), \ldots, X(t_n))^\top$ is full–dimensional and the covariance function $C$ of the Gaussian component is continuous then the COL predictor for sub–Gaussian random fields is continuous.

The proof is similar to the proof of Theorem 7.

**Theorem 9.** Let $1 < \alpha \leq 2$. For Gaussian and sub–Gaussian random fields, the COL and LSL predictors coincide.

**Proof.** Introduce the notation $t_0 = t$. Put $\hat{\lambda}_0(t_0) = -1$ and

$$
\hat{X}(t_0) - X(t_0) = A^{1/2} \sum_{i=0}^n \lambda_i(t_0) G(t_i).
$$

The characteristic function of random vector $(X(t_0), \ldots, X(t_n))^\top$ is given by

$$
E \exp \left\{ i \sum_{k=0}^n \theta_k X(t_k) \right\} = \exp \left\{ - \frac{1}{2} \sum_{i=0}^n \sum_{j=0}^n \theta_i \theta_j C(t_i - t_j) \right\} \left( \sigma_{X(t_0)} \right)^{\alpha/2}.
$$

(35)

for all $\theta_1, \ldots, \theta_n \in \mathbb{R}$, cf. [32, Proposition 2.5.2]. Now it is simple to see that

$$
\sigma_{\hat{X}(t_0) - X(t_0)} = \left( \frac{1}{2} \text{var} \left( \sum_{i=0}^n \lambda_i(t_0) G(t_i) \right) \right)^{1/2} = \left( \frac{1}{2} \sum_{i,j=0}^n \lambda_i \lambda_j C(t_i - t_j) \right)^{1/2}.
$$

Thus, the LSL optimization problem is equivalent to

$$
\sum_{i,j=0}^n \lambda_i \lambda_j C(t_i - t_j) \rightarrow \min_{\lambda_1, \ldots, \lambda_n}.
$$

Taking derivatives we obtain $\sum_{j=0}^n C(t_k - t_j) \lambda_j = 0, k = 1, \ldots, n$ which coincides with the COL extrapolation system (34).

**Remark 5.** It follows from the proof of Theorem 9 (which is valid for all $\alpha \in (0, 2)$) that the weights of the LSL predictor for sub–Gaussian random fields are a solution of the system (34) also in the case $\alpha \in (0, 1]$. The statement of Theorem 8 holds as
well. To summarize, the LSL predictor for stationary sub–Gaussian random fields \( X \) exists and is unique and exact for all \( \alpha \in (0, 2] \) if the covariance function \( C \) of the Gaussian component \( G \) is positive definite. If \( C \) is additionally continuous then this LSL predictor is also continuous.

### 4.3 Maximization of Covariation

In this section, we assume that \( X \) is an \( \alpha \)–stable random field with \( \alpha \in (1, 2] \). The predictor \( \hat{X}(t) = \sum_{i=1}^n \lambda_i(t) X(t_i) \), whose weights \( \lambda_1(t), \ldots, \lambda_n(t) \) solve the following optimization problem

\[
\begin{align*}
&\left[ \hat{X}(t), X(t) \right]_\alpha = \sum_{i=1}^n \lambda_i(t) \left[ X(t_i), X(t) \right]_\alpha \rightarrow \max_{\lambda_1, \ldots, \lambda_n}, \\
&\sigma_{\hat{X}(t)} = \sigma_X(t)
\end{align*}
\]

for \( t \in \mathbb{R}^d \), is called **Maximization of Covariation Linear (MCL) predictor**.

The Lagrange function of the optimization problem (36) is given by

\[
L(\lambda, \gamma) = \sum_{i=1}^n \lambda_i [X(t_i), X(t)]_\alpha + \gamma \left( \sigma^{\alpha}_{\sum_{i=1}^n \lambda_i X(t_i)} - \sigma^{\alpha}_{\hat{X}(t)} \right), \quad \lambda \in \mathbb{R}^n, \quad \gamma \in \mathbb{R}.
\]

By taking partial derivatives and setting them equal to zero, we get

\[
\begin{align*}
&\left[ X(t_j), X(t) \right]_\alpha + \gamma \cdot \partial \sigma^{\alpha}_{\sum_{i=1}^n \lambda_i X(t_i)} / \partial \lambda_j = 0, \quad j = 1, \ldots, n, \\
&\sigma^{\alpha}_{\sum_{i=1}^n \lambda_i X(t_i)} = \sigma_X(t).
\end{align*}
\]

Analogously to formula (28) one can obtain

\[
\frac{\partial \sigma^{\alpha}_{\sum_{i=1}^n \lambda_i X(t_i)}}{\partial \lambda_j} = \alpha \cdot \left[ X(t_j), \sum_{i=1}^n \lambda_i(t) X(t_i) \right]_\alpha.
\]

Since \( \gamma = -1/\alpha \), \( \lambda_i(t_j) = \delta_{ij} \) is obviously a solution of system (37) for \( t = t_j, \ j = 1, \ldots, n \), the MCL predictor is exact.

Let us discuss the properties of the MCL predictor. Notice that here no direct analogy with kriging can be drawn. For instance, a counterpart \( \sigma_{\hat{X}(t)} \leq \sigma_X(t) \) of the shrinkage property (18) is deliberately mutated to the additional condition \( \sigma_{\hat{X}(t)} = \sigma_X(t) \). The reason for this is that both conditions lead to the same solutions due to the convexity of the optimization problem (36).

Introduce the following notation: \( \zeta(t) = (\kappa(t_1, t), \ldots, \kappa(t_n, t))^\top, \ t \in \mathbb{R}^d \), the function \( \sigma_0 : \mathbb{R}^d \rightarrow \mathbb{R}_+ \) is \( \sigma_0(t) = \sigma_X(t) = \kappa(t, t) \). The function \( \Psi : \mathbb{R}^n \rightarrow \mathbb{R}_+ \) is defined by
\[ \Psi(\lambda) = \sigma_{\hat{x}(t)} = \left\| \sum_{i=1}^{n} \lambda_i f_i \right\|_\alpha. \]

Denote the level set of function \( \Psi \) at level \( u \in \mathbb{R} \) by \( B_u = \{ \lambda \in \mathbb{R}^n : \Psi(\lambda) \leq u \} \). The support set of any convex set \( B \subset \mathbb{R}^n \) at a point \( x \in \mathbb{R}^n \) is defined by

\[ T(B, x) = \left\{ y \in B : \langle y, x \rangle = \sup_{z \in B} \langle z, x \rangle \right\}. \]

It is known that for strictly convex sets \( B \) and any non-zero \( x \in \mathbb{R}^n \) the support set \( T(B, x) \) is a singleton. We denote this single point by \( y_{B, x} \).

**Theorem 10.** Assume that the \( \alpha \)-stable random vector \( X = (X(t_1), \ldots, X(t_n))^\top \) is full-dimensional.

1. The solution of the optimization problem (36) exists for all \( t \in \mathbb{R}^d \). If \( \kappa(t, t) \neq 0 \) for some \( i = 1, \ldots, n \) then the MCL predictor \( \hat{x}(t) \) is unique.

2. If \( \kappa \) is a continuous function on \( \mathbb{R}^d \times \mathbb{R}^d \) and \( \kappa(t, t) \neq 0 \) for some \( i = 1, \ldots, n \) then the MCL predictor is continuous in \( t \).

**Proof.** For the proof of the existence and uniqueness of MCL we refer the reader to the paper [10]. It is also shown there that the vector of MCL weights

\[ \lambda(t) = (\lambda_1(t), \ldots, \lambda_n(t))^\top \]

is equal to \( y_{B_{\alpha(t)}} \xi(t) \) for any \( t \in \mathbb{R}^d \) whereas the set \( B_{\alpha(t)} \) is strictly convex. Let us prove that \( \lambda : \mathbb{R}^d \to \mathbb{R}^n \) is a continuous function. It is easy to see that \( B_{\alpha(t)} = \frac{1}{\alpha(t)} B_1 \), because the sets \( B_{\alpha(t)}, t \in \mathbb{R}^d \) are homothetic, i.e. \( a B_{\alpha(t)} = B_{\alpha(t)}/a, a > 0 \). Thus by simple geometric considerations

\[ T(B_{\alpha(t)}, \xi(t)) = T \left( \frac{1}{\alpha(t)} B_1, \xi(t) \right) = \frac{1}{\alpha(t)} T(B_1, \xi(t)), \]

thus \( \lambda(t) = \frac{1}{\alpha(t)} y_{B_1, \xi(t)} \). Put \( B = B_1 \) and \( x(s) = y_{B, \xi(s)} \) for any \( s \in \mathbb{R}^d \). Show that

\[ \lim_{s \to t} x(s) = x(t). \]

This limit exists by the definition of the support set and continuity of the scalar product. We know that \( \xi(s) \to \xi(t) \) as \( s \to t \) since \( \kappa \) is a continuous function. Moreover, \( B \) is a compact, and \( x(s) \in B \) for all \( s \). Choose a convergent sequence \( s_m \to t \) as \( m \to \infty \) such that \( x(s_m) \to y \) as \( m \to \infty \), where \( y \in B \). Show that \( y = x(t) \). It is clear that \( \langle x(s_m), \xi(s_m) \rangle \to \langle y, \xi(t) \rangle \) as \( m \to \infty \). And for any \( x \in B \) it holds

\[ \langle x, \xi(t) \rangle = \lim_{m \to \infty} \langle x(s_m), \xi(s_m) \rangle \leq \lim_{m \to \infty} \langle x(s_m), \xi(s_m) \rangle = \langle y, \xi(t) \rangle. \]

The inequality here is due to the fact that \( \{ x(s_m) \} = T(B, \xi(s_m)) \) for any \( m \in \mathbb{N} \).

Thus \( y = y_{B, \xi(t)} \).
4.4 Case $\alpha \in (0, 1]$

As noticed in Section 2.4, the covariation function is not defined for $\alpha \in (0, 1]$. Moreover, the function $\| \cdot \|_\alpha$ for $\alpha < 1$ defined in (25) is not a norm anymore since the triangle inequality fails to hold. The property of strict convexity of $L^\alpha(E, m)$ does not hold as well.

To cope with these drawbacks, one may come to an idea that the codifference (cf. Definition 17) can be used instead of the covariation in COL and MCL methods. However, it does not seem to make advances in extrapolation. For instance, replacing the covariation by the codifference in the MCL method leads to the optimization problem

\[
\begin{align*}
\begin{cases}
\tau(\hat{X}(t), X(t)) = \sigma_{\hat{X}(t)} + \sigma_{X(t)} - \sigma_{\hat{X}(t) - X(t)} & \rightarrow \max_{\lambda_1, \ldots, \lambda_n} \\
\sigma_{\hat{X}(t)} = \sigma_{X(t)}.
\end{cases}
\end{align*}
\]

Using the constraint $\sigma_{\hat{X}(t)} = \sigma_{X(t)}$, the first relation rewrites

\[
\tau(\hat{X}(t), X(t)) = 2\sigma_{X(t)} - \sigma_{\hat{X}(t) - X(t)}.
\]

Hence, the method (38) is equivalent to LSL extrapolation, i.e., to minimizing the scale parameter

\[
\sigma_{\hat{X}(t) - X(t)} = \| f_t - \sum_{i=1}^n \lambda_i f_{t_i} \|_\alpha
\]

of $\hat{X}(t) - X(t)$.

Replacing the covariation by the codifference in the COL method (29), one arrives at the system of nonlinear equations

\[
\begin{align*}
\tau_{\hat{X}(t), X(t)}(\hat{X}(t), X(t)) = \tau_{\hat{X}(t), X(t)}(X(t)), \quad i = 1, \ldots, n.
\end{align*}
\]

Here the numerical computation of a solution is necessary, which can be very time consuming. Furthermore, it is shown in [11] that the solution of the system (39) is not unique. For this reason, we shall not pursue the method (39) in future.

Neither leads the maximization of $\tau_{\hat{X}(t), X(t)}$ with respect to weights $\lambda_1, \ldots, \lambda_n$ to a unique predictor (24). In particular, its existence is not really clear. As an example consider a random field (23) with the kernel function $f_t$ of compact support such that the supports of $f_t$ and $f_{t_1}, \ldots, f_{t_n}$ do not overlap. Then it is easy to see that $\tau_{\hat{X}(t), X(t)} = 0$ allowing for an arbitrary choice of weights $\lambda_1, \ldots, \lambda_n$.

In the remainder of this Section, we focus on the properties of the LSL method for $\alpha$–stable random fields with $\alpha \in (0, 1]$. First of all, the fundamental question of existence has to be answered. Here we follow [11] and do this in a more general setting of $r$–normed vector spaces.

**Definition 19.** Let $V$ be a vector space over a field $\mathbb{K}$. A map $\| \cdot \|_r : V \to \mathbb{R}_+$ is called an $r$–norm, if there exists $K \geq 1$ and $r > 0$ such that
Let $V$ be a vector space over $\mathbb{R}$ with $r$-norm $\| \cdot \|_{(r)}$ and let $f_1, \ldots, f_n \in V$ be linearly independent. For any $f_0 \in V$, there exist real numbers $\lambda^*_1, \ldots, \lambda^*_n$ such that

$$
\| f_0 - \sum_{i=1}^n \lambda^*_i f_i \|_{(r)} = \inf_{\lambda_1, \ldots, \lambda_n} \| f_0 - \sum_{i=1}^n \lambda_i f_i \|_{(r)}.
$$

If we set $V = L^\alpha(E, m)$ and note that $\| \cdot \|_{\alpha} = \| \cdot \|_{(r)}$ defined in (25) is an $\alpha$-norm on $L^\alpha(E, m)$ (even a norm if $\alpha \geq 1$), the existence of the LSL predictor follows immediately from Theorem 11. In contrast to the case $\alpha \in (1, 2]$ (Theorem 4), the uniqueness of the LSL weights $\lambda^* := (\lambda^*_1, \ldots, \lambda^*_n)^\top$ in Theorem 11 is not guaranteed. We illustrate this by the following example. Introduce the notation $H_\alpha(\lambda) = \sigma_{X(t) - X(i)}$ for $\lambda = (\lambda_1, \ldots, \lambda_n)^\top \in \mathbb{R}^n$.

**Example 5.** Consider the measurable space $(E, m) = ([0, 1], \mathcal{V}_1)$ and the kernel function $f_t(x) = 1 \{ x \in (t + \frac{1}{2}, t + \frac{3}{2}) \}$. Given $t_1 = \frac{1}{4}$, predict the value of the symmetric $\alpha$-stable process $X(t) = \int_{[0,1]} f_t(x) M(dx)$ at the point $t = 0$. By elementary calculations we obtain

$$
H_\alpha(\lambda) = \int_{[0,1]} |f_t(x) - \lambda f_{t_1}(x)|^\alpha dx = \frac{1}{4} (1 + |1 - \lambda|^\alpha + |\lambda|^\alpha).
$$

It is easy to see that for $0 < \alpha < 1$, $H_\alpha$ has two global minima at $\lambda = 0$ and $\lambda = 1$. If $\alpha = 1$ the set of all global minimum points equals the interval $[0, 1]$. For values $\alpha > 1$, the function $H_\alpha$ has a unique global minimum at $\lambda = 0.5$.

In order to get unbiased prediction (provided that the first moment of $X$ is finite), the parameter space is often restricted to $\{ (\lambda_1, \ldots, \lambda_n)^\top \in \mathbb{R}^n : \sum_{i=1}^n \lambda_i = 1 \}$. S. Hagel showed in [11] that this restriction does not cause uniqueness of LSL prediction for $\alpha \in (0, 1)$. Alternatively, the following algorithmic approach to choose a unique global minimum in the LSL optimization problem is proposed:

**Algorithm 1** Let $\{ X(t), t \in T \}$ be an $\alpha$–stable random field (25) with $0 < \alpha < 1$ and $T \subset \mathbb{R}^d$. Let $t_1, \ldots, t_n \in T$ be fixed such that functions $f_{t_1}, \ldots, f_{t_n}$ are linearly independent.

1. Order the points $t_1, \ldots, t_n$ so that

$$
\| t - t_1 \| \leq \| t - t_2 \| \leq \ldots \leq \| t - t_n \|.
$$
Extrapolation of Stationary Random Fields

and if \( \|t - t_i\| = \|t - t_{i+1}\| \) for some \( i \in \{1, \ldots, n-1\} \) then

\[
\begin{align*}
\hat{i}_i^{(p)} &= i_{i+1}^{(p)} \quad \text{for all } p = 1, \ldots, k-1 \\
\hat{i}_i^{(k)} &= i_{i+1}^{(k)}
\end{align*}
\]

(40)

(41)

for some \( k \in \{1, \ldots, m\} \), where \( t_i^{(p)} \) is the \( p \)-th component of \( t_i \).

2. Determine the set \( A_0 \) of all critical points

\[
A_0 = \{ (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n : H_\alpha(\lambda_1, \ldots, \lambda_n) = \inf_{(\mu_1, \ldots, \mu_n) \in \mathbb{R}^n} H_\alpha(\mu_1, \ldots, \mu_n) \}
\]

3. Reduce \( A_0 \) step by step to sets \( A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \) given by

\[
A_j = \{ (\lambda_1, \ldots, \lambda_n) \in A_{j-1} : \lambda_j = \max_{(\mu_1, \ldots, \mu_n) \in A_{j-1}} \mu_j \}, \quad j = 1, \ldots, n.
\]

Clearly, the set \( A_n \) consists of just one element.

**Definition 20.** We call \( \hat{X}(t) = \sum_{j=1}^n \hat{\lambda}_j^* X(t_j) \) the **best LSL predictor** if \( (\hat{\lambda}_1^*, \ldots, \hat{\lambda}_n^*) \in A_n \).

The above construction has a simple intuitive meaning. The points \( t_1, \ldots, t_n \) are ordered with respect to their distance to \( t \). To get a unique ordering, conditions (40) and (41) are required. Points with a smaller distance to \( t \) are regarded to exert more influence on the value of \( X \) at \( t \), so their weights should be maximized first.

To show that \( A_j \neq \emptyset \), \( j = 1, \ldots, n \) we notice that \( A_0 \) is nonempty and compact. Therefore, the projection mapping \( (x_1, \ldots, x_n) \mapsto x_1 \) takes its maximum on \( A_0 \). Hence, \( A_1 \) is nonempty and compact as well. Sets \( A_2, \ldots, A_n \) are not empty by induction.

It can be easily proved that the best LSL predictor is exact. To see this, let \( t = t_i \) for some \( i \in \{1, \ldots, n\} \) and let \( t_1, \ldots, t_n \in \mathbb{R}^d \) be as in Algorithm 1. Relations (40) and (41) then imply that \( t = t_i \). Trivially, \( (1, 0, \ldots, 0) \in A_0 \) holds. Due to the linear independence of \( f_1, \ldots, f_n \), it holds that \( A_n = A_0 = \{(1, 0, \ldots, 0)\} \).

For \( 1 < \alpha \leq 2 \), Theorem 5 stated the continuity of LSL prediction. In contrast, the best LSL predictor is not necessarily continuous for \( 0 < \alpha \leq 1 \) as the next example shows.

**Example 6.** Let \( X = \{X(t), t \in \mathbb{R}^2\} \) be an \( \alpha \)-stable random field \( (2.3) \) with \( 0 < \alpha < 1 \), \( f_i(x) = 1 \left( x \in \left( \min\{t^{(1)}, t^{(2)}\}, \max\{t^{(1)}, t^{(2)}\} \right) \right) \) for \( t = (t^{(1)}, t^{(2)}) \in \mathbb{R}^2 \), \( E = \mathbb{R} \) and \( M \) being a \( \mathcal{S} \alpha \mathcal{S} \) random measure on \( \mathbb{R} \) with Lebesgue control measure. It follows from relations (8), (9) and Markov inequality that \( X \) is stochastically continuous, i.e., it has a.s. no jumps at fixed locations \( t \). For \( n = 1 \), introduce \( t_0 = \left( \frac{1}{2}, \frac{1}{2} \right), \ t_1 = (0,1), t = t_0 + \varepsilon, \) where \( \varepsilon = (\delta, \delta) \in \mathbb{R}^2 \) for some \( \delta \in (-\frac{1}{2}, \frac{1}{2}) \) and \( \delta \in (-\frac{1}{2}, \frac{1}{2}) \). Consider the best LSL predictor \( \hat{X}(t) \) of \( X(t) \) based on the data \( X(t_1) \). It holds
Let the function $f_i$ have a global minimum at $\lambda = 0$ and if $\delta < 0$ it has a global minimum at $\lambda = 1$. So $X(t)$ is discontinuous at $t = t_0$.

In addition to the best LSL prediction, it is possible to treat the case $\alpha = 1$ similar to the case $1 < \alpha < 2$. The following approach is proposed in [11]. For a symmetric $1$–stable field $\{X(t) : t \in T\}$ with integral representation

$$X(t) = \int_E f_i(x) M(dx)$$

let the function $f_i \in L^1(E, \mathcal{E}, m) \cap L^\delta(E, \mathcal{E}, m)$ for some $\delta > 1$. Then we have

$$\int_E |f_i(x) - \sum_{i=1}^n \lambda_i f_i(x)|^\gamma m(dx) < \infty$$

for all $\gamma \in [1, \delta]$, $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and $t, t_1, \ldots, t_n \in T$. Now fix $t, t_1, \ldots, t_n \in T$ and chose an arbitrary sequence $\langle \gamma_k \rangle_{k \in \mathbb{N}} \subset (1, \delta)$ which converges to 1 as $k \to \infty$. Let $\langle \lambda_1^{(\gamma_k)}, \ldots, \lambda_n^{(\gamma_k)} \rangle$ be the unique solution of

$$\int_E |f_i(x) - \sum_{i=1}^n \lambda_i f_i(x)|^{\gamma_k} m(dx) \to \min_{\lambda_1, \ldots, \lambda_n}.$$ (42)

Applying the stability theorem in [19, p.225] it follows the convergence

$$\int_E |f_i(x) - \sum_{i=1}^n \lambda_i^{(\gamma_k)} f_i(x)|^{\gamma_k} m(dx) \to \inf_{\mu_1, \ldots, \mu_n} \int_E |f_i(x) - \sum_{i=1}^n \mu_i f_i(x)| m(dx)$$ (43)

as $k \to \infty$. Moreover, it can be shown that

$$\langle \lambda_1^{(\gamma_k)}, \ldots, \lambda_n^{(\gamma_k)} \rangle \to \langle \lambda_1^*, \ldots, \lambda_n^* \rangle, \quad k \to \infty.$$  

This set of weights $\langle \lambda_1^*, \ldots, \lambda_n^* \rangle$ exists and is unique if all LSL prediction problems (42) with stability indices $\gamma_k > 1$ do so. It also does not depend on the choice of the sequence $\langle \gamma_k \rangle_{k \in \mathbb{N}} \subset (1, \delta)$ such that $\gamma_k \to 1$ as $k \to \infty$.

**Definition 21.** The predictor $\hat{X}^*(t) = \sum_{i=1}^n \lambda_i^* X(t_i), t \in T$ is called an index–continuous LSL predictor (ICLSL) for the symmetric $1$–stable random field $X$.

It is still an open problem to explore the statistical properties of ICLSIL.

---

3 Personal communication of Adrian Zimmer
4.5 Numerical Examples

In this section, LSL, COL and MCL extrapolation methods (as well as Maximum Likelihood extrapolation and conditional simulation for sub–Gaussian random fields) are applied to simulated data of various \( \alpha \)--stable random processes and fields \( X \) for \( \alpha \in (0, 2) \).

The random fields are simulated and extrapolated on an equidistant 50 \times 50 –grid of points within \( T = [0, 1]^2 \). In Examples 1 and 2, the simulated field \( X = \{X(t), t \in [0, 1]^2\} \) is observed at the points \( t_1, \ldots, t_{16} \) given by their coordinates:

- \( t_1 = (0, 0), \quad t_2 = (0, 0.3), \quad t_3 = (0, 0.6), \quad t_4 = (0, 0.9), \)
- \( t_5 = (0.3, 0), \quad t_6 = (0.3, 0.3), \quad t_7 = (0.3, 0.6), \quad t_8 = (0.3, 0.9), \)
- \( t_9 = (0.6, 0), \quad t_{10} = (0.6, 0.3), \quad t_{11} = (0.6, 0.6), \quad t_{12} = (0.6, 0.9), \)
- \( t_{13} = (0.9, 0), \quad t_{14} = (0.9, 0.3), \quad t_{15} = (0.9, 0.6), \quad t_{16} = (0.9, 0.9). \)

1. Sub–Gaussian Random Fields

Consider a stationary sub–Gaussian random field \( X \) described in Example 5 of Section 2.5 with \( \alpha = 1.2 \). The Gaussian part \( G \) of this field has a Whittle–Matérn covariance function (cf. Section 2.2.1, Example 6) with parameters as in Figure 1(a). Figure 7(a) shows a realization of \( X \). The corresponding LSL (coinciding with COL by Theorem 9) and MCL predictors can be seen in Figures 7(b) and 7(c). Both predictions are smoother than the realization of the field itself. Since predictions in Figures 7(b) and 7(c) look quite similar and can not be told one from another by eye, their difference is given in Figure 7(d).

Figure 8(a) shows a realization of the stationary sub–Gaussian field with \( \alpha = 0.8 \) and covariance function \( C \) of the Gaussian part as above. A Maximum Likelihood (ML) predictor for sub–Gaussian random fields is introduced in [16]. It is shown in Theorem 11 of that paper that LSL, COL and ML methods coincide if \( \alpha \in (1, 2) \). However, its proof does not depend on \( \alpha \) covering (with regard to Remark 5 of this chapter) the range of all \( \alpha \in (0, 2) \). Thus, LSL and ML predictors coincide for sub–Gaussian random fields with any stability index \( \alpha \in (0, 2) \). A possibility of extrapolation of sub–Gaussian random fields \( X \) by conditional simulation (CS) of the Gaussian component \( G \) of \( X \) and the subsequent scaling by \( \sqrt{A} \) is straightforward; see e.g. [20] and [14, p. 112]. Algorithms for the conditional simulation of \( G \) are given in [21]. Corresponding extrapolation results for LSL (ML) and CS methods are given in Figures 8(b) and 8(c). Notice that the ML prediction for this realization of \( X \) is much smoother than CS prediction.

2. Skewed stable Lévy Motion

Consider the two–dimensional 1.5–stable Lévy motion \( X \) defined by
44 Evgeny Spodarev, Elena Shmileva and Stefan Roth

(a) Realization of a sub–Gaussian random field with $\alpha = 1.2$.

(b) Corresponding LSL (COL) predictor

(c) Corresponding MCL predictor

(d) Pointwise difference ((b)-(c)) between LSL (b) and MCL (c) predictors

Fig. 7 Realization of a sub–Gaussian random field for $\alpha > 1$ and different predictors

\[ X(t) = \int_0^1 \int_0^1 \mathbf{1}(x_1 \leq t_1, x_2 \leq t_2) M(d(x_1, x_2)), \quad t = (t_1, t_2)^\top \in [0, 1]^2, \]

where $M$ is a non–symmetric centered 1.5–stable random measure with skewness intensity $\beta = 1$. Comparing a realization of $X$ (Figure 9(a)) with its LSL, COL and MCL predictors (Figures 9(b), 9(c) and 9(d)) one can see that prediction has a smoothing effect.

3. Stable Ornstein–Uhlenbeck Process

Let $X$ be a 1.6-stable Ornstein–Uhlenbeck process with $\lambda = 0.5$ defined in Example 2 of Section 2.5. Figure 10 shows a trajectory of this process and different interpolators. The process $X$ is observed at positions $t_i = 1, \ldots, 10$ within $[0, 10]$. It can be seen that LSL interpolation is very smooth. In contrast, the COL predictor is piecewise smooth and continuous on the whole interval.
Extrapolation of Stationary Random Fields

(a) Realization of a sub-Gaussian random field with $\alpha = 0.8$.

(b) Corresponding LSL (ML) predictor

(c) Prediction by conditional simulation

Fig. 8 Realization of a sub-Gaussian random field for $\alpha < 1$ and different predictors

4. Stable Moving Average

Let $X = \{X(t), t \in [0, 0.49]^2\}$ be a moving average field (cf. Example 2 of Section 2.5) with the kernel function

$$f(x) = 0.5 \left(0.04 - \|x\|^2\right) \mathbf{1}(\|x\| \leq 0.2),$$

stability index $\alpha = 0.5$ and skewness intensity $\beta = 0.8$. Random field $X$ is simulated on an equidistant $50 \times 50$–grid of points within $[0, 0.49]^2$ using the step function approach from paper [15] with an accuracy ($L^\alpha$-error) $\varepsilon = 0.01$. The field is observed at points

$$t_1 = (0, 0), \quad t_2 = (0, 0.25), \quad t_3 = (0, 0.49),$$
$$t_4 = (0.25, 0), \quad t_5 = (0.25, 0.25), \quad t_6 = (0.25, 0.49),$$
$$t_7 = (0.49, 0), \quad t_8 = (0.49, 0.25), \quad t_9 = (0.49, 0.49).$$

To solve the optimization problems for the best LSL prediction (cf. Section 4.4) numerically, an average of 8 realizations of the simulated annealing algorithm from [17] is used. Figures 11(a) and 11(b) show a realization of $X$ and its best LSL predic-
Evgeny Spodarev, Elena Shmileva and Stefan Roth

(a) Realization of stable Lévy motion with skewness intensity $\beta = 1$ and $\alpha = 1.5$

(b) Corresponding LSL predictor

(c) Corresponding COL predictor

(d) Corresponding MCL predictor

Fig. 9 Realization of a skewed stable Lévy motion field and different predictors

The numerical optimization procedure is quite time consuming with 136 min. of computation time (Pentium Dual Core E5400, 2.70 GHz, 8 GB RAM) per extrapolation.

5 Open problems

In contrast to kriging methods, there is no common methodology of measuring prediction errors in the stable case. We propose the following measures

$$\sup_{t \in \mathbb{R}^d} \left( \mathbb{E} |X(t) - \hat{X}(t)|^p \right)^{1/p} = c_\alpha(p) \sup_{t \in \mathbb{R}^d} \|f_t - \sum_{i=1}^n \lambda_i f_i\|_\alpha,$$

where $1 < p < \alpha$ and $c_\alpha(p) > 0$ is a constant from relation (8), or
Fig. 10 A trajectory (black) of the stable Ornstein–Uhlenbeck process together with LSL (red), COL (green) and MCL (blue) predictors, $\alpha = 1.6$.

Fig. 11 Realization of a skewed 0.5–stable moving average random field.

(a) Realization of a skewed 0.5–stable moving average random field

(b) Corresponding best LSL predictor

Realization of a skewed moving average field with $\alpha = 0.5$ and its best LSL predictor.
\[
P \left( \sup_{t \in \mathbb{R}^d} |X(t) - \hat{X}(t)| > \varepsilon \right), \quad \varepsilon > 0.
\] (45)

It is an open problem to find lower and upper bounds for these errors as well as minimax bounds where the infimum over a subclass of stable random fields \(X\) is additionally considered in relations (44) and (45). Alternatively, one can be interested in the asymptotic behavior of \(P \left( \sup_{t \in \mathbb{R}^d} |X(t) - \hat{X}(t)| < \varepsilon \right)\) as \(\varepsilon \to 0\) which is related to small deviation problems.

**Acknowledgements** This research was partially supported by the DFG – RFBR grant 09–01–91331. The second author was also supported by the Chebyshev Laboratory (Department of Mathematics and Mechanics, St.-Petersburg State University) within RF government grant 11.G34.31.0026.

**References**

1. Bochner, S.: Lectures on Fourier integrals. Princeton Univ. Press (1959)
2. Breiman, L.: Probability. Addison-Wesley, Massachusetts (1968)
3. Bulinski, A.V., Shiryaev, A.N.: Theory of Stochastic Processes. Fizmatlit (2005). (in Russian)
4. Chilès, J.P., Delfiner, P.: Geostatistics: Modeling Spatial Uncertainty. Wiley, New York (1999)
5. Cressie, N.: Statistics for Spatial Data. Wiley, New York (1991)
6. DeVore, R.A., Lorentz, G.G.: Constructive Approximation. Springer-Verlag, Berlin, Heidelberg (1993)
7. Foss, S., Korshunov, D.A., Zachary, S.: An Introduction to Heavy-tailed and Subexponential Distributions. Springer (2011)
8. Franke, R.: Scattered data interpolation: Tests of some methods. Mathematics of computation \textbf{38}, 181–200 (1982)
9. Gandin, L.S.: Objektivnyj Analiz Meteorologiceskikh Polej. Gidrometeoiizadosat, Leningrad (1963)
10. Gneiting, T., Sasvári, Z., Schlather, M.: Analogies and correspondences between variograms and covariance functions. Adv. Appl. Probab. \textbf{33}, 617–630 (2001)
11. Hagel, S.: Extrapolation von stabilen Zufallsfeldern. Master’s thesis, Ulm University (2012)
12. Ibragimov, I.A., Rozanov, Y.A.: Gaussian Random Processes, Applications of Mathematics, vol. 9, Springer, New York-Berlin (1978)
13. Jeulin, D., Jeulin, P.: Synthesis of rough surfaces by random morphological models. Stereologica Iugoslavia \textbf{3}, 239–246 (1981)
14. Karcher, W.: On infinitely divisible random fields with an application in insurance. Ph.D. thesis, Ulm University, Ulm (2012)
15. Karcher, W., Scheffler, H.P., Spodarev, E.: Simulation of infinitely divisible random fields. Commun. Statist. Sim. Comput. \textbf{42}, 215–246 (2013)
16. Karcher, W., Shmileva, E., Spodarev, E.: Extrapolation of stable random fields. J. Multivar. Anal. \textbf{115}, 516–536 (2013). URL \url{http://dx.doi.org/10.1016/j.jmva.2012.11.003}
17. Kirkpatrick, S., Gelatt, C., Vecchi., M.: Optimization by simulated annealing. Science \textbf{220}, 671–680 (1983)
18. Kolmogorov, A.: Interpolation und Extrapolation von stationären zufälligen Folgen. Izv. Akad. Nauk SSSR \textbf{5}, 3–14 (1941)
19. Kosmol, P.: Optimierung und Approximation. Walter de Gruyter, Berlin (1991)
20. Krige, D.G.: A statistical approach to some basic mine valuation problems on the Witwatersrand. J. Chem. Metal. Min. Soc. S. Afr. 52(6), 119–139 (1951)
21. Lantuéjoul, C.: Geostatistical Simulation: Models and Algorithms. Springer-Verlag, Berlin Heidelberg (2002)
22. Laslett, G.: Kriging and splines: An empirical comparison of their predictive performance in some applications. J. Amer. Statist. Assoc. 89(426), 391–400 (1994)
23. Lifshits, M.: Lectures on Gaussian Processes. Springer (2012)
24. Magnus, W., Oberhettinger, F., Soni, R.P.: Formulas and Theorems for the Special Functions of Mathematical Physics, Die Grundlehren der mathematischen Wissenschaften in Einzel­darstellungen, vol. 52. Springer-Verlag, Berlin (1966)
25. Markovich, N.: Nonparametric Analysis of Univariate Heavy-Tailed data: Research and Practice. Wiley (2007)
26. Matern, B.: Spatial variation. Meddelanden från Statens Skogsforskningsinstitut 49, 1–144 (1960)
27. Matheron, G.: Traite de Geostatistique Appliquee, Tome I, Memoires de Bureau de Recherche Geologiques et Minieres, vol. 14. Editions Technip, Paris (1962)
28. Matheron, G.: Traite de Geostatistique Appliquee, Tome II: Le Krigeage, Memoires de Bureau de Recherche Geologiques et Minieres, vol. 24. Editions Bureau de Recherche Geologiques et Minieres, Paris (1963)
29. Nolan, J.: Stable Distributions – Models for Heavy Tailed Data. Birkhauser, Boston (2013). In progress, Chapter 1 online at http://academic2.american.edu/~jpnolan
30. Painter, S.: Numerical method for conditional simulation of Lévy random fields. Math. Geol. 30(2), 163–179 (1998)
31. Pelletier, R., Lévy-Véhel, J.: Multifractional Brownian motion: definition and preliminary results. Tech. Rep. RR-2645, INRIA, Le Chesnay, France (1995)
32. Samorodnitsky, G., Taqqu, M.S.: Stable Non-Gaussian Random Processes. Chapman & Hall, Boca Raton (1994)
33. Sato, K.I.: Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press (1999)
34. Schlather, M.: Some covariance models based on normal scale mixtures. Bernoulli 16(3), 780–797 (2010)
35. Schoenberg, I.J.: Metric spaces and complete monotone functions. Ann. Math. 39(4), 811–841 (1938)
36. Schoenberg, I.J.: Metric spaces and positive definite functions. Trans. Amer. Math. Soc. 44(3), 522–536 (1938)
37. Sibson, R.: A brief description of natural neighbor interpolation. In: V. Barnett (ed.) Interpreting Multivariate Data, pp. 21–36. John Wiley, Chichester (1981)
38. Spodarev, E. (ed.): Stochastic Geometry, Spatial Statistics and Random Fields. Asymptotic Methods, Lecture Notes in Mathematics, vol. 2068. Springer, Heidelberg (2013)
39. Stoev, S., Taqqu, M.: Stochastic properties of the linear multifractional stable motion. Adv. Appl. Probab. 36(4), 1085–1115 (2004)
40. Stoev, S., Taqqu, M.: Path properties of the linear multifractional stable motion. Fractals 13(2), 157–178 (2005)
41. Tempelman, A.A.: On ergodicity of Gaussian homogeneous random fields on homogeneous spaces. Theory Probab. Appl. 18(1), 173–175 (1973)
42. Wackernagel, H.: Multivariate Geostatistics. 2 edn. Springer, Berlin (2003)
43. Webster, R., Oliver, M.: Geostatistics for Environmental Scientists. Wiley (2007)
44. Wentzell, A.D.: A Course in the Theory of Stochastic Processes. McGraw-Hill (1981)
45. Wiener, N.: Extrapolation, Interpolation and Smoothing of Stationary Time Series. Wiley, New York (1949)
46. Yaglom, A.M.: Correlation Theory of Stationary and Related Random Functions, Volume I. Springer (1987)
47. Yaglom, A.M.: Correlation Theory of Stationary and Related Random Functions, Volume II. Springer (1987)
48. Zolotarev, V.M.: One-dimensional Stable Distributions. Translations of Mathematical Monographs, vol 65, American Mathematical Society, Providence (1986)
Index

$r$–norm, 39
Bessel family, 8
best LSL predictor, 41
Boolean random function, 5
Brownian Lévy field, 4

Cauchy family, 9
centered, 3
codifference, 22
conditionally negative semi–definite function, 7
covariance function, 7
covariation, 20
function, 34
Covariation Orthogonal Linear (COL) predictor, 34
cyclone model, 10
exponential model, 9
fractional Brownian field, 4
full-dimensional random vector, 15
Gaussian covariance family, 9
Gaussian linear random function, 3
geometric anisotropy, 10
hole effect model, 8
Hurst index, 4
independently scattered random measure, 18
index–continuous LSL predictor, 42
intrinsic stationarity of order two, 2
invariance in strict sense, 2
invariance in wide sense, 2
isotropy, 2

kernel function, 23
kriging, 25
Lévy process, 5
least scale linear (LSL) predictor, 32
level set, 38
linear multifractional stable motion, 23
linear predictor, 24
Maximization of Covariation Linear (MCL) predictor, 37
mixed anisotropy, 13
motion invariance, 2
normal scale mixture, 8
nugget effect, 12
ordinary kriging, 28
Ornstein-Uhlenbeck process, 3
Poisson shot noise field, 5
positive definite function, 6
positive semi–definite function, 6
predictor
  continuous, 25
  exact, 25
  stochastically continuous, 31
  unbiased, 25
  weakly consistent, 31
purely zonal anisotropy, 12
random field, 2
separable in probability random field, 20
shift of a stable vector, 15
shrinkage property, 27
sill, 12
| Term                                                      | Page(s) |
|-----------------------------------------------------------|---------|
| simple kriging                                            | 26      |
| conditional unbiasedness                                 | 28      |
| homoscedasticity                                          | 28      |
| singular random vector                                    | 15      |
| Gaussian                                                  | 15      |
| spectral density                                          | 7       |
| spectral measure                                          | 7       |
| of a stable vector                                        | 15      |
| spectral representation                                   | 7       |
| spherical model                                           | 10      |
| stability index                                           | 15      |
| stable family                                             | 9       |
| stable Lévy process                                       | 22      |
| stable moving average random field                        | 23      |
| stable Ornstein–Uhlenbeck process                         | 23      |
| stable random field                                       | 20      |
| stable random measure                                     | 18      |
| control measure                                           | 18      |
| skewness function                                         | 18      |
| stable random variable                                    | 17      |
| totally skewed                                            | 17      |
| stable random vector                                      | 15      |
| stable Riemann–Liouville process                          | 24      |
| stable subordinator                                       | 23      |
| stationarity                                              | 2       |
| strict convexity of $L^q(E, m)$                           | 33      |
| sub-Gaussian random field                                 | 24      |
| support set                                               | 38      |
| symmetric distribution of a random vector                 | 16      |
| variogram                                                 | 4, 11   |
| white noise                                               | 8       |
| Whittle-Matér family                                      | 9       |