Adv. Appl. Prob. **50,** 1155–1175 (2018)
*doi:*10.1017/apr.2018.54
© *Applied Probability Trust* 2018

EFFICIENT SIMULATION OF BROWN–RESNICK PROCESSES BASED ON VARIANCE REDUCTION OF GAUSSIAN PROCESSES

MARCO OESTING,* *University of Siegen*
KIRSTIN STROKORB,** *Cardiff University*

Abstract

Brown–Resnick processes are max-stable processes that are associated to Gaussian processes. Their simulation is often based on the corresponding spectral representation which is not unique. We study to what extent simulation accuracy and efficiency can be improved by minimizing the maximal variance of the underlying Gaussian process. Such a minimization is a difficult mathematical problem that also depends on the geometry of the simulation domain. We extend Matheron's (1974) seminal contribution in two directions: (i) making his description of a minimal maximal variance explicit for convex variograms on symmetric domains, and (ii) proving that the same strategy also reduces the maximal variance for a huge class of nonconvex variograms representable through a Bernstein function. A simulation study confirms that our noncostly modification can lead to substantial improvements among Gaussian representations. We also compare it with three other established algorithms.

Keywords: Brown–Resnick process; Gaussian process; max-stable process; simulation; spatial extremes; variance reduction; variogram

2010 Mathematics Subject Classification: Primary 60G70; 60G15
Secondary 60G60

1. Introduction

Many powerful tools in geostatistics are conveniently based on Gaussian processes as an underlying probabilistic model for uncertainty (Chilès and Delfiner (2012), Gelfand *et al.* (2010)). By contrast, assessing the extreme values of spatial data genuinely requires statistical methodology that goes beyond such tools. A common approach from extreme value analysis is the usage of max-stable models instead. In particular, the class of Brown–Resnick processes (Brown and Resnick (1977), Kabluchko *et al.* (2009)) has emerged as a now widely adopted class of processes considered in the analysis of spatial data, cf. e.g. Asadi *et al.* (2015), Buhl and Klüppelberg (2016), Davison *et al.* (2013), Einmahl *et al.* (2016), Engelke *et al.* (2015), Gaume *et al.* (2013), Oesting *et al.* (2017), Oesting and Stein (2017), Sang and Genton (2014), and Thibaud *et al.* (2016).

There is a strong connection between Brown–Resnick processes and Gaussian processes. First, Brown–Resnick processes arise as the only possible nondegenerate limits of maxima of appropriately rescaled independent Gaussian processes (Kabluchko *et al.* (2009), Kabluchko (2011)). Second, they can be represented as maxima of a convolution of the points of a

Received 21 September 2017; revision received 17 October 2018.
* Postal address: Department of Mathematics, University of Siegen, 57072 Siegen, Germany.
Email address: oesting@mathematik.uni-siegen.de
** Postal address: School of Mathematics, Cardiff University, Cardiff CF24 4AG, UK.
Email address: strokorbk@cardiff.ac.uk

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

Poisson point process and Gaussian processes. As the number of Gaussian processes involved in the maximum is locally finite, Brown–Resnick processes still inherit various properties from Gaussian processes on a local level. They are parsimonious models in the sense that their law is fully specified by a bivariate quantity, namely, the variogram of the underlying Gaussian process. On the other hand, they are still very flexible in the sense that various features such as smoothness, scale, or nugget effect can be controlled by the choice of variogram family. All of this makes Brown–Resnick processes popular, consistent spatial models and marks their status as a benchmark in spatial extremes.

To extract probabilistic properties of interest from a fitted Brown–Resnick model, it is usually necessary to be able to efficiently simulate from the fitted model. Meanwhile, several approaches for this task have been developed. Starting from the basic threshold stopping approach based on the work of Schlather (2002) using plainly the original definition of a Brown–Resnick process, Dieker and Mikosch (2015), Oesting *et al.* (2012), and Oesting *et al.* (2018) achieved further improvements that are based on modified spectral representations. Based on different techniques, Dombry *et al.* (2016) and Liu *et al.* (2016) proposed the extremal functions and the record-breakers approach, respectively, both of which together with the normalization method of Oesting *et al.* (2018) can now be seen as state-of-the-art algorithms for the exact simulation of Brown–Resnick processes.

When dealing with spatial data, the study area on which the process should be simulated may be large with respect to its spatial extent or the number of locations therein. In such situations, exact simulation of Brown–Resnick processes via one of the state-of-the-art algorithms can be very time consuming and, depending on the purpose of the application, it can be more appropriate to admit (desirably small) simulation errors. Once a trade-off between accuracy and efficiency is necessary, it is no longer clear which of the previously considered simulation approaches performs 'best'. Keeping this in mind, we return to the initial threshold stopping approach devised by Schlather (2002) from a new perspective in this work. We explore to what extent a modified choice of Gaussian spectral representation of Brown–Resnick processes can improve the efficiency or accuracy of the simulation and whether an improved threshold stopping approach can compete with state-of-the-art simulation when an error is admitted. Dealing with these questions ultimately leads us to a classical minimization problem for Gaussian processes, namely, to find a Gaussian process whose maximal variance across the simulation domain is minimized, while its variogram on this domain is fixed. In this regard, we extent Matheron's (1974) seminal contribution in two directions: (i) making his description of a minimal maximal variance explicit for convex variograms on symmetric domains, and (ii) proving that the same strategy also reduces the maximal variance for the huge class of nonconvex variograms that can be represented via Bernstein functions.

The paper is organized as follows. In Section 2 we recall the spectral representation of Brown–Resnick processes and revisit Schlather's threshold stopping algorithm for the simulation of max-stable processes on a compact domain. We explain why it is beneficial for the underlying Gaussian process to have a reduced maximal variance across the simulation domain. Subsequently, in Section 3, the main contribution, we provide two results complementing Matheron (1974) for the corresponding minimization problem and elaborate on discretization effects. Throughout the text, the most popular family of Brown–Resnick processes that are associated to fractional Brownian sheets figure as an example. In Section 4 we report the setup and results of a simulation study for this family of models, where we also compare our approach with three other methods. Finally, we end with a discussion of our findings in Section 5. Proofs and additional auxiliary results are deferred to Appendix A.

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

2. Spectral representations and threshold stopping

Let $Z = \{Z(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathbb{R}^d}$ be a *simple max-stable process* on $\mathbb{R}^d$, which means that, for each $n \in \mathbb{N}$ and n independent and identically distributed (i.i.d.) copies $Z_1, \dots, Z_n$ of Z, the process $Z_1 \vee \cdots \vee Z_n$ of the pointwise maxima has the same law as nZ and that Z has *standard Fréchet* margins: $\mathbb{P}(Z(\boldsymbol{x}) \le z) = \exp(-1/z)$ for $z > 0$, $\boldsymbol{x} \in \mathbb{R}^d$. It was shown by de Haan (1984) that, if Z is continuous in probability, there exists a nonnegative stochastic process $V = \{V(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathbb{R}^d}$ on $\mathbb{R}^d$ with $\mathbb{E}(V(\boldsymbol{x})) = 1$, $\boldsymbol{x} \in \mathbb{R}^d$, such that the law of the simple max-stable process Z is recovered from the following max-series:

$$\{Z(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathbb{R}^d} \overset{\mathrm{D}}{=} \left\{\bigvee_{k=1}^{\infty} U_k V_k(\boldsymbol{x})\right\}_{\boldsymbol{x}\in\mathbb{R}^d}. \tag{2.1}$$

Here $\{U_k\}_{k=1}^{\infty}$ denotes a Poisson point process on $(0, \infty)$ with intensity measure $u^{-2}\,\mathrm{d}u$, which is independent of the i.i.d. sequence $\{V_k\}_{k=1}^{\infty}$ of copies of V. The process V is often called a *spectral process* and representation (2.1) is referred to as a *spectral representation* of Z. By Giné *et al.* (1990), if Z has continuous sample paths, then the trajectories of V will also be continuous, and vice versa, such that the sequence $\{(U_k, V_k)\}_{k=1}^{\infty}$ may be considered as an independently marked Poisson process Π on $(0, \infty) \times C(\mathbb{R}^d)$ with intensity measure

$$\Lambda((u, \infty) \times A) = u^{-1}\mathbb{P}(V \in A), \qquad u > 0,\ A \subset C(\mathbb{R}^d) \text{ Borel},$$

where $C(\mathbb{R}^d)$ is endowed with the usual topology of uniform convergence on compact subsets.

One popular choice for the spectral process V of a max-stable process Z is a log-Gaussian process of the form

$$V(\boldsymbol{x}) = \mathrm{e}^{\eta(\boldsymbol{x})} \quad \text{with} \quad \eta(\boldsymbol{x}) = W(\boldsymbol{x}) - \frac{\operatorname{var}(W(\boldsymbol{x}))}{2}, \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{2.2}$$

where W is a zero-mean Gaussian process W with stationary increments. The associated max-stable process Z, called the *Brown–Resnick process*, was introduced and many of its properties were analysed in Kabluchko *et al.* (2009). The requirement that W has *stationary increments* means that the law of the process $\{W(\boldsymbol{x}+\boldsymbol{h}) - W(\boldsymbol{x})\}_{\boldsymbol{h}\in\mathbb{R}^d}$ does not depend on $\boldsymbol{x} \in \mathbb{R}^d$. Since W is zero-mean Gaussian, it is equivalent to the *intrinsic stationarity* of the process W, that is, the stationarity of the process $\{W(\boldsymbol{x}+\boldsymbol{h}) - W(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathbb{R}^d}$ for all $\boldsymbol{h} \in \mathbb{R}^d$, cf. e.g. Strokorb (2013, p. 108). It ensures that the resulting max-stable process Z is stationary and its law is uniquely specified by the variogram

$$\gamma(\boldsymbol{x} - \boldsymbol{y}) = \mathbb{E}(W(\boldsymbol{x}) - W(\boldsymbol{y}))^2, \qquad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d.$$

Gneiting *et al.* (2001) showed that a function $\gamma\colon \mathbb{R}^d \to \mathbb{R}$ is a (not necessarily centered) variogram of an intrinsically stationary Gaussian random field W if and only if $\gamma(\boldsymbol{0}) = 0$, and γ is *negative definite* in the sense that $\gamma(-\boldsymbol{x}) = \gamma(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^d$ and

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i \gamma(\boldsymbol{x}_i - \boldsymbol{x}_j) a_j \le 0 \tag{2.3}$$

for all finite systems $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathbb{R}^d$ and $a_1, \dots, a_n \in \mathbb{R}$ with $a_1 + \cdots + a_n = 0$. The most popular family of variograms used in practice is $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}/s\|^{\alpha}$ for some $\alpha \in (0, 2]$ and $s > 0$ corresponding to *fractional Brownian sheets*. This family of variograms will also serve as an illustrating example throughout this text. It can be shown that all fractional Brownian sheets admit continuous trajectories; cf. for instance, Theorem 1.4.1. of Adler and Taylor (2009). Hence, the resulting max-stable process is also continuous.

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

Henceforth, we will restrict our attention to a compact subset $K \subset \mathbb{R}^d$. Let $\{E_i\}_{i=1}^{\infty}$ and $\{V_k\}_{k=1}^{\infty}$ be two independent i.i.d. sequences of standard exponential random variables and copies of V, respectively, and set $U_k = (\sum_{i=1}^{k} E_i)^{-1}$, $k = 1, 2, \ldots$. As the points $\{(U_k, V_k)\}_{k=1}^{\infty}$ form a Poisson point process with intensity measure Λ, the finite approximations

$$Z^{(n)}(\boldsymbol{x}) = \bigvee_{k=1}^{n} U_k V_k(\boldsymbol{x}), \qquad \boldsymbol{x} \in K,$$

almost surely converge to a limit process $Z^{(\infty)}$ (as $n \to \infty$) satisfying $Z^{(\infty)} \overset{\mathrm{D}}{=} Z$. To obtain an approximation built from a finite number of exponential random variables E_k and stochastic processes V_k, often the process $Z^{(T)}$ is considered, where T is a stopping time defined as

$$T = T_{K,\tau} = \min\left\{k \geq 1 \colon U_{k+1}\tau \leq \inf_{\boldsymbol{x}\in K} Z^{(k)}(\boldsymbol{x})\right\} \quad \text{for some } \tau > 0.$$

Such an approximation $Z^{(T)}$ yields an exact simulation of a general max-stable process Z, i.e. $Z^{(T)} = Z^{(\infty)}$ almost surely (a.s.), if the condition $\sup_{\boldsymbol{x}\in K} V(\boldsymbol{x}) < \tau$ is a.s. satisfied; for such situations, this approach has been proposed in Schlather (2002). In the case when V is log-Gaussian, however, as in (2.2), an approximation error may occur. The probability that such an error occurs is given by

$$\begin{aligned}\mathcal{P}_{K,\tau} &= \mathbb{P}(Z^{(T)} \neq Z^{(\infty)} \text{ on } K) \\ &= 1 - \mathbb{E}_{Z^{(T)}}\left\{\exp\left(-\mathbb{E}_V\left(\sup_{\boldsymbol{x}\in K} \frac{V(\boldsymbol{x})}{Z^{(T)}(\boldsymbol{x})} - \sup_{\boldsymbol{x}\in K} \frac{\tau}{Z^{(T)}(\boldsymbol{x})}\right)_+\right)\right\} \qquad (2.4)\end{aligned}$$

(Oesting and Strokorb (2018)), where the symbol $\mathbb{E}$ with a stochastic process as a subscript means that the expectation is meant with respect to this process.

The situation is particularly difficult when the spectral functions' variance $\mathrm{var}(V(\boldsymbol{x}))$ is large, since trajectories $U_k V_k(\boldsymbol{x})$ that are generated late in a hypothetical unstopped simulation can have an outsize influence on the sample path, i.e. stopping too soon has profound consequences on the quality of the output. Specifically, for Brown–Resnick processes, the variance $\mathrm{var}(V(\boldsymbol{x}))$ of the log-Gaussian spectral functions in (2.2) is of order $\exp(\mathrm{var}(W(\boldsymbol{x})))$ and this order can be excessive depending on the values of $\mathrm{var}(W(\boldsymbol{x}))$ on the simulation domain K. For instance, if W is the standard ('original') fractional Brownian sheet that vanishes at the origin $\mathbf{0} \in \mathbb{R}^d$ (i.e. $W(\mathbf{0}) = 0$ a.s.), the variance $\mathrm{var}(V(\boldsymbol{x}))$ is exponential in $\|\boldsymbol{x}\|^{\alpha}$. This poses a major challenge for stopping-time-based approximate simulations.

What might however mitigate this challenge to some extent is the fact that there is still some choice among the log-Gaussian representations of Brown–Resnick processes. While the variogram γ uniquely determines the law of the associated Brown–Resnick process Z, it does not uniquely determine the law of the Gaussian process W in its spectral representation (2.2). Various covariance functions $C(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{cov}(W(\boldsymbol{x}), W(\boldsymbol{y}))$ on K share the same variogram

$$\gamma(\boldsymbol{x} - \boldsymbol{y}) = C(\boldsymbol{x}, \boldsymbol{x}) - 2C(\boldsymbol{x}, \boldsymbol{y}) + C(\boldsymbol{y}, \boldsymbol{y}), \qquad \boldsymbol{x}, \boldsymbol{y} \in K.$$

In what follows, we seek to find such Gaussian representations W for a prescribed variogram γ, whose variances $\mathrm{var}(V(\boldsymbol{x}))$ are uniformly small over the simulation domain K. More precisely, we address the following problem in Section 3 and study the consequences of using low-variance Gaussian representations in the threshold stopping procedure in Section 4.

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

Problem 2.1. Let $\mathcal{W}_K(\gamma)$ be the set of Gaussian processes with variogram γ on the compact simulation window K. If it exists, identify a Gaussian process $W_{\min} \in \mathcal{W}_K(\gamma)$ that minimizes the functional

$$W \mapsto \sup_{\boldsymbol{x}\in K} \operatorname{var}(W(\boldsymbol{x})).$$

Remark 2.1. Addressing Problem 2.1 will also be beneficial for reducing the error probability $\mathcal{P}_{K,\tau}$ in (2.4) for high thresholds τ or several other error terms in Oesting and Strokorb (2018). To see this, note that Proposition A.1 given in Appendix A ensures that, for each positive $z \in C(K)$, the tail probabilities

$$\mathbb{P}\left(\sup_{\boldsymbol{x}\in K} \frac{V(\boldsymbol{x})}{z(\boldsymbol{x})} > u\right) = \mathbb{P}\left(\sup_{\boldsymbol{x}\in K} \exp\left\{W(\boldsymbol{x}) - \frac{\operatorname{var}(W(\boldsymbol{x}))}{2} - \log z(\boldsymbol{x})\right\} > u\right)$$

decay as fast as possible as $u \uparrow \infty$ if V is built from the solution of Problem 2.1. This property entails that

$$\mathbb{E}\left(\sup_{\boldsymbol{x}\in K} \frac{V(\boldsymbol{x})}{z(\boldsymbol{x})} - \sup_{\boldsymbol{x}\in K} \frac{\tau}{z(\boldsymbol{x})}\right)_+ = \int_{\tau \sup_{\boldsymbol{x}\in K} 1/z(\boldsymbol{x})}^{\infty} \mathbb{P}\left(\sup_{\boldsymbol{x}\in K} \frac{V(\boldsymbol{x})}{z(\boldsymbol{x})} > u\right) \mathrm{d}u$$

becomes smaller for each positive function $z \in C(K)$ and for high thresholds $\tau > 0$.

3. Minimal log-Gaussian representations

To the best of the authors' knowledge, Problem 2.1 was first addressed by Matheron (1974), who introduced the notion of the minimal representation of an intrinsically stationary process. Starting from the fact that, given a specific centered Gaussian process W_0 on K with variogram γ, any other centered Gaussian process W on K with the same variogram is of the form $W(\boldsymbol{x}) = W_0(\boldsymbol{x}) + B$ for some square-integrable random variable B, he defined the minimal representation of W_0 as the process $W_{\min}(\boldsymbol{x}) = W_0(\boldsymbol{x}) + B_{\min}$ such that

$$\sup_{\boldsymbol{x}\in K} \operatorname{var}(W_{\min}(\boldsymbol{x})) = \inf_{B\in L^2(\Omega,\mathcal{A},\mathbb{P})} \sup_{\boldsymbol{x}\in K} \operatorname{var}(W_0(\boldsymbol{x}) + B),$$

i.e. the process whose covariance function $C_{\min}$ is the solution of Problem 2.1. Matheron (1974) further showed that, for any sample-continuous, intrinsically stationary Gaussian process, a unique minimal representation exists and has the form

$$W_{\min}(\boldsymbol{x}) = W_0(\boldsymbol{x}) - \int_K W_0(\tilde{\boldsymbol{x}})\lambda_{\min}(\mathrm{d}\tilde{\boldsymbol{x}}), \qquad \boldsymbol{x} \in K, \tag{3.1}$$

for some probability measure $\lambda_{\min}$ on K. The resulting covariance function $C_{\min}$ on $K \times K$ can be obtained from

$$\begin{aligned} 2C_{\min}(\boldsymbol{x}, \boldsymbol{y}) = {} & -\gamma(\boldsymbol{x} - \boldsymbol{y}) + \int_K \gamma(\boldsymbol{x} - \tilde{\boldsymbol{x}})\lambda_{\min}(\mathrm{d}\tilde{\boldsymbol{x}}) + \int_K \gamma(\boldsymbol{y} - \tilde{\boldsymbol{y}})\lambda_{\min}(\mathrm{d}\tilde{\boldsymbol{y}}) \\ & - \int_K \int_K \gamma(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{y}})\lambda_{\min}(\mathrm{d}\tilde{\boldsymbol{x}})\lambda_{\min}(\mathrm{d}\tilde{\boldsymbol{y}}). \end{aligned} \tag{3.2}$$

Note that the covariance and, consequently, the law of the minimal representation does not depend on the initially chosen random field W_0.

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

3.1. Minimal solution for convex variograms

If the variogram γ is convex and regular in the sense that, for two probability measures λ and μ on the compact domain K, the equality

$$\int_K \int_K \gamma(\boldsymbol{x}-\boldsymbol{y})(\lambda-\mu)(\mathrm{d}\boldsymbol{x})(\lambda-\mu)(\mathrm{d}\boldsymbol{y}) = 0$$

implies that $\lambda = \mu$, then Matheron (1974) characterizes the minimizing probability measure $\lambda_{\min}$ by the following two properties:

$$\mathrm{Supp}(\lambda_{\min}) \subset \mathrm{Ex}(K), \tag{3.3}$$

$$\int_K \gamma(\boldsymbol{x}-\tilde{\boldsymbol{y}})\lambda_{\min}(\mathrm{d}\tilde{\boldsymbol{y}}) \le \int_K \int_K \gamma(\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{y}})\lambda_{\min}(\mathrm{d}\tilde{\boldsymbol{x}})\lambda_{\min}(\mathrm{d}\tilde{\boldsymbol{y}}) \quad \text{for all } \boldsymbol{x} \in \mathrm{Ex}(K). \tag{3.4}$$

Here, $\mathrm{Ex}(K)$ denotes the set of *extremal points* of K, i.e. the elements of K that cannot be decomposed nontrivially as a convex combination of any two other points of K. For instance, the 2^d vertices $(\pm R_1, \pm R_2, \ldots, \pm R_d)$ of a d-dimensional hyperrectangle $K = \prod_{i=1}^d [-R_i, R_i]$ form its extremal points. In fact, hyperrectangles are the most natural simulation domains that we consider in practice and we will be mainly interested in this case. For a hyperrectangle $K = \prod_{i=1}^d [-R_i, R_i]$, it is also often convenient to label its vertex set $\mathrm{Ex}(K)$ by subsets A of $\{1, \ldots, d\}$ through

$$\mathrm{Ex}\left(\prod_{i=1}^d [-R_i, R_i]\right) = \{\boldsymbol{v}_A \colon A \subset \{1, \ldots, d\}\}, \qquad \boldsymbol{v}_A = (\sigma_i^A R_i)_{i=1}^d, \tag{3.5a}$$

$$\sigma_i^A = \begin{cases} +1, & i \in A, \\ -1, & i \notin A. \end{cases} \tag{3.5b}$$

When the variogram γ and the simulation domain K are sufficiently symmetric, the measure $\lambda_{\min}$ can be made even more explicit. To this end, we refer to the set of orthogonal transformations $\boldsymbol{M} \in O(\mathbb{R}^d)$, such that $\boldsymbol{M}A = \{\boldsymbol{M}\boldsymbol{a} \colon \boldsymbol{a} \in A\}$ coincides with A, as the *symmetry group* $\mathrm{Sym}(A)$ of a set $A \subset \mathbb{R}^d$.

Proposition 3.1. *Let W_0 be a sample-continuous, intrinsically stationary Gaussian process on $\mathbb{R}^d$ with convex variogram $\gamma(\boldsymbol{h}) = \psi(\|\boldsymbol{h}\|^2)$, $\boldsymbol{h} \in \mathbb{R}^d$. Let $K \subset \mathbb{R}^d$ be a compact domain, whose symmetry group* $\mathrm{Sym}(\mathrm{Ex}(K))$ *acts transitively on the set of extremal points* $\mathrm{Ex}(K)$. *Then the minimizing measure $\lambda_{\min}$ in the sense of (3.1) is the uniform distribution on* $\mathrm{Ex}(K)$.

Example 3.1. Let W_0 be a fractional Brownian sheet on $\mathbb{R}^d$ with variogram $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}/s\|^\alpha$ for some $\alpha \in [1, 2)$, $s > 0$. Let $K = \prod_{i=1}^d [-R_i, R_i]$ be a d-dimensional hyperrectangle, whose vertices $\boldsymbol{v}_A$ are labelled by subsets $A \subset \{1, \ldots, d\}$ as in (3.5). Then Proposition 3.1 applies and states that the process

$$W(\boldsymbol{x}) = W_0(\boldsymbol{x}) - \frac{1}{2^d} \sum_{A \subset \{1,\ldots,d\}} W_0(\boldsymbol{v}_A), \qquad \boldsymbol{x} \in K, \tag{3.6}$$

possesses the smallest maximal variance on K among the family of intrinsically stationary Gaussian processes with variogram γ.

Remark 3.1. Matheron (1974) provided the argument for (3.3) and (3.4) being necessary and sufficient for $\lambda_{\min}$ to be the minimal measure for compact and convex sets K. It is however easily checked that convexity of K can be dispensed with.

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

3.2. Reduction of maximal variance for nonconvex variograms

For nonconvex variograms, it is not so clear how to obtain the minimizing measure $\lambda_{\min}$ explicitly, not even for hyperrectangles. However, in many situations it is still possible to apply the same strategy as in Example 3.1 to substantially reduce the maximal variance. At least we show below that this is possible when the variogram γ can be represented as $\gamma(\boldsymbol{h}) = \psi(\|\boldsymbol{h}\|^2)$, $\boldsymbol{h} \in \mathbb{R}^d$, for a Bernstein function ψ.

There are several definitions of Bernstein functions, and various properties and examples have been summarized in the recent monograph Schilling *et al.* (2010). For us, it will be convenient to define a *Bernstein function* as a function $\psi : \mathbb{R}_+ \to \mathbb{R}$ on the positive real line $\mathbb{R}_+ = [0, \infty)$ that is bounded from below and *negative definite* in the sense that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \psi(s_i + s_j) a_j \leq 0 \tag{3.7}$$

for all finite systems $s_1, \dots, s_n \in \mathbb{R}_+$ and $a_1, \dots, a_n \in \mathbb{R}$ with $a_1 + \cdots + a_n = 0$; cf. Berg *et al.* (1984, pp. 113–114). Note that we deviate from most of the literature where continuity at 0 is additionally required. The following lemma shows that considering only variograms that can be represented by Bernstein functions is not a serious restriction and, secondly, that Bernstein functions obey certain monotonicity properties defined as follows. A function $f : \mathbb{R}_+ \to \mathbb{R}$ is *n-alternating* if it satisfies

$$\Delta_{s_1} \Delta_{s_2} \cdots \Delta_{s_n} f(s) = \sum_{A \subset \{1,\dots,n\}} (-1)^{|A|} f\left(s + \sum_{i \in A} s_i\right) \leq 0 \tag{3.8}$$

for $s, s_1, s_2, \dots, s_n \in \mathbb{R}_+$. In particular, f is 1-alternating if and only if it is nondecreasing.

Lemma 3.1. *Let $\psi : \mathbb{R}_+ \to \mathbb{R}$ be bounded from below. Then the following statements are equivalent.*

(i) *ψ is a Bernstein function.*

(ii) *$\psi(\|\boldsymbol{h}\|^2)$, $\boldsymbol{h} \in \mathbb{R}^d$, is negative definite in the sense of (2.3) for any dimension $d \in \mathbb{N}$.*

(iii) *ψ is n-alternating of any order $n \in \mathbb{N}$.*

In particular, the equivalence of (i) and (ii) means that ψ is a Bernstein function if and only if $\gamma(\boldsymbol{h}) = \psi(\|\boldsymbol{h}\|^2)$, $\boldsymbol{h} \in \mathbb{R}^d$, is a valid (not necessarily centred) variogram in any dimension $d \in \mathbb{N}$. Finally, this allows us to transfer the strategy from Example 3.1 to nonconvex variograms as follows.

Proposition 3.2. *Let W_0 be an intrinsically stationary Gaussian process on $\mathbb{R}^d$ with variogram $\gamma(\boldsymbol{h}) = \psi(\|\boldsymbol{h}\|^2)$, $\boldsymbol{h} \in \mathbb{R}^d$, for a Bernstein function ψ, and let $W_0(\mathbf{0}) = 0$ a.s. Let $K = \prod_{i=1}^{d} [-R_i, R_i]$ be a d-dimensional hyperrectangle, whose vertices $\boldsymbol{v}_A$ are labelled by subsets $A \subset \{1, \dots, d\}$ as in (3.5). Then the maximal variance of the modified process W in (3.6) is at least as small as the maximal variance of the original process W_0 on the domain K, i.e.*

$$\sup_{\boldsymbol{x} \in K} \operatorname{var}(W(\boldsymbol{x})) \leq \sup_{\boldsymbol{x} \in K} \operatorname{var}(W_0(\boldsymbol{x})).$$

Example 3.2. Proposition 3.2 applies to the situation of Example 3.1 with $\alpha \in [1, 2)$ replaced by *arbitrary* $\alpha \in (0, 2)$ and $W_0(\mathbf{0}) = 0$ a.s. In particular, the maximal variance on the hyperrectangle K can also be reduced for $\alpha \in (0, 1)$ by the same simple trick of subtracting the vertices of K with equal weights that we applied already for $\alpha \in [1, 2)$ in Example 3.1. Summarizing, we see that this trick *reduces* the maximal variance on K for *any* fractional Brownian sheet. For $\alpha \in [1, 2)$, it even *minimizes* this maximal variance; cf. Example 3.1.

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

Remark 3.2. (a) In fact, the proof of Proposition 3.2 only requires n-alternation of ψ for $n \leq 3$ and not necessarily that ψ is a Bernstein function.

(b) In Examples 3.1 and 3.2 and Proposition 3.2 the variogram $\gamma(\boldsymbol{h})$ can be replaced by $\tilde{\gamma}(\boldsymbol{h}) = \gamma(\boldsymbol{M}\boldsymbol{h})$, where $\boldsymbol{M}$ is any invertible symmetric $d \times d$ matrix. This is possible since such a transformation of $\mathbb{R}^d$ maps hyperrectangles that are centered at the origin again into hyperrectangles that are centered at the origin, whence Propositions 3.1 and 3.2 can be applied to the transformed processes $W_0(\boldsymbol{M}^{-1}\boldsymbol{x})$ and $W(\boldsymbol{M}^{-1}\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$, and their variogram $\tilde{\gamma}(\boldsymbol{M}^{-1}\boldsymbol{h}) = \gamma(\boldsymbol{h})$, $\boldsymbol{h} \in \mathbb{R}^d$.

3.3. Minimal K-stationary representations

There are some situations in which we can even compare the improvements of the variance reduction of Proposition 3.2 to the true minimal maximal variance on the simulation domain K. This is the case when the minimizing measure $\lambda_{\min}$ leads to a *K-stationary* solution, that is, when $C_{\min}(\boldsymbol{x}, \boldsymbol{y})$ in (3.2) depends only on $\boldsymbol{x} - \boldsymbol{y}$ for $\boldsymbol{x}, \boldsymbol{y} \in K$. In general, it is not clear that a given variogram $\gamma(\boldsymbol{x} - \boldsymbol{y})$, $\boldsymbol{x}, \boldsymbol{y} \in K$, possesses a K-stationary representation at all; cf. Remark 3.2.5 of Berg *et al.* (1984) for a counterexample. For fractional Brownian sheets on a d-dimensional Euclidean ball $B_R(\boldsymbol{0}) = \{\boldsymbol{h} \in \mathbb{R}^d : \|\boldsymbol{h}\| \leq R\}$ however, the following proposition, restating Gneiting (2000) and Matheron (1974), confirms their existence and makes their covariance explicit.

Proposition 3.3. (a) (Gneiting (2000)) *For $\alpha \in (0, 2)$, $s > 0$, and $R > 0$, the function*

$$C(\boldsymbol{x}, \boldsymbol{y}) = a - \frac{1}{2}\left\|\frac{\boldsymbol{x} - \boldsymbol{y}}{s}\right\|^{\alpha}, \qquad \boldsymbol{x}, \boldsymbol{y} \in B_R(\boldsymbol{0}),$$

is a covariance function if and only if $a \geq (A_{\alpha,d}/2)(R/s)^{\alpha}$, where

$$A_{\alpha,d} = \Gamma\left(\frac{2-\alpha}{2}\right)\Gamma\left(\frac{d+\alpha}{2}\right)\Gamma\left(\frac{d}{2}\right)^{-1}.$$

(b) (Matheron (1974)) *Choosing $a = (A_{\alpha,d}/2)(R/s)^{\alpha}$ minimizes the maximal variance*

$$\sup_{\boldsymbol{x} \in B_R(\boldsymbol{0})} C(\boldsymbol{x}, \boldsymbol{x}) = \sup_{\boldsymbol{x} \in B_R(\boldsymbol{0})} \operatorname{var}(W(\boldsymbol{x}))$$

among all intrinsically stationary Gaussian representations W on the domain $B_R(\boldsymbol{0})$ for the variogram $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}/s\|^{\alpha}$ if and only if $d = 1$ and $\alpha \in (0, 1]$.

Generally, the choice in (b) only minimizes the maximal variance among the *K-stationary* Gaussian representations of the variogram $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}/s\|^{\alpha}$ on the domain $K = B_R(\boldsymbol{0})$, not necessarily among *all* intrinsically stationary Gaussian representations of γ on $B_R(\boldsymbol{0})$.

Example 3.3. Let $\gamma(h) = |h/s|^{\alpha}$, $\alpha \in (0, 2)$, $s > 0$, be the family of fractional Brownian motion variograms in dimension 1. Collectively, the results from Examples 3.1 and 3.2 and Proposition 3.3(b) provide a full description of the intrinsically stationary Gaussian representations $W_{\min}$ of γ on the domain $[-R, R] \subset \mathbb{R}$ that minimize the maximal variance therein.

- For $\alpha \in [1, 2)$, it is given by the minimizing measure $\lambda_{\min} = \frac{1}{2}(\delta_{-R} + \delta_R)$ as described in Proposition 3.1 and Example 3.1. If W_0 is any intrinsically stationary Gaussian representation of γ, it can be obtained as the $\lambda_{\min}$-*modified process*

$$W_{\min}(x) = W_0(x) - \tfrac{1}{2}(W_0(-R) + W_0(R)), \qquad x \in [-R, R]. \tag{3.9}$$

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

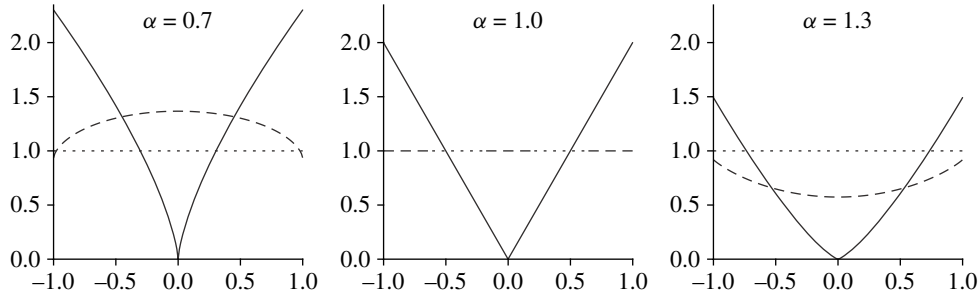


FIGURE 1: Variances $\sigma^2(t)$ of the Gaussian representations of the variogram $\gamma(h) = |h/s|^\alpha$ on the domain $K = [-1, 1]$. The plots show the variance for the original representation with $W_0(0) = 0$ (*solid line*), the minimal K-stationary representation (*dotted line*), and the λ-modified representation with $\lambda = \mathrm{Unif}(\mathrm{Ex}(K))$ (*dashed line*). For $\alpha = 1$, the last two coincide. The scale $s > 0$ is chosen such that the variance of the minimal K-stationary representation (*dotted line*) is normalized to 1.

- For $\alpha \in (0, 1]$, Proposition 3.3(b) tells us that the covariance function of $W_{\min}$ is

$$C_{\min}(x, y) = \frac{A_{\alpha,1} - |x - y|^\alpha}{2s^\alpha}, \qquad x, y \in [-R, R], \tag{3.10}$$

  which is the minimal K-*stationary* representation of γ on $K = B_R(0) = [-R, R]$.

This description was already obtained by Matheron (1974). In particular, the $\alpha = 1$ case in dimension 1 is very special in the sense that (i) d-dimensional balls and d-dimensional hyperrectangles are the same, and (ii) the functions $h \mapsto |h/s|^\alpha$, $s > 0$, are both concave and convex on $[0, \infty)$ for $\alpha = 1$. This leads to the situation that the minimizing representation can be obtained either way. The covariance of (3.9) is precisely (3.10) for $\alpha = 1$.

As an illustration of the different cases, in Figure 1 we present the variances of the original representation (with $W_0(0) = 0$), the λ-modified representation (with $\lambda = \frac{1}{2}(\delta_{-1} + \delta_1)$), and the minimal K-stationary representation of Gaussian random fields on $K = [-1, 1]$ with variogram $\gamma(h) = |h/s|^\alpha$ for $\alpha \in \{0.7, 1, 1.3\}$.

Example 3.4. Let $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}/s\|^\alpha$, $\alpha \in (0, 2)$, $s > 0$, be the family of variograms of fractional Brownian sheets in dimension $d \geq 2$, and consider the (hyper-)rectangular simulation domain $K = \prod_{i=1}^d [-R_i, R_i]$. Again we are interested in the intrinsically stationary Gaussian representation $W_{\min}$ of γ on the domain $K \subset \mathbb{R}^d$ that minimize the maximal variance therein.

- When $\alpha \in [1, 2)$, Proposition 3.1 and Example 3.1 again provide the minimizing measure

$$\lambda_{\min} = \frac{1}{2^d} \sum_{A \subset \{1,\dots,d\}} \delta_{\boldsymbol{v}_A}, \qquad \boldsymbol{v}_A = (\sigma_i^A R_i)_{i=1}^d, \qquad \sigma_i^A = \begin{cases} +1, & i \in A, \\ -1, & i \notin A, \end{cases}$$

  and $W_{\min}$ can be obtained as the λ-*modified* process from (3.6).

- When $\alpha \in (0, 1)$, however, we do not know the minimizing measure $\lambda_{\min}$ or the corresponding covariance $C_{\min}$, not even if we replace the domain K by the d-dimensional ball $B_R(\mathbf{0})$. At least, Proposition 3.3 identifies on $\widetilde{K} = B_R(\mathbf{0})$ the minimal $\widetilde{K}$-*stationary* Gaussian representations of the variogram γ for $\alpha \in (0, 2)$. For $R = (R_1^2 + \cdots + R_d^2)^{1/2}$, the domain $\widetilde{K} = B_R(\mathbf{0})$ is the smallest ball that contains K. Therefore, the covariance

$$C(\boldsymbol{x}, \boldsymbol{y}) = \frac{A_{\alpha,d} R^\alpha - \|(\boldsymbol{x} - \boldsymbol{y})\|^\alpha}{2s^\alpha}, \qquad \boldsymbol{x}, \boldsymbol{y} \in K, \tag{3.11}$$

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

also describes a K-stationary Gaussian representation of γ, but we do not know if it is minimal among K-stationary Gaussian representations of γ on K. However, in Figure 2 we illustrate that it is certainly not minimal among general Gaussian representations of γ. In each situation the λ-modified process (3.6) has an even smaller maximal variance on K than the K-stationary process derived from (3.11). While this is not a surprise for $\alpha = 1$ or $\alpha = 1.3$, where we know that (3.6) is minimal, it seems quite remarkable how much the maximal variance of (3.6) is reduced even in the $\alpha = 0.7$ case compared to the K-stationary process from (3.11). This is a major difference to the one-dimensional case that we considered in Example 3.3, where the K-stationary process provides the minimal representation.

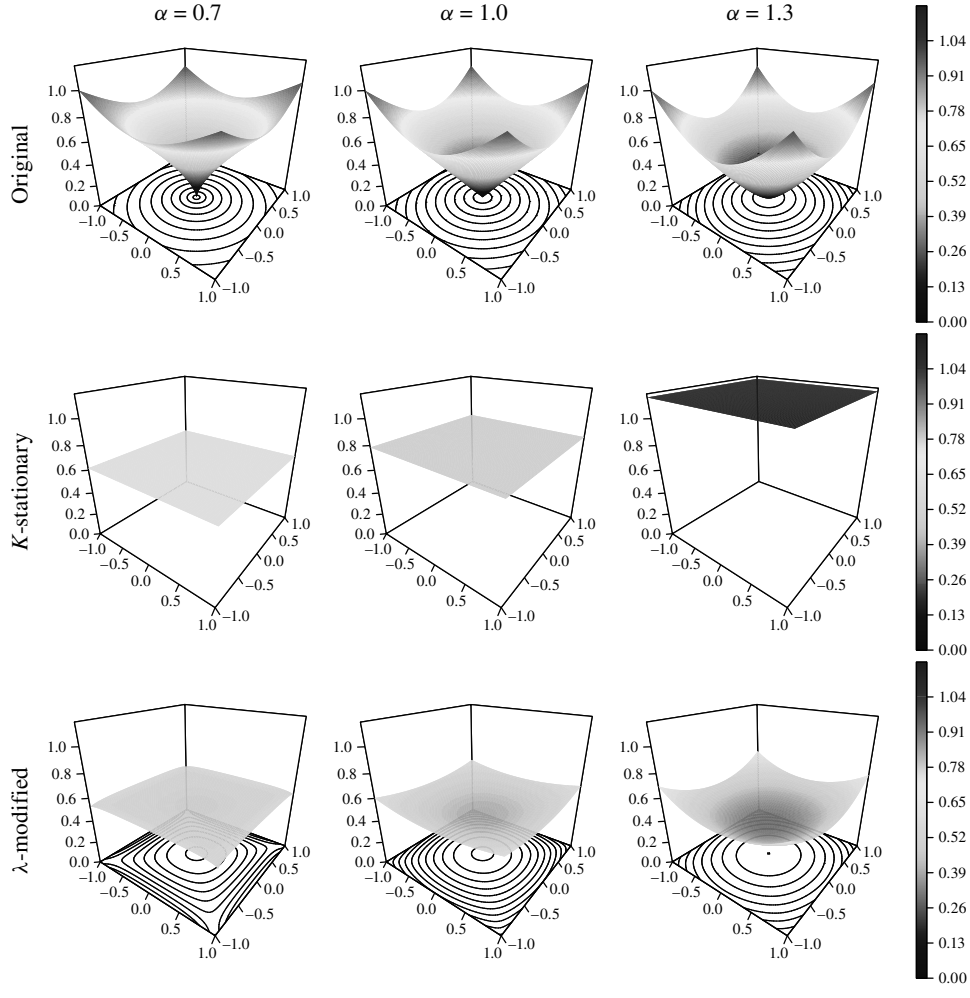


FIGURE 2: Variances $\sigma^2(t)$ of the Gaussian representations of the variogram $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}/\sqrt{2}\|^\alpha$ on the domain $K = [-1, 1]^2$ for $\alpha \in \{0.7, 1.0, 1.3\}$ (in order from left to right). The plots show the variance for the original representation with $W_0(\mathbf{0}) = 0$, the minimal K-stationary representation, and the λ-modified representation with $\lambda = \mathrm{Unif}(\mathrm{Ex}(K))$ (in order from top to bottom). Minimality of the K-stationary representation refers to the minimal ball $B_{\sqrt{2}}(\mathbf{0})$ containing K.

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

Remark 3.3. In the literature there are different terms related to the concept of *K-stationarity*. In e.g. Gneiting *et al.* (2001) a *locally equivalent stationary covariance* is defined as a covariance function C on $\mathbb{R}^d \times \mathbb{R}^d$ such that $C(\boldsymbol{x}, \boldsymbol{y})$ depends only on $\boldsymbol{x} - \boldsymbol{y}$ for *all* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, and

$$C(\boldsymbol{x}, \boldsymbol{y}) = C(\boldsymbol{0}) - \tfrac{1}{2}\gamma(\boldsymbol{x} - \boldsymbol{y}), \qquad \boldsymbol{x}, \boldsymbol{y} \in K. \tag{3.12}$$

In general, this assumption is stronger than that in our definition, as it additionally requires that the locally defined covariance function C on $K \times K$ can be extended to a stationary covariance function on $\mathbb{R}^d \times \mathbb{R}^d$. There are numerous examples of variograms γ and corresponding functions of the form (3.12) that are positive definite on a small (potentially discrete and finite) domain K for sufficiently large $C(\boldsymbol{0})$, but cannot be extended to a positive definite function on $\mathbb{R}^d \times \mathbb{R}^d$. In the case of an isotropic variogram and a domain of the type $K = B_R(\boldsymbol{0})$ for some $R > 0$, i.e. in the case considered in Proposition 3.3, however, such an extension is always possible by Rudin's theorem (Rudin (1970)). Thus, the terms *locally equivalent stationary covariance* and *covariance of a K-stationary representation* are equivalent in the case $K = B_R(\boldsymbol{0})$.

3.4. Discretization effects

In our examples above we always considered processes on *convex* d-dimensional domains K with *nontrivial interiors* (such as d-dimensional hyperrectangles or d-dimensional balls of some radii). Indeed, the regions of interest are usually of such kind. On the other hand, the simulation domain K in a computer experiment is necessarily a finite set of points $K = \{\boldsymbol{x}_i\}_{i=1,\dots,N}$ (such as a subgrid of a hyperrectangle) and the minimizing measure $\lambda_{\min}$ can depend on this choice of discretization. In the case that the variogram γ is convex, not much changes. Proposition 3.1 still applies and, in particular, if the discretization consists of a subgrid of some rectangular domain, the minimizing measure $\lambda_{\min}$ will be equal to the uniform distribution on the vertices of the chosen subgrid. The modified process W as in (3.6) still yields the minimal representation of the variogram γ.

It is again the case of a nonconvex variogram γ which is more intricate. First, we know that Proposition 3.2 also still applies and, typically, we can reduce the maximal variance by a significant amount by the same strategy as before, i.e. by considering the modification W in (3.6) instead of the original representation W_0. In the example of an α-fractional Brownian sheet we found this strategy useful when α is close to 1. Second, since we discretized our space, it is usually possible to solve

$$\int_K \gamma(\boldsymbol{x} - \boldsymbol{y})\lambda_0(\mathrm{d}\boldsymbol{y}) = 1, \qquad \boldsymbol{x} \in K,$$

explicitly for λ_0. If λ_0 is a nonnegative measure (and not a signed measure), normalizing λ_0 so that it becomes a probability measure will then provide the minimal (K-stationary) solution $\lambda_{\min}$; see Matheron (1974, p. 8). For an α-fractional Brownian sheet we found this strategy useful for $d = 1$ or α close to 0 and our numerical experiments led us to the following conjecture.

Conjecture 3.1. *Let* $K = \{\boldsymbol{x}_i\}_{i=1,\dots,N}$ *be a finite set of distinct points in* $\mathbb{R}^d$*, which contains at least two elements. Set* $\boldsymbol{\Gamma}_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^\alpha$*,* $i, j = 1, \dots, N$*,* $\alpha \in (0, 1]$*, and let* $\boldsymbol{e} \in \mathbb{R}^N$ *be the vector whose entries are all equal to* 1.

- *For* $d = 1$ *and* $\alpha \in (0, 1]$*, all entries of* $\boldsymbol{\Gamma}^{-1}\boldsymbol{e}$ *are nonnegative (we write* $\boldsymbol{\Gamma}^{-1}\boldsymbol{e} \geq 0$*).*
- *For* $d \geq 2$*, there exists* $\alpha_{\text{critical}} = \alpha_{\text{critical}}(K) \in (0, 1]$ *such that* $\boldsymbol{\Gamma}^{-1}\boldsymbol{e} \geq 0$ *for* $\alpha \in (0, \alpha_{\text{critical}}]$.
- *We have* $\alpha_{\text{critical}}(K) = 1$ *if and only if the points in K are collinear.*

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

For the remaining cases, the minimal solution $\lambda_{\min} = \sum_{i=1}^{N} \lambda_i \delta_{\boldsymbol{x}_i}$ can be obtained from the solution $\boldsymbol{\lambda} \in \mathbb{R}^N$ of the optimization problem

$$\min_{\boldsymbol{\lambda}} \max_{i=1}^{N} \tfrac{1}{2}(\boldsymbol{\lambda} - \boldsymbol{e}^i)^\top (-\boldsymbol{\Gamma})(\boldsymbol{\lambda} - \boldsymbol{e}^i) \quad \text{subject to} \quad \boldsymbol{e}^\top \boldsymbol{\lambda} = 1,\ \boldsymbol{\lambda} \geq \boldsymbol{0}, \tag{3.13}$$

where the vector $\boldsymbol{e}^i$ is the ith unit vector in $\mathbb{R}^N$, the matrix $\boldsymbol{\Gamma}$ has entries $\boldsymbol{\Gamma}_{ij} = \gamma(\boldsymbol{x}_i - \boldsymbol{x}_j)$, and $\boldsymbol{e} = \sum_{i=1}^{N} \boldsymbol{e}^i$. This is a nonlinear convex optimization problem that can be solved via standard optimization techniques.

4. Numerical results

Our numerical study focuses on Brown–Resnick processes on the hyperrectangle $K = [-1, 1]^d$ in dimensions $d \in \{1, 2\}$ with the underlying variograms belonging to the fractional Brownian sheet family $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}/s\|^\alpha$. It compares the performance of the threshold stopping algorithms that are based on

(i) the *original* Gaussian representation W_0 of the variogram γ with $W_0(\boldsymbol{0}) = 0$,

(ii) the minimal K*-stationary* Gaussian representation from (3.11),

(iii) the λ*-modified* Gaussian representation from (3.6).

In dimension 1 we fix the actual simulation domain as the grid $\{-1, -0.996, \ldots, 1\} \subset [-1, 1] = K$ that consists of 501 equally spaced points and vary both the smoothness parameter $\alpha \in (0, 2)$ and scale $s > 0$, as shown in Table 1. In each column the scale is chosen such that the minimal K-stationary Gaussian representation of the variogram (3.10) has the same variance σ_K^2 across the domain $K = [-1, 1]$, that is,

$$s = \left(\frac{1}{2\sqrt{\pi}\sigma_K^2} \Gamma\left(\frac{2-\alpha}{2}\right) \Gamma\left(\frac{1+\alpha}{2}\right) \right)^{1/\alpha}, \qquad \sigma_K^2 \in \{0.5, 1, 2\}. \tag{4.1}$$

In the $\alpha = 1$ case we could equivalently fix the scale as $s = 1$ and vary the domain $K = [-R, R]$ across $R = 1$, $R = 2$, and $R = 4$. In dimension 2 we fix the simulation domain as the grid $\{-1, -0.9, \ldots, 1\}^2 \subset [-1, 1]^2 = K$ that consists of 21×21 points and consider as a variogram the classical two-dimensional Brownian sheet variogram $\gamma(\boldsymbol{h}) = (2\sqrt{2}/\pi)\|\boldsymbol{h}\|$, where $\alpha = 1$. The scale $s = \pi/(2\sqrt{2})$ is chosen in such a way that the variance σ_K^2 of the associated K-stationary representation from (3.11) equals 1.

All simulation experiments are repeated 50 000 times. Based on the 50 000 simulated realizations of $Z^{(T)}$ and 50 000 independent realizations of the corresponding spectral process V,

TABLE 1: Simulation scenarios in dimension 1 on the interval $K = [-1, 1]$ for Brown–Resnick processes associated to the variogram $\gamma(h) = |h/s|^\alpha$; see Section 4. The scale s is obtained from (4.1).

α	Feature of variogram	Minimal Gaussian representation	s		
			Scale 1 $\sigma_K^2 = 0.5$	Scale 2 $\sigma_K^2 = 1.0$	Scale 3 $\sigma_K^2 = 2.0$
0.7	Concave	K-stationary	0.818	0.304	0.113
1.0	Linear	K-stationary / $\lambda = \mathrm{Unif}(\mathrm{Ex}(K))$	1.000	0.500	0.250
1.3	Convex	$\lambda = \mathrm{Unif}(\mathrm{Ex}(K))$	1.253	0.735	0.431

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

TABLE 2: Benchmark error terms $\widehat{\mathcal{P}}_{K,\tau}$ for threshold stopping algorithms based on Gaussian representations for the simulation scenarios of Table 1, where $\alpha \in \{0.7, 1.0, 1.3\}$; see Section 4.

Scenario	Original definition	K-stationary	$\lambda = \mathrm{Unif}(\mathrm{Ex}(K))$ modification
$\sigma_K^2 = 0.5, \alpha = 0.7$	0.19	**0.05**	0.10
$\sigma_K^2 = 0.5, \alpha = 1.0$	0.17	**0.10**	**0.10**
$\sigma_K^2 = 0.5, \alpha = 1.3$	0.16	0.17	**0.10**
$\sigma_K^2 = 1.0, \alpha = 0.7$	0.25	**0.04**	0.11
$\sigma_K^2 = 1.0, \alpha = 1.0$	0.22	**0.11**	**0.11**
$\sigma_K^2 = 1.0, \alpha = 1.3$	0.20	0.20	**0.10**
$\sigma_K^2 = 2.0, \alpha = 0.7$	0.32	**0.03**	0.10
$\sigma_K^2 = 2.0, \alpha = 1.0$	0.24	**0.11**	**0.09**
$\sigma_K^2 = 2.0, \alpha = 1.3$	0.21	0.24	**0.10**

TABLE 3: Benchmark error terms $\widehat{\mathcal{P}}_{K,\tau}$ for threshold stopping algorithms based on Gaussian representations for the simulation of Brown–Resnick processes on the square $K = [-1, 1]^2$ associated to the variogram $\gamma(\boldsymbol{h}) = (2\sqrt{2}/\pi)\|\boldsymbol{h}\|$.

Scenario	Original definition	K-stationary	$\lambda = \mathrm{Unif}(\mathrm{Ex}(K))$ modification
$\sigma_K^2 = 1.0, \alpha = 1.0$	0.23	0.30	**0.10**

we estimate the benchmark simulation error $\mathcal{P}_{K,\tau}$ according to (2.4). The estimated errors are reported in Table 2 for dimension 1 and Table 3 for dimension 2. The smallest error term in each scenario is always marked bold. To ensure consistency across the different scenarios, we first fix $\tau_{\lambda\text{-mod}}$ such that the resulting benchmark error term $\widehat{\mathcal{P}}_{K,\tau}$ assumes a value around 0.1, i.e. the third column in Tables 2 and 3 is always fixed (up to small deviations). To ensure a fair comparison of the different algorithms, we then choose the other thresholds τ_{orig} and $\tau_{K\text{-stat}}$ in such a way that the mean number $\mathbb{E}T_{K,\tau}$ of Gaussian processes that need to be simulated to obtain one single approximation to the Brown–Resnick process stays fixed within each scenario (up to relative deviations smaller than 1%).

The results in Tables 2 and 3 allow us to compare the performance of the different algorithms according to the error term $\widehat{\mathcal{P}}_{K,\tau}$ within a given scenario (α, s). As explained in Example 3.3, in dimension 1 we know for each scenario (α, s) which of the Gaussian representations of the variogram $\gamma(h) = |h/s|^\alpha$ leads to the minimal maximal variance across the simulation domain $[-1, 1]$. It depends only on the value of α as listed in Table 1. Indeed, the simulation results in Table 2 show that the smallest error term $\widehat{\mathcal{P}}_{K,\tau}$ is always attained by the algorithm that corresponds to the minimal representation. This confirms our previous theoretical considerations on the influence of the maximal variance of the spectral process on the performance of the threshold stopping algorithms. In dimension 2 we focus on the classical scenario of a two-dimensional Brownian sheet. As we can see from Table 3 and as anticipated, simulations via the threshold stopping algorithm based on the λ-modified representation exhibit smaller errors than simulations based on the K-stationary representation and simulations based on the original definition. This observation is well in line with our theoretical considerations,

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

since the maximal variance of the λ-modified representation is the smallest among those three (cf. Figure 2). It is even minimal in the sense of Problem 2.1; cf. Proposition 3.1. In the $\alpha = 1.3$ case in dimension 1 and the $\alpha = 1$ case in dimension 2, the threshold-stopping algorithm that is based on the K-stationary representation performs slightly worse than the original definition. Here, Figures 1 and 2 show (for both cases) that the variance of the original field is much smaller than the K-stationary variance on a large proportion of the relevant domain $[-1, 1]^d$, which can explain this phenomenon as a reasonable subasymptotic effect. Otherwise, the ranking of the threshold stopping algorithms according to the error terms even corresponds precisely to the ranking of the maximal variance on $[-1, 1]$ for each of the remaining scenarios in dimension 1.

In addition, we compare the performance of the above threshold stopping algorithms that are based on Gaussian processes to the performance of three other established algorithms for simulation of max-stable processes, namely,

(iv) the *random shift* approach (cf. Oesting *et al.* (2012)), a threshold stopping algorithm, where the original log-Gaussian spectral process $V^{(\mathrm{orig})}$ is additionally shifted uniformly across the (finite!) simulation domain $K = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$, i.e. its spectral process is

$$V^{(\mathrm{shift})}(\boldsymbol{x}) = V^{(\mathrm{orig})}(\boldsymbol{x} - \boldsymbol{S}), \qquad \boldsymbol{x} \in K = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\},$$

where $\boldsymbol{S}$ is uniformly distributed on $K = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ and independent of $V^{(\mathrm{orig})}$;

(v) the *Dieker–Mikosch* approach (cf. Dieker and Mikosch (2015) and Dombry *et al.* (2016)), another threshold stopping algorithm, which is based on the (sum-)normalized spectral process of the form

$$V(\boldsymbol{x}) = N \frac{\exp(W^{(\mathrm{orig})}(\boldsymbol{x} - \boldsymbol{S}))}{\sum_{k=1}^{N} \exp(W^{(\mathrm{orig})}(\boldsymbol{x} - \boldsymbol{x}_k))}, \qquad \boldsymbol{x} \in K = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}, \tag{4.2}$$

where $\boldsymbol{S}$ is again uniformly distributed on $K = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ and independent of $W^{(\mathrm{orig})}$;

(vi) the *extremal functions* approach (cf. Dombry *et al.* (2016)) with exact simulation taking place on a subset of N_{extrfun} prespecified equispaced locations in the simulation domain only.

To ensure a fair comparison, the thresholds for the random shift approach and the Dieker–Mikosch algorithm are again chosen in such a way that the mean number $\mathbb{E}T_{K,\tau}$ of Gaussian processes is fixed within each scenario (up to relative deviations smaller than 1%). Likewise, we fix the expected number of simulated Gaussian processes for the extremal functions approach, for which it corresponds to the expected number N_{extrfun} of locations for which exact simulation is ensured (cf. Dombry *et al.* (2016)). Since all scales in Table 2 seem to reproduce a similar qualitative behaviour, in this experiment we focus on the scenarios in Table 1 corresponding to scale 1 ($\sigma_K^2 = 0.5$) and explore a slightly larger range of $\alpha \in \{0.1, 0.4, \dots, 1.9\}$. The results are reported in Table 4 and plotted in Figure 3.

A first observation (that we found surprising at first sight) is the poor performance of the extremal functions approach and the strong performance of the Dieker–Mikosch approach in this experiment. We did not expect this previously, since, for *exact* simulation, the extremal functions approach always outperforms the Dieker–Mikosch algorithm in terms of the expected number of simulated Gaussian processes (see Dombry *et al.* (2016) and Oesting and Strokorb (2018)), i.e. one should prefer the extremal functions approach in this case. We conclude that allowing for a simulation error makes the Dieker–Mikosch approach competitive again with

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

TABLE 4: Benchmark error terms $\widehat{\mathcal{P}}_{K,\tau}$ for threshold stopping algorithms based on Gaussian representations and three further algorithms for the simulation scenarios of Table 1 with scale 1 ($\sigma_K^2 = 0.5$); see Figure 3 for an illustration.

Scenario	Original definition	K-stationary	$\lambda = \mathrm{Unif}(\mathrm{Ex}(K))$ modification	Random shift	Dieker–Mikosch	Extremal functions
$\sigma_K^2 = 0.5, \alpha = 0.1$	0.244	**0.007**	0.098	0.263	**0.005**	0.899
$\sigma_K^2 = 0.5, \alpha = 0.4$	0.204	**0.018**	0.095	0.245	**0.003**	0.822
$\sigma_K^2 = 0.5, \alpha = 0.7$	0.186	**0.051**	0.103	0.265	**0.006**	0.686
$\sigma_K^2 = 0.5, \alpha = 1.0$	0.168	**0.102**	**0.101**	0.236	**0.017**	0.521
$\sigma_K^2 = 0.5, \alpha = 1.3$	0.161	0.172	**0.102**	0.223	**0.037**	0.364
$\sigma_K^2 = 0.5, \alpha = 1.6$	0.130	0.241	**0.102**	0.171	**0.063**	0.231
$\sigma_K^2 = 0.5, \alpha = 1.9$	0.109	0.362	**0.101**	0.121	**0.094**	0.210

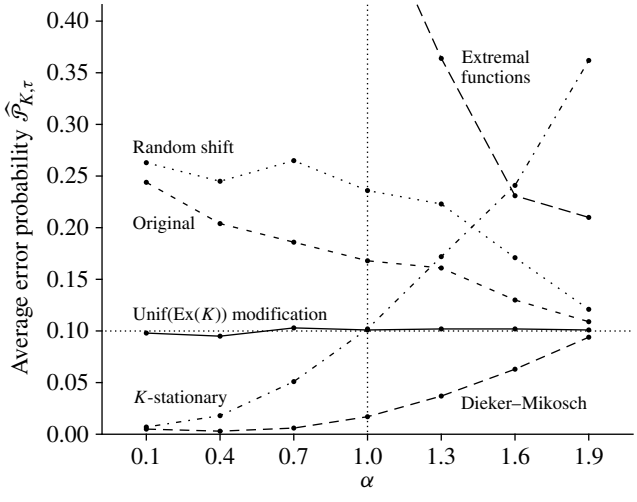


FIGURE 3: Benchmark error terms $\widehat{\mathcal{P}}_{K,\tau}$ for threshold stopping algorithms based on Gaussian representations ('Original', 'Unif(Ex(K)) modification', 'K-stationary') and three further algorithms ('Random shift', 'Dieker–Mikosch', 'Extremal functions') for the simulation scenarios of Table 1 with scale 1 ($\sigma_K^2 = 0.5$); see Table 4 for the estimated values.

other methods. Intuitively, this can be explained by its spectral functions (4.2) being bounded and probabilistically homogeneous in space, which makes its convergence rate quite fast at the beginning. While the random shift approach also enforces homogeneity of the spectral functions, this is however done in this case at the cost of increasing the maximal variance of the spectral functions, which explains why its performance is even worse than the performance of the original Gaussian representation. Generally, the three threshold-stopping algorithms that are based on *Gaussian* representation all lie between the random shift approach (apart from too high values of α, when the K-stationary approach is the worst) and the Dieker–Mikosch approach. Among them, we can also see the phase transition at $\alpha = 1$ between the K-stationary approach performing better for $\alpha < 1$ and the λ-modification taking over for $\alpha > 1$ quite well. As α gets close to 0, the K-stationary approach and the Dieker–Mikosch approach show a similar

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

error rate, whereas the λ-modification and the Dieker–Mikosch approach perform similarly as α gets close to 2.

5. Discussion

Efficient simulation of Brown–Resnick processes is an important task that is often needed to describe the extremal behaviour of spatial random fields. As exact simulation of such processes can be very time consuming, in particular when the simulation domain consists of a large number of points, it is often necessary to resort to the simulation of approximations of these processes. This can be done by either cutting an exact algorithm short or by running an inexact stopping algorithm. *A priori* it is unclear which algorithm will lead to the smallest error term for a fixed simulation time, or vice versa. Our focus here lies on the classical threshold-stopping algorithm (Schlather (2002)), where we consider different (log-)Gaussian spectral representations. Trying to mitigate the problem of the occurrence of excessively large variances of the spectral functions has led us to the following optimization problem.

Minimization Problem. *Among Gaussian processes with prescribed variogram γ, find a process W such that* $\sup_{x\in K} \operatorname{var}(W(x))$ *is minimal.*

This is a difficult mathematical problem that has been of independent interest for the case of K-stationary solutions in geostatistics (or locally equivalent stationary solutions, respectively); see, e.g. Gneiting *et al.* (2001), Gneiting (1999), (2000), Stein (2001), Chilès and Delfiner (2012), and Matheron (1973). To the best of the authors' knowledge, an explicit solution is known only in very few cases, such as for the variogram of a fractional Brownian motion on an interval (cf. Example 3.3 stating Matheron (1974)). Solutions also depend quite heavily on the geometry of the simulation domain K. Generally, Matheron's (1974) contribution does not seem to have received very much attention in the literature so far.

Here, we first make Matheron's description of a solution for convex variograms more explicit for symmetric domains (cf. Proposition 3.1), and then prove that the strategy that is employed for symmetric domains can even be applied to most practically relevant nonconvex variograms on hyperrectangular domains in order to achieve a substantial reduction of the maximal variance $\sup_{x\in K} \operatorname{var}(W(x))$, albeit not a minimal solution (cf. Proposition 3.2). We also consider discretization effects and conjecture for the fractional Brownian sheet family with variogram $\|h\|^\alpha$ the existence of a critical value of α, below of which the solution to the minimization problem is K-stationary (cf. Conjecture 3.1).

One of the nice features of the variance reduction is that it can be very easily implemented. Our simulation study confirms that the proposed modification of threshold-stopping algorithms that are based on a Gaussian representation can lead to significant improvements of the probability that no approximation error occurs while the expected running time is fixed. We expect similar performance improvements also for other features, e.g. when other types of errors are considered; cf. Oesting and Strokorb (2018). Further improvements are possible when taking into account discretization effects, i.e. modifying the Gaussian representation of the variogram γ according to (3.1) using the discrete minimizing measure (3.13). In our numerical study, we always considered a relatively dense grid, such that discretization effects (not reported) did not play a significant role. However, we conclude from further experiments (not reported) that they can become relevant for coarser designs.

A comparison with three other established simulation algorithms demonstrates that our approach can compete with the (potentially exact) extremal functions method (Dombry *et al.* (2016)) and always outperforms the random shift modification (Oesting *et al.* (2012)) due to

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

its increased variance. We were unable to detect a scenario in which our approach would outperform the Dieker–Mikosch algorithm (Dieker and Mikosch (2015)). The latter seems to converge relatively fast at the beginning stages of a simulation due to the boundedness and probabilistic homogeneity of the spectral functions in this threshold-stopping approach. However, when increasing the required accuracy, one should expect a phase transition, in which the Dieker–Mikosch approach performs better first, before the extremal functions approach takes over again. If we consider only the threshold-stopping approach based on *Gaussian* spectral representations, the proposed variance reduction leads to the best performances as expected. Further comparisons of different simulation algorithms are beyond the scope of this paper, but will be addressed in Oesting and Strokorb (2018).

Appendix A. Proofs

A.1. Proof for Section 2

Proposition A.1. *Let* $\{W_1(\boldsymbol{x}),\ \boldsymbol{x} \in K\}$ *and* $\{W_2(\boldsymbol{x}),\ \boldsymbol{x} \in K\}$ *be two centered Gaussian processes with a.s. bounded sample paths and bounded variance functions* $\sigma_i^2(\boldsymbol{x}) = \operatorname{var}(W_i(\boldsymbol{x}))$, $i = 1, 2$. *Furthermore, let* $\sigma_{1,\max}^2 < \sigma_{2,\max}^2$, *where* $\sigma_{i,\max}^2 = \sup_{\boldsymbol{x}\in K} \sigma_i^2(\boldsymbol{x}) < \infty$ *for* $i = 1, 2$. *Then, for any function* $y \in C(K)$,

$$\lim_{u\to\infty} \frac{\mathbb{P}(\sup_{\boldsymbol{x}\in K} W_1(\boldsymbol{x}) - \sigma_1^2(\boldsymbol{x})/2 - y(\boldsymbol{x}) > u)}{\mathbb{P}(\sup_{\boldsymbol{x}\in K} W_2(\boldsymbol{x}) - \sigma_2^2(\boldsymbol{x})/2 - y(\boldsymbol{x}) > u)} = 0.$$

Proof. By Proposition 1 of Debicki *et al.* (2010), for $i = 1, 2$, we have

$$-\log \mathbb{P}\Big(\sup_{\boldsymbol{x}\in K} W_i(\boldsymbol{x}) - \frac{\sigma_i^2(\boldsymbol{x})}{2} - y(\boldsymbol{x}) > u\Big) \sim -\log \mathbb{P}\Big(\sup_{\boldsymbol{x}\in K} W_i(\boldsymbol{x}) > u\Big) \sim \frac{u^2}{2\sigma_{i,\max}^2}$$

as $u \to \infty$. Consequently, for each $\varepsilon > 0$, there is some $u_i(\varepsilon) > 0$ such that, for all $u > u_i(\varepsilon)$,

$$\exp\Big(-\frac{u^2(1+\varepsilon)}{2\sigma_{i,\max}^2}\Big) \le \mathbb{P}\Big(\sup_{\boldsymbol{x}\in K} W_i(\boldsymbol{x}) - \frac{\sigma_i^2(\boldsymbol{x})}{2} - y(\boldsymbol{x}) > u\Big) \le \exp\Big(-\frac{u^2(1-\varepsilon)}{2\sigma_{i,\max}^2}\Big).$$

Now, let $\varepsilon > 0$ be sufficiently small such that $(1-\varepsilon)\sigma_{2,\max}^2 > (1+\varepsilon)\sigma_{1,\max}^2$. Then, for all $u > \max\{u_1(\varepsilon), u_2(\varepsilon)\}$, we have

$$\frac{\mathbb{P}(\sup_{\boldsymbol{x}\in K} W_1(\boldsymbol{x}) - \sigma_1^2(\boldsymbol{x})/2 - y(\boldsymbol{x}) > u)}{\mathbb{P}(\sup_{\boldsymbol{x}\in K} W_2(\boldsymbol{x}) - \sigma_2^2(\boldsymbol{x})/2 - y(\boldsymbol{x}) > u)} \le \exp\Big(-\frac{u^2}{2}\Big(\frac{1-\varepsilon}{\sigma_{1,\max}^2} - \frac{1+\varepsilon}{\sigma_{2,\max}^2}\Big)\Big),$$

which implies the assertion. □

A.2. Proofs for Section 3

Proof of Proposition 3.1. First note that $\operatorname{Ex}(K)$ is a (locally compact) homogeneous space with respect to the action of $\operatorname{Sym}(\operatorname{Ex}(K)) \subset O(\mathbb{R}^d)$, and so a unique normalized left-invariant Haar measure λ exists on $\operatorname{Ex}(K)$, which we call the uniform distribution; cf. e.g. Mardia and Khatri (1977) or Nachbin (1976). Its support $\operatorname{Supp}(\lambda) = \operatorname{Ex}(K)$ is necessarily a subset of $\operatorname{Ex}(K)$, which establishes (3.3) for λ. Furthermore, observe that the assignment

$$\boldsymbol{x} \mapsto \int_K \gamma(\boldsymbol{x} - \tilde{\boldsymbol{y}})\lambda(\mathrm{d}\tilde{\boldsymbol{y}}) = \int_{\operatorname{Ex}(K)} \gamma(\boldsymbol{x} - \tilde{\boldsymbol{y}})\lambda(\mathrm{d}\tilde{\boldsymbol{y}})$$

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

is also a convex function (since γ is convex). In particular, it attains its maximal value m on K for some $v \in \mathrm{Ex}(K)$. Since $\mathrm{Sym}(\mathrm{Ex}(K)) \subset O(\mathbb{R}^d)$ acts transitively on $\mathrm{Ex}(K)$, any element in $\mathrm{Ex}(K)$ can be represented as $\boldsymbol{M}\boldsymbol{v}$ for some $\boldsymbol{M} \in O(\mathbb{R}^d)$. Since $\gamma(\boldsymbol{h})$ depends only on $\|\boldsymbol{h}\|$, this gives

$$\int_{\mathrm{Ex}(K)} \gamma(\boldsymbol{M}\boldsymbol{v} - \tilde{\boldsymbol{y}})\lambda(\mathrm{d}\tilde{\boldsymbol{y}}) = \int_{\mathrm{Ex}(K)} \gamma(\boldsymbol{M}\boldsymbol{v} - \boldsymbol{M}\tilde{\boldsymbol{y}})\lambda(\mathrm{d}\tilde{\boldsymbol{y}}) = \int_{\mathrm{Ex}(K)} \gamma(\boldsymbol{v} - \tilde{\boldsymbol{y}})\lambda(\mathrm{d}\tilde{\boldsymbol{y}}).$$

So, in fact, all elements of $\mathrm{Ex}(K)$ attain this maximal value m, which implies that

$$\int_K \int_K \gamma(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{y}})\lambda(\mathrm{d}\tilde{\boldsymbol{x}})\lambda(\mathrm{d}\tilde{\boldsymbol{y}}) = \int_{\mathrm{Ex}(K)} \int_{\mathrm{Ex}(K)} \gamma(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{y}})\lambda(\mathrm{d}\tilde{\boldsymbol{x}})\lambda(\mathrm{d}\tilde{\boldsymbol{y}}) = \int_{\mathrm{Ex}(K)} m\lambda(\mathrm{d}\tilde{\boldsymbol{y}}) = m.$$

Finally, this gives

$$\int_K \gamma(\boldsymbol{x} - \tilde{\boldsymbol{y}})\lambda(\mathrm{d}\tilde{\boldsymbol{y}}) \le m \le \int_K \int_K \gamma(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{y}})\lambda(\mathrm{d}\tilde{\boldsymbol{x}})\lambda(\mathrm{d}\tilde{\boldsymbol{y}})$$

for all $\boldsymbol{x} \in \mathrm{Ex}(K)$, as desired (cf. Condition (3.4)). □

Proof of Lemma 3.1. First note that (i) is equivalent to ψ being negative definite in the sense of (3.7). The equivalence (i) ⇔ (ii) is then immediate from Corollary 5.1.8 of Berg *et al.* (1984, p. 150) and the fact that $(\mathbb{R}, \mathbb{R}_+, x^2)$ is a Schoenberg triple; cf. Example 5.1.3 of Berg *et al.* (1984, p. 146). Second, the equivalence (i) ⇔ (iii) follows from Corollary 4.6.8 of Berg *et al.* (1984, p. 133) and the 2-divisibility of the semigroup $(\mathbb{R}_+, +)$. □

Lemma A.1. *Let $\psi : \mathbb{R}_+ \to \mathbb{R}$ be n-alternating up to $n \le 2$ and let $0 \le x \le R$ as well as $0 \le x_i \le R_i$ for $i = 1, \dots, d$. Then the following inequalities hold for any $a \ge 0$.*

(a) $\psi(a + (R - x)^2) + \psi(a + (R + x)^2) \le \psi(a + R^2) + \psi(a + 3R^2)$.

(b) $$\sum_{A \subset \{1,\dots,d\}} \psi\Bigg(a + \sum_{i \in A}(x_i - R_i)^2 + \sum_{j \in A^{\mathrm{c}}}(x_j + R_j)^2\Bigg) \le \sum_{A \subset \{1,\dots,d\}} \psi\Bigg(a + 3\sum_{i \in A} R_i^2 + \sum_{j \in A^{\mathrm{c}}} R_j^2\Bigg).$$

Proof. (a) By the alternation properties of ψ, we have

$$\begin{aligned}
&\psi(a + (R - x)^2) + \psi(a + (R + x)^2) - \psi(a + R^2) - \psi(a + 3R^2) \\
&\quad \le \psi(a + (R - x)^2) + \psi(a + (R + x)^2) - \psi(a + R^2) - \psi(a + R^2 + 2x^2) \\
&\quad = \Delta_{2Rx - x^2}\Delta_{2Rx + x^2}\psi(a + (R - x)^2) \\
&\quad \le 0.
\end{aligned}$$

(b) The assertion follows by induction on d. For $d = 1$, it is evident from (a). For the step from d to $d + 1$, note that

$$\begin{aligned}
&\sum_{A \subset \{1,\dots,d+1\}} \psi\Bigg(a + \sum_{i \in A}(x_i - R_i)^2 + \sum_{j \in A^{\mathrm{c}}}(x_j + R_j)^2\Bigg) \\
&\quad = \sum_{B \subset \{1,\dots,d\}} \psi\Bigg(a + \sum_{i \in B}(x_i - R_i)^2 + \sum_{j \in B^{\mathrm{c}}}(x_j + R_j)^2 + (x_{d+1} - R_{d+1})^2\Bigg) \\
&\qquad + \sum_{B \subset \{1,\dots,d\}} \psi\Bigg(a + \sum_{i \in B}(x_i - R_i)^2 + \sum_{j \in B^{\mathrm{c}}}(x_j + R_j)^2 + (x_{d+1} + R_{d+1})^2\Bigg).
\end{aligned}$$

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

According to (a), the latter is less than or equal to

$$\sum_{B\subset\{1,\dots,d\}} \psi\Bigg(a+\sum_{i\in B}(x_i-R_i)^2+\sum_{j\in B^{\mathrm{c}}}(x_j+R_j)^2+3R_{d+1}^2\Bigg)$$
$$+\sum_{B\subset\{1,\dots,d\}} \psi\Bigg(a+\sum_{i\in B}(x_i-R_i)^2+\sum_{j\in B^{\mathrm{c}}}(x_j+R_j)^2+R_{d+1}^2\Bigg).$$

Now the induction hypothesis can be applied and gives that the latter is less than or equal to

$$\sum_{B\subset\{1,\dots,d\}} \psi\Bigg(a+3\sum_{i\in B}R_i^2+\sum_{j\in B^{\mathrm{c}}}R_j^2+R_{d+1}^2\Bigg)$$
$$+\sum_{B\subset\{1,\dots,d\}} \psi\Bigg(a+3\sum_{i\in B}R_i^2+\sum_{j\in B^{\mathrm{c}}}R_j^2+3R_{d+1}^2\Bigg)$$
$$=\sum_{A\subset\{1,\dots,d+1\}} \psi\Bigg(a+3\sum_{i\in A}R_i^2+\sum_{j\in A^{\mathrm{c}}}R_j^2\Bigg). \qquad \square$$

Lemma A.2. *Let $\psi:\mathbb{R}_+\to\mathbb{R}$ be n-alternating up to $n\le 3$ and let $\psi(0)=0$. Then we have, for $a,b\ge 0$,*

$$\psi(3a+b)-\tfrac{1}{2}\psi(4a)\le\psi(a+b).$$

Proof. The assertion follows from

$$\begin{aligned}
&-\psi(a+b)+\psi(3a+b)-\tfrac{1}{2}\psi(4a)\\
&\quad=[\psi(a)-\psi(3a)-\psi(a+b)+\psi(3a+b)]\\
&\qquad+\tfrac{1}{2}[\psi(0)-\psi(2a)-\psi(a)-\psi(a)+\psi(3a)+\psi(3a)+\psi(2a)-\psi(4a)]\\
&\quad=\Delta_{2a}\Delta_b\psi(a)+\tfrac{1}{2}\Delta_{2a}\Delta_a\Delta_a\psi(0)\\
&\quad\le 0. \qquad \square
\end{aligned}$$

Proof of Proposition 3.2. We need to show that

$$\operatorname{var}\Bigg(W_0(\boldsymbol{x})-\frac{1}{2^d}\sum_{A\subset\{1,\dots,d\}}W_0(\boldsymbol{v}_A)\Bigg)\le\sup_{\boldsymbol{x}\in K}\operatorname{var}(W_0(\boldsymbol{x})) \tag{A.1}$$

for all $\boldsymbol{x}\in K$. Since

$$\begin{aligned}
&2\operatorname{cov}(W_0(\boldsymbol{x})-W_0(\boldsymbol{y}),W_0(\boldsymbol{x}')-W_0(\boldsymbol{y}'))\\
&\qquad=\gamma(\boldsymbol{x}-\boldsymbol{y}')+\gamma(\boldsymbol{y}-\boldsymbol{x}')-\gamma(\boldsymbol{x}-\boldsymbol{x}')-\gamma(\boldsymbol{y}-\boldsymbol{y}'),
\end{aligned}$$

the left-hand side of (A.1) equals

$$\begin{aligned}
&\operatorname{var}\Bigg(W_0(\boldsymbol{x})-\frac{1}{2^d}\sum_{A\subset\{1,\dots,d\}}W_0(\boldsymbol{v}_A)\Bigg)\\
&\quad=\frac{1}{2^{2d}}\sum_{A\subset\{1,\dots,d\}}\sum_{B\subset\{1,\dots,d\}}\operatorname{cov}(W_0(\boldsymbol{x})-W_0(\boldsymbol{v}_A),W_0(\boldsymbol{x})-W_0(\boldsymbol{v}_B))\\
&\quad=\frac{1}{2^{2d+1}}\sum_{A\subset\{1,\dots,d\}}\sum_{B\subset\{1,\dots,d\}}\gamma(\boldsymbol{x}-\boldsymbol{v}_A)+\gamma(\boldsymbol{x}-\boldsymbol{v}_B)+\gamma(\boldsymbol{v}_A-\boldsymbol{v}_B)\\
&\quad=\frac{1}{2^d}\sum_{A\subset\{1,\dots,d\}}\gamma(\boldsymbol{x}-\boldsymbol{v}_A)-\frac{1}{2}\gamma(\boldsymbol{v}_\varnothing-\boldsymbol{v}_A).
\end{aligned}$$

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

On the other hand, since $W_0(\mathbf{0}) = 0$ and $\gamma(\boldsymbol{h}) = \psi(\|\boldsymbol{h}\|^2)$ for a monotonously increasing function ψ, the right-hand side of (A.1) coincides with $\gamma(\boldsymbol{v}_A)$ for any vertex $\boldsymbol{v}_A$ of K. Hence, we need to show that, for all $\boldsymbol{x} \in K$,

$$\frac{1}{2^d} \sum_{A \subset \{1,\dots,d\}} \gamma(\boldsymbol{x} - \boldsymbol{v}_A) - \frac{1}{2}\gamma(\boldsymbol{v}_\varnothing - \boldsymbol{v}_A) \le \gamma(\boldsymbol{v}_\varnothing),$$

which is equivalent to

$$\frac{1}{2^d} \sum_{A \subset \{1,\dots,d\}} \psi\left(\sum_{i \in A}(x_i - R_i)^2 + \sum_{j \in A^c}(x_j + R_j)^2\right) - \frac{1}{2}\psi\left(4\sum_{i \in A} R_i^2\right) \le \psi(R_1^2 + \cdots + R_d^2).$$

By Lemma A.1, it suffices to show that

$$\begin{aligned}\frac{1}{2^d} \sum_{A \subset \{1,\dots,d\}} \psi\left(3\sum_{i \in A} R_i^2 + \sum_{j \in A^c} R_j^2\right) - \frac{1}{2}\psi\left(4\sum_{i \in A} R_i^2\right) \\ \le \frac{1}{2^d} \sum_{A \subset \{1,\dots,d\}} \psi\left(\sum_{i \in A} R_i^2 + \sum_{j \in A^c} R_j^2\right).\end{aligned}$$

The latter follows from Lemma A.2. □

Acknowledgements

The authors would like to thank an anonymous referee for a lot of well-conceived feedback on their work, including the suggestion to include the Dieker–Mikosch algorithm in our simulation study. The authors would also like to thank Stilian Stoev and Holger Drees for their very helpful comments during EVA 2017 in Delft, which led to a rethinking of the error assessment. The sole responsibility for all directions taken lies, of course, with the authors.

References

Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*. Springer, New York.

Asadi, P., Davison, A. C. and Engelke, S. (2015). Extremes on river networks. *Ann. Appl. Statist.* **9,** 2023–2050.

Berg, C., Christensen, J. P. R. and Ressel, P. (1984). *Harmonic Analysis on Semigroups* (Graduate Texts Math. **100**). Springer, New York.

Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Prob.* **14,** 732–739.

Buhl, S. and Klüppelberg, C. (2016). Anisotropic Brown-Resnick space-time processes: estimation and model assessment. *Extremes* **19,** 627–660.

Chilès, J. P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*, 2nd edn. John Wiley, Hoboken, NJ.

Davison, A. C., Huser, R. and Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Math. Geosci.* **45,** 511–529.

Dębicki, K., Kosínski, K. M., Mandjes, M. and Rolski, T. (2010). Extremes of multidimensional Gaussian processes. *Stoch. Process. Appl.* **120,** 2289–2301.

De Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Prob.* **12,** 1194–1204.

Dieker, A. B. and Mikosch, T. (2015). Exact simulation of Brown-Resnick random fields at a finite number of locations. *Extremes* **18,** 301–314.

Dombry, C., Engelke, S. and Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika* **103,** 303–317.

Einmahl, J. H. J., Kiriliouk, A., Krajina, A. and Segers, J. (2016). An M-estimator of spatial tail dependence. *J. R. Statist. Soc. B* **78,** 275–298.

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press

Engelke, S., Malinowski, A., Kabluchko, Z. and Schlather, M. (2015). Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *J. R. Statist. Soc. B* **77,** 239–265.

Gaume, J. *et al.* (2013). Mapping extreme snowfalls in the French Alps using max-stable processes. *Water Resources Res.* **49,** 1079–1098.

Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (eds) (2010). *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL.

Giné, E., Hahn, M. G. and Vatan, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Prob. Theory Relat. Fields* **87,** 139–165.

Gneiting, T. (1999). Isotropic correlation functions on d-dimensional balls. *Adv. Appl. Prob.* **31,** 625–631.

Gneiting, T. (2000). Addendum to 'Isotropic correlation functions on d-dimensional balls'. *Adv. Appl. Prob.* **32,** 960–961.

Gneiting, T., Sasvári, Z. and Schlather, M. (2001). Analogies and correspondences between variograms and covariance functions. *Adv. Appl. Prob.* **33,** 617–630.

Kabluchko, Z. (2011). Extremes of independent Gaussian processes. *Extremes* **14,** 285–310.

Kabluchko, Z., Schlather, M. and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Prob.* **37,** 2042–2065.

Liu, Z., Blanchet, J. H., Dieker, A. B. and Mikosch, T. (2018). On logarithmically optimal exact simulation of max-stable and related random fields on a compact set. To appear in *Bernoulli*.

Mardia, K. V. and Khatri, C. G. (1977). Uniform distribution on a Stiefel manifold. *J. Multivariate Anal.* **7,** 468–473.

Matheron, G. (1973). The intrinsic random functions and their applications. *Adv. Appl. Prob.* **5,** 439–468.

Matheron, G. (1974). Représentations stationnaires et représentations minimales pour les F.A.I.-k. Tech. Rep. (Note Géostatistique 125). Centre de Géostatistique, École des Mines de Paris, France.

Nachbin, L. (1976). *The Haar Integral*. Robert E. Krieger Publishing, Huntington, NY.

Oesting, M. and Stein, A. (2018). Spatial modeling of drought events using max-stable processes. *Stoch. Environ. Res. Risk Assess.* **32,** 63–81.

Oesting, M. and Strokorb, K. (2018). A comparative tour through the simulation algorithms for max-stable processes. Preprint. Available at http://arxiv.org/abs/1809.09042v1.

Oesting, M., Kabluchko, Z. and Schlather, M. (2012). Simulation of Brown-Resnick processes. *Extremes* **15,** 89–107.

Oesting, M., Schlather, M. and Friederichs, P. (2017). Statistical post-processing of forecasts for extremes using bivariate Brown-Resnick processes with an application to wind gusts. *Extremes* **20,** 309–332.

Oesting, M., Schlather, M. and Zhou, C. (2018). Exact and fast simulation of max-stable processes on a compact set using the normalized spectral representation. *Bernoulli* **24,** 1497–1530.

Rudin, W. (1970). An extension theorem for positive-definite functions. *Duke Math. J* **37,** 49–53.

Sang, H. and Genton, M. G. (2014). Tapered composite likelihood for spatial max-stable models. *Spat. Statist.* **8,** 86–103.

Schilling, R. L., Song, R. and Vondraček, Z. (2010). *Bernstein Functions: Theory and Applications* (De Gruyter Stud. Math. **37**). Walter de Gruyter, Berlin.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes* **5,** 33–44.

Stein, M. L. (2001). Local stationarity and simulation of self-affine intrinsic random functions. *IEEE Trans. Inform. Theory* **47,** 1385–1390.

Strokorb, K. (2013). Characterization and construction of max-stable processes. Doctoral Thesis. Georg-August-Universität Göttingen.

Thibaud, E. *et al.* (2016). Bayesian inference for the Brown-Resnick process, with an application to extreme low temperatures. *Ann. Appl. Statist.* **10,** 2303–2324.

https://doi.org/10.1017/apr.2018.54 Published online by Cambridge University Press