Quantifying uncertainties on excursion sets under a Gaussian random field prior

Dario Azzimonti, Julien Bect, Clément Chevalier, David Ginsbourger

Quantifying uncertainties on excursion sets under a Gaussian random field prior

Dario Azzimonti [*], Julien Bect [†], Clément Chevalier [‡], David Ginsbourger[*]

Abstract

We focus on the problem of estimating and quantifying uncertainties on the excursion set of a function under a limited evaluation budget. We adopt a Bayesian approach where the objective function is assumed to be a realization of a random field, typically assumed Gaussian. In this setting, the posterior distribution on the objective function gives rise to a posterior distribution of excursion sets. Several approaches exist to summarize the distribution of the excursion sets based on random closed set theory. While the recently proposed Vorob'ev approach leads to analytically tractable expectations, further notions of variability require Monte Carlo estimators relying on Gaussian random field conditional simulations. In the present work we propose a method to choose simulation points and obtain realizations of the conditional field at fine designs through affine predictors. The points are chosen optimally in the sense that they minimize the expected distance in measure between the posterior excursion set and its reconstruction. The proposed method reduces the computational costs due to simulations and enables the prediction of realizations on fine designs even in large dimensions. We apply this reconstruction approach to obtain realizations of an excursion set on a fine grid which allow us to give a new measure of uncertainty based on the distance transform of the excursion set. Finally we present an application of the method where the distribution of the volume of excursion is estimated in a six-dimensional example.

Keywords: Set estimation, distance transform, Gaussian processes, conditional simulations

1 Introduction

In a number of application fields where mathematical models are used to predict the behavior of some parametric system of interest, practitioners not only wish to get the response for a given set of inputs (forward problem) but are

[*]Department of Mathematics and Statistics, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland

[†]Supélec, Gif-sur-Yvette, France

[‡]Institute of Mathematics, University of Zurich, Switzerland

interested in recovering the set of inputs values leading to a prescribed value or range of values for the output (inverse problem). Such problems are especially common in cases where the response is a scalar quantifying the degree of danger or abnormality of a system, or equivalently represents a score measuring some performance or pay-off. Examples include applications in reliability engineering, where the focus is often put on describing the set of parameter configurations leading to an unsafe design (mechanical engineering [15], [4], nuclear criticality [8], etc.), but also in natural sciences, where conditions leading to dangerous phenomena in climatological [17] or geophysical [3] settings are of crucial interest.

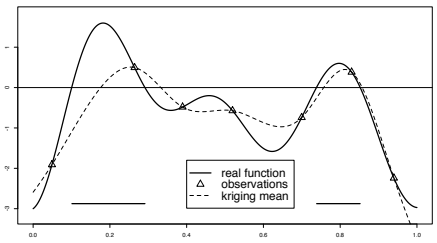
When tackling inverse problems involving costly-to-evaluate forward models, the number of model evaluations affordable during a study is typically a limiting factor. As a consequence, a systematic exploration of the input space, e.g. on a fine grid, is generally out of reach, even in small dimensions. Therefore reconstructions of the sets of interest have to be performed based on a scarce number of evaluations, thus implying some uncertainty.

Various methods are available to interpolate or approximate an objective function relying on a sample of pointwise evaluations, including polynomial approximations, splines, neural nets, and more. Here we mainly focus on the so-called *Gaussian Random Field* modeling approach.

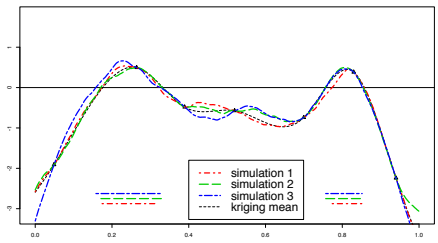
Gaussian Random Field (GRF) models have become very popular in engineering and further application areas to approximate, or *predict*, expensive-to-evaluate functions relying on a drastically limited number of observations (see, e.g., [38], [30], [5] and references therein). A major advantage of GRF models over deterministic approximation models is to deliver a posterior probability distribution on functions, not only enabling to get predictions of the objective function at any point, but also to quantify uncertainties on the associated predictions. Furthermore, posterior distributions of quantities non-linearly involving the objective function may be estimated by conditional simulations.

The idea of appealing to conditional simulation in the context of set estimation has already been introduced in various contexts (see, e.g., [25], [12]). Instead of having a single estimate of the excursion set like in most *set estimation* approaches (see, e.g., [14], [20], [29] and references therein), it is possible to get a distribution of sets. For example, Figure 1 shows some realizations of the excursion set (above the threshold $t = 0$) obtained by simulating the Gaussian random field Z conditional on few observations of the true function f at locations $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. A natural question arising in practice is how to summarize this distribution by appealing to simple concepts, analogues to notions of expectation and variance (or location and scatter) in the framework of random variables and vectors. For example the notions of Vorob'ev expectation and Vorob'ev deviation have been recently revisited in [9] in the context of excursion set estimation and uncertainty quantification with GRF models.

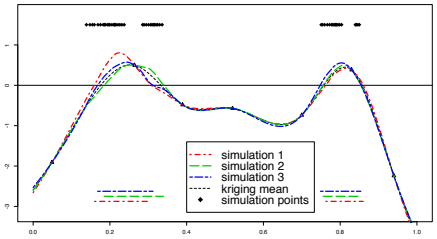
One of the key contributions of the present paper is a method to generate conditional realizations of the random excursion set based on simulations of the underlying Gaussian random field at few points. By contrast, in the literature, Monte Carlo simulations of excursion sets are often obtained by sim-



(a) Real function, kriging mean and true excursion set.



(b) 3 realizations of the conditional GRF and the associated excursion set. The realizations are obtained with simulations at 1000 points in $[0, 1]$.



(c) 3 realizations of the conditional random field and the associated random set generated by simulating at 30 optimally-chosen points and predicting the field in the 1000 grid points design.

Figure 1: Example of a Gaussian random field model with few realizations of a function defined on $[0, 1]$.

ulating the underlying field at space filling designs, as shown in Figure 1b. While this approach is straightforward to implement, it might be too cumbersome when fine designs are needed, especially in high dimensions. The proposed approach reduces the simulation costs by choosing few appropriate points $E_m = \{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ where to simulate the field and by approximating the field on the full design with a suitable affine predictor. Approximations of the excursion set realizations are obtained as excursion sets of the approximate field. Figure 1c shows realizations of the excursion set approximated with the best linear unbiased predictor from simulations at $m = 30$ points, for the example introduced in Figure 1. Simulation points are chosen in an optimal way in the sense that they minimize a specific distance between the reconstructed random set and the true random set. With this method it is possible to obtain arbitrarily fine approximations of the excursion set realizations while retaining control on how close those approximations are to the true random set distribution.

The paper is divided into 6 Sections. In Section 2 we introduce the framework and the fundamental definitions needed for our method. In Section 3 we give an explicit formula to compute the distance between the reconstructed random excursion set and the true random excursion set. In this section we also present a result on the consistency of the method when a dense sequence points is considered as simulation points, the proof are in appendix. Section 4 explains the computational aspects and introduces two algorithms to calculate the optimized points. In this section we also discuss the limits and the advantages of two algorithms.

Sections 5 and 6 present two applications that benefit from the optimized selection of simulation points. In the first application we show that the distance average of a set (see, e.g., [2]) over a two-dimensional grid can be computed accurately by simulating the field at few simulation points and by predicting the field over an arbitrary design. This approach allows us to compute the distance transform variability, an uncertainty quantification estimate based on the distance average. Since the distance transform variability heavily relies on conditional simulations, it has not been used before as uncertainty quantification technique. In the second application we apply the proposed method to the problem of estimating the distribution of the excursion volume in a six-dimensional example. The excursion volumes are generally estimated with Monte Carlo methods because the analytical distribution is unknown (see, e.g., [37], [1]). We empirically show that our method reduces the computational times for generating Monte Carlo estimates while preserving the accuracy.

2 Preliminaries

In this section we recall two definitions coming from the theory of random closed sets that help us to define an expected excursion set. The main focus here is on a real-valued continuous objective function $f : D \subset \mathbb{R}^d \longrightarrow \mathbb{R}$ where $d \geq 1$ and D is a compact of $\mathbb{R}^d$. f is modeled by a Gaussian random field with continuous paths, $Z = (Z_{\mathbf{x}})_{\mathbf{x} \in D}$, whose mean function and covariance kernel are denoted by

m and k. The range of critical response values of interest and the corresponding excursion set are denoted by $T \in \mathcal{B}(\mathbb{R})$, a measurable element of the Borel σ-algebra of $\mathbb{R}$, and $\Gamma^\star = f^{-1}(T) = \{\mathbf{x} \in D : f(\mathbf{x}) \in T\}$. In most applications, T is a closed set in $\mathbb{R}$ of the form $[t, \infty)$ for some $t \in \mathbb{R}$. Here we solely need to assume that T is closed in $\mathbb{R}$, but we will stick to the settings where $T = [t, \infty)$, for simplicity. Generalizations to unions of intervals are straightforward. Under such assumptions, the excursion set $\Gamma^\star$ is closed in D as pre-image of a closed set by a continuous function. Similarly, $\Gamma = \{\mathbf{x} \in D : Z(\mathbf{x}) \in T\}$ defines a random closed set. In the following, we will appeal to a number of concepts from the theory of random closed sets [27].

2.1 Vorob'ev approach

The notion of *distance in measure* plays a key role in the proposed approach. Let μ be a Borel measure on the Borel σ-algebra $\mathcal{B}(D)$ and $S_1, S_2 \in \mathcal{B}(D)$. Their distance in measure (with respect to μ) is defined as $\mu(S_1 \Delta S_2)$, where $S_1 \Delta S_2 = (S_1 \cap S_2^c) \cup (S_2 \cap S_1^c)$ is the symmetrical difference between S_1 and S_2. Similarly, for two random sets Γ_1 and Γ_2 and μ as before, one can define a distance as follows.

Definition 1 (Expected distance in measure)**.** *The expected distance in measure between two random sets Γ_1, Γ_2 with respect to a Borel measure μ is defined as*

$$d_\mu(\Gamma_1, \Gamma_2) = \mathbb{E}[\mu(\Gamma_1 \Delta \Gamma_2)] \tag{1}$$

Several notions of expectation have been proposed for random closed sets. In particular, the Vorob'ev expectation is related to the notion of expected distance in measure. Consider the coverage function of a random closed set Γ, $p_\Gamma : \mathbf{x} \in D \longrightarrow [0, 1]$ defined as $p_\Gamma(\mathbf{x}) = P(\mathbf{x} \in \Gamma)$. The Vorob'ev expectation Q_α of a random set Γ is defined as the α level set of its coverage function, i.e. $Q_\alpha = \{\mathbf{x} \in D : p_\Gamma(\mathbf{x}) \geq \alpha\}$, where the level α is determined so as to satisfy $\mu(Q_\beta) \leq \mathbb{E}[\mu(\Gamma)] \leq \mu(Q_\alpha)$ for all $\beta > \alpha$. It is a well-known fact [27] that, in the particular case $\mathbb{E}[\mu(\Gamma)] = \mu(Q_\alpha)$, the Vorob'ev expectation minimizes the distance in measure to Γ among all measurable (deterministic) sets M such that $\mu(M) = \mathbb{E}[\mu(\Gamma)]$. In the following we review another notion of expectation for a random closed set: the distance average and its related notion of variability.

2.2 Distance average approach

Given a *distance function* $d : (\mathbf{x}, S) \in D \times \mathcal{F}' \longrightarrow \mathbb{R}$ where $\mathcal{F}'$ is the space of all non-empty closed sets (see [27] pp. 179–180 for details) and assuming that $d(\mathbf{x}, \Gamma)$ has finite expectation for all $\mathbf{x} \in D$, one defines the *mean distance function* as $\bar{d} : \mathbf{x} \in D \longrightarrow \mathbb{E}[d(\mathbf{x}, \Gamma)]$. Assuming further that some metric δ (e.g., the L^2 distance) is available over an appropriate functional space, the *distance average* of Γ [27] is defined as the set that has the closest (with respect to the metric δ) distance function to $\bar{d}$.

Definition 2 (Distance average and distance function variability)**.** *Let $\bar{u}$ be the value of $u \in \mathbb{R}$ that minimizes the δ-distance $\delta(d(\cdot, \{\bar{d} \leq u\}), \bar{d})$ between the distance function of $\{\bar{d} \leq u\}$ and the mean distance function of Γ. If $\delta(d(\cdot, \{\bar{d} \leq u\}), \bar{d})$ achieves its minimum in several points we assume $\bar{u}$ to be their infimum. The set*

$$\mathbb{E}_{DA}(\Gamma) = \{\mathbf{x} \in D : \bar{d}(\mathbf{x}) \leq \bar{u}\} \tag{2}$$

is said to be the distance average of Γ with respect to δ. In addition, we define the distance function variability of Γ with respect to δ as $\mathrm{DTV}(\Gamma) = \mathbb{E}[\delta(\bar{d}, d(\cdot, \Gamma))]$.

These notions will be at the heart of the application section, where a method is proposed for estimating discrete counterparts of $\mathbb{E}_{\mathrm{DA}}(\Gamma)$ and $\mathrm{DTV}(\Gamma)$ relying on approximate GRF simulations. Taking a standard matrix decomposition approach for GRF simulations, a straightforward way for simulating Γ is to simulate Z at a fine design, e.g. a grid in moderate dimensions, $G = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \subset D$ with large $r \in \mathbb{N}$, and then to represent Γ as $\{\mathbf{u} \in G : Z_{\mathbf{u}} \in T\}$. A drawback of this procedure, however, is that it may become impractical for a high resolution r, as the covariance matrix involved may rapidly become close to singular and also cumbersome if not impossible to store.

3 Main results

The proposed approach consists in replacing conditional GRF simulations at the finer design G by approximate simulations that rely on a smaller simulation design $E_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$, with $m \ll r$, and to use these approximate simulations as basis for quantifying uncertainties on Γ. Even though such approach might seem somehow heuristic at first, it is actually possible to control the effect of the approximation on the end result, as we show in this section. Let us first expose the proposed workflow.

3.1 A Monte-Carlo approach with predicted conditional simulations

Performing Monte Carlo simulations of Γ (or of its trace on a fine design G) necessitates to simulate Z. Here we propose to replace Z by a simpler random field denoted by $\widetilde{Z}$, which simulations at any design should remain at an affordable cost. In particular, we aim at constructing $\widetilde{Z}$ in such a way that the associated $\widetilde{\Gamma}$ is as close as possible to Γ in expected distance in measure. Let us assume that the random field Z has been evaluated at $n \geq 0$ locations $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset D$. Consider $1 \leq m \leq r$ points $E_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\} \subset D$ and denote by $Z(E_m) = (Z_{\mathbf{e}_1}, \ldots, Z_{\mathbf{e}_m})^T$ the random vector of values of Z at E_m. The essence of the proposed approach is to appeal to affine predictors of Z, i.e. to consider $\widetilde{Z}$ of the form

$$\widetilde{Z}(\mathbf{x}) = a(\mathbf{x}) + \mathbf{b}^T(\mathbf{x}) Z(E_m) \qquad (\mathbf{x} \in D), \tag{3}$$

where $a : D \longrightarrow \mathbb{R}$ is a trend function and $\mathbf{b} : D \longrightarrow \mathbb{R}^m$ is a vector-valued function of deterministic weights. Note that usual kriging predictors are particular cases of Equation (3) with adequate choices of the functions a and $\mathbf{b}$, see for example [13] for an extensive review. Re-interpolating conditional simulations by kriging is an idea that has been already proposed in different contexts, notably by [28] in the context of Bayesian uncertainty analysis for complex computer codes. However, to the best of our knowledge, the derivation of optimal criteria for choosing simulation points has not been addressed until now, be it for excursion set estimation or for further purposes. Here we focus exclusively on excursion set estimation and we take as criterion the expected distance in measure between Γ and $\widetilde{\Gamma}$.

3.2 Distance in measure of the approximated random set to the original one

In the next proposition we show how it is possible to express the expected distance in measure as a function of the probability distribution of the underlying field only.

Proposition 3 (Distance in measure between Γ and $\widetilde{\Gamma}$)**.** *a) Assuming Z and $\widetilde{Z}$ are random fields such that Γ and $\widetilde{\Gamma}$ are random closed sets, $D \subset \mathbb{R}^d$ and μ is a finite Borel measure on D, we have*

$$d_{\mu,n}(\Gamma, \widetilde{\Gamma}) = \int \rho_{n,m}(\mathbf{x}) \mu(\mathrm{d}\mathbf{x}) \tag{4}$$

with

$$\begin{aligned} \rho_{n,m}(\mathbf{x}) &= \mathbb{P}_n\big(\mathbf{x} \in \Gamma \Delta \widetilde{\Gamma}\big) \\ &= \mathbb{P}_n(Z(\mathbf{x}) \geq t, \widetilde{Z}(\mathbf{x}) < t) + \mathbb{P}_n(Z(\mathbf{x}) < t, \widetilde{Z}(\mathbf{x}) \geq t). \end{aligned} \tag{5}$$

where $\mathbb{P}_n(\mathbf{x} \in \Gamma \Delta \widetilde{\Gamma})$ denotes the conditional probability $\mathbb{P}\big(\mathbf{x} \in \Gamma \Delta \widetilde{\Gamma} \;\big|\; Z(\mathbf{X}_n)\big)$.
b) Moreover if we denote the mean of Z conditional on $Z(\mathbf{X}_n)$ with m_n and the conditional covariance kernel with k_n, we get

$$\mathbb{P}_n(Z(\mathbf{x}) \geq t, \widetilde{Z}(\mathbf{x}) < t) = \Phi_2\left(\mathbf{c}_n(\mathbf{x}, E_m), \Sigma_n(\mathbf{x}, E_m)\right), \tag{6}$$

where $\Phi_2(\,\cdot\,, \Sigma)$ is the cumulative distribution function of a centered bivariate Gaussian with covariance Σ, with

$$\mathbf{c}_n(\mathbf{x}, E_m) = \begin{pmatrix} m_n(\mathbf{x}) - t \\ t - a(\mathbf{x}) - \mathbf{b}(\mathbf{x})^T m_n(E_m) \end{pmatrix}$$

and

$$\Sigma_n(\mathbf{x}, E_m) = \begin{pmatrix} k_n(\mathbf{x}, \mathbf{x}) & -\mathbf{b}(\mathbf{x})^T k_n(E_m, \mathbf{x}) \\ -\mathbf{b}(\mathbf{x})^T k_n(E_m, \mathbf{x}) & \mathbf{b}(\mathbf{x})^T k_n(E_m, E_m) \mathbf{b}(\mathbf{x}) \end{pmatrix}.$$

c) Particular case: If $\mathbf{b}(\mathbf{x})$ *is chosen as the kriging predictor* $\mathbf{b}(\mathbf{x}) = k_n(E_m, E_m)^{-1}k_n(E_m, \mathbf{x})$*, then*

$$\Sigma_n(\mathbf{x}, E_m) = \begin{pmatrix} k_n(\mathbf{x}, \mathbf{x}) & -\gamma_n(\mathbf{x}, E_m) \\ -\gamma_n(\mathbf{x}, E_m) & \gamma_n(\mathbf{x}, E_m) \end{pmatrix}$$

where $\gamma_n(\mathbf{x}, E_m) = \mathrm{Var}_n[\widetilde{Z}(\mathbf{x})] = k_n(E_m, \mathbf{x})^T k_n(E_m, E_m)^{-1}k_n(E_m, \mathbf{x})$.

Proof. a) Interchanging integral and expectation by Tonelli, we get

$$\begin{aligned} d_{\mu,n}(\Gamma, \widetilde{\Gamma}) &= \mathbb{E}_n[\mu(\Gamma \backslash \widetilde{\Gamma})] + \mathbb{E}_n[\mu(\widetilde{\Gamma} \backslash \Gamma)] \\ &= \mathbb{E}_n\left[\int \mathbf{1}_{Z(\mathbf{x})\geq t}\mathbf{1}_{\widetilde{Z}(\mathbf{x})<t}\mu(\mathrm{d}\mathbf{x}) + \int \mathbf{1}_{\widetilde{Z}(\mathbf{x})\geq t}\mathbf{1}_{Z(\mathbf{x})<t}\mu(\mathrm{d}\mathbf{x})\right] \\ &= \int \left[\mathbb{P}_n(Z(\mathbf{x}) \geq t, \widetilde{Z}(\mathbf{x}) < t) + \mathbb{P}_n(Z(\mathbf{x}) < t, \widetilde{Z}(\mathbf{x}) \geq t)\right] \mu(\mathrm{d}\mathbf{x}) \end{aligned}$$

b) Since the random field Z is assumed to be Gaussian, the vector $(Z(\mathbf{x}), \widetilde{Z}(\mathbf{x}))$ is also Gaussian conditionally on $Z(\mathbf{X}_n)$, and proving the property boils down to calculating its conditional moments.

Now we directly get $\mathbb{E}_n[Z(\mathbf{x})] = m_n(\mathbf{x})$ and $\mathbb{E}_n[\widetilde{Z}(\mathbf{x})] = a(\mathbf{x}) + \mathbf{b}(\mathbf{x})^T m_n(E_m)$. Similarly, $\mathrm{Var}_n[Z(\mathbf{x})] = k_n(\mathbf{x}, \mathbf{x})$ and $\mathrm{Var}_n[\widetilde{Z}(\mathbf{x})] = \mathbf{b}(\mathbf{x})^T k_n(E_m, E_m)\mathbf{b}(\mathbf{x})$. Finally, $\mathrm{Cov}_n[Z(\mathbf{x}), \widetilde{Z}(\mathbf{x})] = \mathbf{b}(\mathbf{x})^T k_n(E_m, \mathbf{x})$ and the conclusion follows from the Gaussianity of the field. □

Remark 1. *The Gaussian assumption on the random field* Z *in Proposition 3 can be relaxed: in part a) it suffices that the excursion sets of the field* Z *are random closed sets and in part b) it suffices that the field* Z *is Gaussian conditionally on* $Z(\mathbf{X}_n)$*.*

3.3 Convergence result

Let $\mathbf{e}_1, \mathbf{e}_2, \ldots$ be a given sequence of simulation points in D and set $E_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ for all m. Assume that Z is, conditionally on $Z(\mathbf{X}_n)$, a Gaussian random field with conditional mean m_n and conditional covariance function k_n. Let $\widetilde{Z}(\mathbf{x}) = \mathbb{E}_n\left(Z(\mathbf{x}) \mid Z(E_m)\right)$ be the best predictor of $Z(\mathbf{x})$ given $Z(\mathbf{X}_n)$ and $Z(E_m)$, which is affine in $Z(E_m)$, and denote by $s^2_{n,m}(\mathbf{x})$ the conditional variance of the prediction error at $\mathbf{x}$:

$$\begin{aligned} s^2_{n,m}(\mathbf{x}) &= \mathrm{Var}_n\left(Z(x) - \widetilde{Z}(x)\right) = \mathrm{Var}_n\left(Z(x) \mid Z(E_m)\right) \\ &= k_n\left(\mathbf{x}, \mathbf{x}\right) - k_n\left(E_m, \mathbf{x}\right)^T k_n\left(E_m, E_m\right)^{-1} k_n\left(E_m, \mathbf{x}\right). \end{aligned}$$

Proposition 4 (Approximation consistency). *Let* $\widetilde{\Gamma}(E_m) = \left\{\mathbf{x} \in D : \widetilde{Z}(\mathbf{x}) \in T\right\}$ *be the random excursion set associated to* $\widetilde{Z}$*. Then, as* $m \to \infty$*,* $d_{\mu,n}(\Gamma, \widetilde{\Gamma}(E_m)) \to 0$ *if and only if* $s^2_{n,m} \to 0$ μ*-almost everywhere.*

Corollary 5. *Assume that the covariance function of* Z *is continuous. a) If the sequence of simulation points is dense in* D*, then the approximation scheme is consistent (in the sense that* $d_{\mu,n}(\Gamma, \widetilde{\Gamma}(E_m)) \to 0$ *when* $m \to \infty$*). b) Assuming further that the covariance function of* Z *has the NEB property [36], the density condition is also necessary.*

The proof of Proposition 4 is in Appendix A

4 Practicalities

In this section we use the previously established results to implement a method that selects appropriate simulation points $E_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\} \subset D$. The conditional field is simulated on E_m and approximated at the design points with ordinary kriging predictors. We present two algorithms to find the set E_m that minimizes the expected distance in measure between Γ and $\widetilde{\Gamma}(E_m)$. We assume that the number of simulation points is fixed in advance and is equal to m. The algorithms were implemented in R with the packages `KrigInv` [11] and `DiceKriging` [30].

4.1 Algorithm A: minimizing $\mathbf{d_{\mu,n}(\Gamma, \widetilde{\Gamma})}$

The first proposed algorithm (Algorithm A) is a sequential minimization of the expected distance in measure $d_{\mu,n}(\Gamma, \widetilde{\Gamma})$. We exploit the characterization in Equation (4) and we assume that the underlying field Z is Gaussian. Under these assumptions, the problem becomes

$$\text{minimize } d_{\mu,n}(\Gamma, \widetilde{\Gamma}) = \int \left[\Phi_2\left(\mathbf{c}_n(\mathbf{x}, E_m), \Sigma_n(\mathbf{x}, E_m)\right) + \Phi_2\left(-\mathbf{c}_n(\mathbf{x}, E_m), \Sigma_n(\mathbf{x}, E_m)\right)\right] \mu(\mathrm{d}\mathbf{x}) \tag{7}$$

with respect to $E_m = \{\mathbf{e}_1, ..., \mathbf{e}_m\}$.

Several classic optimization techniques have already been employed to solve similar problems for optimal designs, for example simulated annealing, [31], genetic algorithms [21] or treed optimization [19]. In our case such global approaches would lead to a $m \times d$ dimensional problem and, since we do not rely on analytical gradients, the full optimization would be very slow. Instead we follow a greedy heuristic approach as in [32], [8] and optimize the criterion sequentially: given $E^*_{i-1} = \{\mathbf{e}^*_1, ..., \mathbf{e}^*_{i-1}\}$ points previously optimized, the ith point $\mathbf{e}_i$ is chosen as the minimizer of $d_{\mu,n}(\Gamma, \widetilde{\Gamma}^*_i)$ where $\widetilde{\Gamma}^*_i = \widetilde{\Gamma}(E^*_{i-1} \cup \{\mathbf{e}_i\})$. The points optimized in previous iterations are considered as fixed parameters and

therefore not modified by the current optimization. The algorithm adds new points until the required number of points is reached.

The parameters of the bivariate normal, $\mathbf{c}_n(\mathbf{x}, E_i)$ and $\Sigma_n(\mathbf{x}, E_i)$, depend on the set E_i and therefore need to be updated each time the optimizer requires an evaluation of the criterion in a new point. Those functions rely on the kriging equations, but recomputing each time the full kriging model is cumbersome numerically. Instead we exploit the sequential nature of the algorithm and use kriging update formulas [10] to compute the new value of the criterion each time a new point is analyzed.

Numerical evaluation of the expected distance in measure poses the issue of approximating both the integral in $\mathbb{R}^d$ and the bivariate normal distribution in Equation (7). The numerical approximation of the bivariate normal distribution is computed with the `pbivnorm` package which relies on the fast Fortran implementation of the standard bivariate normal CDF introduced in [18]. The integral is approximated via quasi-Monte Carlo method: the integrand is evaluated in points from a space filling sequence (Sobol, [6]) and then approximated with a weighted sample mean of the values.

The criterion is optimized with the function `genoud` [26], a genetic algorithm with BFGS descents that finds the optimum by evaluating the criterion over a population of points spread in the domain of reference and by evolving the population in sequential generations to achieve a better fitness.

The evaluation of the criterion in Equation (7) can become computationally expensive because it requires a high number of evaluation of the bivariate normal CDF in order to properly estimate the integral. This consideration led us to develop a second optimization algorithm.

4.2 Algorithm B: maximizing $\boldsymbol{\rho}_{\mathbf{n},\mathbf{m}}(\mathbf{x})$

We follow closely the reasoning used in [32] and [4] for the development of a heuristic method to minimize the integrated squared error by maximizing the mean squared error. The characterization of the expected distance in measure in Equation (4) is the integral of the sum of two probabilities. They are non-negative continuous functions of $\mathbf{x}$ as the underlying Gaussian field is continuous. The integral, therefore, is large if the integrand takes large values. Moreover, $\widetilde{Z}$ interpolates Z in E hence the integrand is zero in the chosen simulation points. The two previous considerations lead to a natural variation of Algorithm A where the simulation points are chosen in order to maximize the integrand.

Algorithm B is based on a sequential maximization of the integrand. Given $E^*_{i-1} = \{\mathbf{e}^*_1, ..., \mathbf{e}^*_{i-1}\}$ points previously optimized, the ith point $\mathbf{e}_i$ is chosen as the maximizer of $\rho^*_{n,i-1}(\mathbf{x})$, where

$$\rho^*_{n,i-1}(\mathbf{x}) = \Phi_2\left(\mathbf{c}_n(\mathbf{x}, E^*_{i-1}), \Sigma_n(\mathbf{x}, E^*_{i-1})\right) + \Phi_2\left(-\mathbf{c}_n(\mathbf{x}, E^*_{i-1}), \Sigma_n(\mathbf{x}, E^*_{i-1})\right).$$

The evaluation of the objective function in algorithm B does not require the approximation of an integral in $\mathbb{R}^d$, thus it requires substantially less approximations of the bivariate normal CDF.

The maximization of the objective function is performed with the L-BFGS-B algorithm [7] implemented in R with the function `optim`. The choice of starting points for the optimization is crucial for gradient descent algorithms; in our case we decided to exploit the analytic formula introduced in [4] for the coverage function of $\Gamma, p_\Gamma(\mathbf{x})$ to obtain a reasonable starting point for each sequential maximization. All points $\mathbf{x}_s$ with high values of $p_\Gamma(\mathbf{x}_s)(1 - p_\Gamma(\mathbf{x}_s))$ are reasonable starting points because they are located in regions of high uncertainty for the excursion set, thus simulations around their locations are more meaningful than in other locations. Before starting the maximization, the function $w(\mathbf{x}) = p_\Gamma(\mathbf{x})(1 - p_\Gamma(\mathbf{x}))$ is evaluated at a fine space filling design and, at each sequential maximization, the starting point is drawn from a distribution proportional to the computed values of w.

4.3 Comparison with non optimized simulation points

In order to quantify the importance of optimizing the simulation points and to show the differences between the two algorithms we first present a 2-d analytical example.

Consider the Branin-Hoo function (see [24]) multiplied by a factor -1 and normalized so that the domain becomes $D = [0, 1]^2$. We are interested in estimating the excursion set $\Gamma^\star = \{\mathbf{x} \in D : f(\mathbf{x}) \geq -10\}$ with $n = 20$ evaluations of f. We consider a Gaussian random field Z with mean and covariance function m and k respectively. The covariance function is chosen as a tensor product Matérn kernel ($\nu = 3/2$), [34], and its parameters are estimated by Maximum Likelihood with the package `DiceKriging` [30]. By following the GRF modeling approach we assume that f is a realization of Z and we condition Z on $n = 20$ evaluations. The evaluation points are chosen with a maximin Latin Hypercube Sample (LHS) design [35] and the conditional field is computed with ordinary kriging equations.

Discrete realization of the random set Γ on a fine grid can be obtained by selecting few optimized simulation points and by interpolating the simulations at those locations on the fine grid. The expected distance in measure is a good indicator of how close the reconstructed set realizations are to the actual realizations. Here we compare the expected distance in measure obtained with optimization algorithms A and B and with two space filling designs, namely a maximin Latin Hypercube Sample [35] and points from the Sobol sequence [6].

Figure 2 shows the expected distance in measure as a function of the number of simulation points. The values were computed only in the dotted points for Algorithms A and B and in each integer for the space filling designs. The optimized designs always achieve a smaller expected distance in measure, but it is clear that the advantage of accurately optimizing the choice of points decreases as the number of points increases, thus showing that the designs tend to become equivalent as the space is filled with points. This effect, linked to the low dimensionality of our example, reduces the advantage of optimizing the points, however in higher dimensions optimizing the points becomes more advantageous because a much larger number of points is required to fill the space. Algorithm

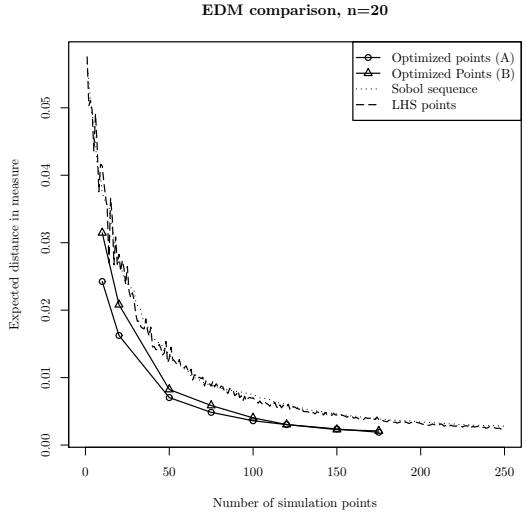


Figure 2: Expected distance in measure for different choices of simulation points

A and B show similar results for low numbers of points and almost identical results for more than 100 simulation points. Even though this effect might be magnified by the low dimension of the example, it is clear that in most situations Algorithm B is preferable to Algorithm A as it achieves similar precision while remaining significantly less computationally expensive, as shown in Figure 4.

5 Application: A new variability measure using the Distance Transform

In this section we deal with the notions of distance average and distance function variability introduced in Section 2 and more specifically we present an application where the interpolated simulations are used to efficiently compute the distance function variability. Let us recall that, given $\Gamma_1, ..., \Gamma_N$ realizations of the random closed set Γ, we can compute an estimator for $\mathbb{E}_{\mathrm{DA}}(\Gamma)$ as

$$\mathbb{E}^*_{DA}(\Gamma) = \{\mathbf{x} \in D : \bar{d}^*(\mathbf{x}) \leq \bar{u}^*\}, \tag{8}$$

where $\bar{d}^*(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} d(\mathbf{x}, \Gamma_i)$ is the empirical distance function and $\bar{u}^*$ is the threshold level for $\bar{d}^*$, chosen in a similar fashion as $\bar{u}$ in Definition 2, see [2] for more detail. The variability of this estimate is measured with the distance average variability $\mathrm{DTV}(\Gamma)$, which, in the empirical case, is defined as $\frac{1}{N}\sum_{i=1}^{N} \delta(\bar{d}^*, d(\cdot, \Gamma_i))$. In the following we will take the usual Lebesgue $L^2(\mathbb{R}^d)$

distance as functional distance δ, thus the distance average variability becomes

$$\mathrm{DTV}(\Gamma) = \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}^d}\left(d(\mathbf{x},\Gamma_i) - \bar{d}^*(\mathbf{x})\right)^2 d\mu(\mathbf{x}). \tag{9}$$

The distance average variability is a measure of uncertainty for excursion set under the postulated GRF model; if the distance average variability is high it means that the excursion set estimated with the distance average has a "high variance" caused by very different realizations of the excursion set, thus the estimate is uncertain. This uncertainty quantification method necessitates conditional simulations of the field on a fine grid to obtain a pointwise estimate. Our simulation method generates approximations of the simulations in a rather inexpensive fashion even on high quality grids, thus it is possible to compute this uncertainty measure.

We consider here a two dimensional example in the same setup as in Section 4 and we show that by selecting few well-chosen simulation points $E_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$, with $m \ll r$, and interpolating the results on G, it is possible to achieve very similar precision to full design simulations. The design considered for both the full simulations and the interpolated simulations is a grid with $r = q \times q$ points, where $q = 50$. The grid design allows us to compute numerically the distance transform with an adaptation for R of the fast distance transform algorithm implemented in [16]. The precision of the estimate $\mathbb{E}^*_{DA}(\Gamma)$ is evaluated with the distance transform variability $\mathrm{DTV}(\Gamma)$, Equation (9), approximated numerically on the grid.

A benchmark estimate of $\mathrm{DTV}(\Gamma; r)$ is first obtained from realizations of Γ stemming from 10.000 conditional Gaussian simulations on the grid of size $r = 50 \times 50$. The distance transform variability is then computed for realizations of Γ obtained via reconstruction from simulations at few points. The conditional Gaussian field is first simulated 10.000 times at a design E_m containing few optimized points, namely $m = 10, 20, 50, 75, 100, 120, 150, 175$, and then the results are interpolated on the $q \times q$ grid used for the benchmark. The experiments are reproduced 100 times, thus obtaining an empirical distribution of $\mathrm{DTV}(\Gamma; r)$, with $r = 2500$, and of $\mathrm{DTV}(\Gamma; m)$ for each m. Three methods are compared to obtain simulation points: Algorithm A and B presented in the previous section and a maximin LHS design. The simulation points obtained with the three methods are interpolated on the grid with the same technique. In particular, the ordinary kriging weights are first computed in each point $\mathbf{u} \in G$ and the value of the interpolated field $\widetilde{Z}(\mathbf{u})$ is obtained as a linear combination of the simulated values $Z(E_m)$ weighted by the kriging weights for each $\mathbf{u} \in G$. This procedure is numerically very fast as it only requires algebraic operations.

Figure 3 shows a comparison of the distributions of DTV obtained with full grid simulations and the distributions obtained with the interpolation over the grid of few simulations.

The distributions of DTV obtained with interpolations of simulations at few locations all approximate very well the benchmark distribution with as little as 100 simulation points, independently of the way simulation points are

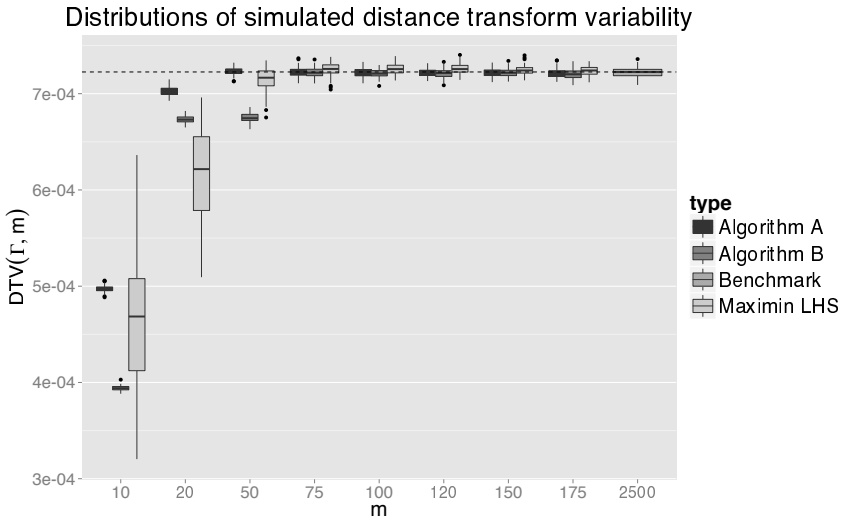


Figure 3: Comparison of the distributions of the simulated DTV(Γ) for different methods, the dashed horizontal line marks the median value of the benchmark distribution.

selected. This effect might be enhanced by the low dimension of the example, but nonetheless it shows substantial savings in simulation costs.

The optimized designs (Algorithm A and B) achieve better approximations with less points than the maximin LHS design. In particular the maximin LHS design is affected by a high variability, while the optimized points converge fast to a good approximation of the benchmark distribution. Interpolation of simulations at $m = 50$ points optimized with Algorithm A results in a relative error of the median estimate with respect to the benchmark of around 0.1%.

Algorithm A and B behave similarly when estimating this variability measure with $m \geq 75$, thus confirming that the reconstructed sets obtained from simulations at points that optimize either one of the criteria are very similar, as already hinted by the result on distance in measure shown in the previous section.

Figure 4 shows the total CPU time for all the simulations in the experiment for Algorithm A, Algorithm B and for the full grid simulations. The CPU times for Algorithm A and B also include the time required to optimize the simulation points. Both interpolation algorithms require less total CPU time than full grid simulations to obtain good approximations of the benchmark distribution ($m >$ 100). If parallel computing is available wall clock time could be significantly reduced by parallelizing operations. In particular the full grid simulation can be parallelized quite easily while the optimization of the simulation points could be much harder to parallelize. The times were computed on the cluster of the University of Bern with an Intel Xeon E5649 2.53GHz CPU with 4GB RAM.

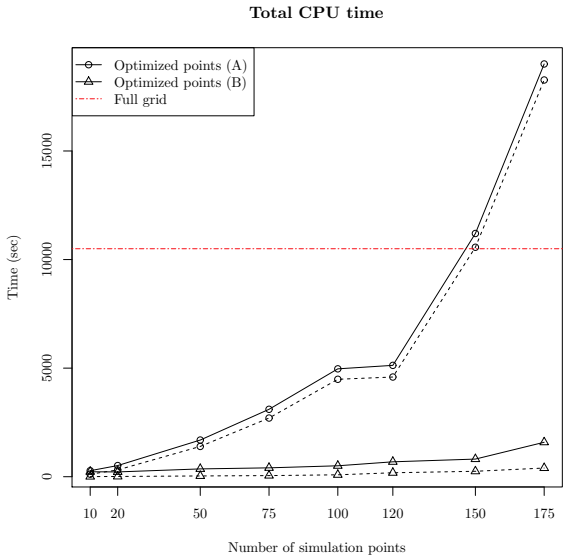


Figure 4: Total CPU time to obtain all realizations of Γ. Full grid simulations only include simulation time (dot-dashed horizontal line), while both algorithms include simulation point optimization (dashed lines) and simulation and interpolation times (solid lines).

6 Application: Estimating the distribution of a volume of excursion in six dimensions

In this section we show how it is possible to approach the conditional distribution of a volume of excursion under a GRF prior by simulating at few well chosen points and predicting over fine designs.

In the framework developed in section 2, the random closed set Γ naturally defines a distribution for the excursion set, thus $\mu(\Gamma)$ can be regarded as a random variable. In the specific case of a Gaussian prior, the expected volume of excursion can be computed analytically by integrating the coverage function, however working out the posterior distribution of this volume requires Monte Carlo simulations (see [37], [1]). In practice, a good estimation of the volume requires a discretization of the random closed set on a fine design. However, already in moderate dimensions ($2 \leq d \leq 10$), a discretization of the domain fine enough to achieve good approximations of the volume might require simulating at a prohibitive number of points. Here we show how the proposed approach can help solving this problem on a six-dimensional analytical example.

We consider the following test function $h(\mathbf{x}) = -\log(\text{Hartman}_6(\mathbf{x}))$, where Hartman_6 is the six-dimensional Hartman function (see [24]) defined on $D = [0,1]^6$ and we are interested in estimating the volume distribution of the excursion set $\Gamma^\star = \{\mathbf{x} \in D : h(\mathbf{x}) \geq t\}$, $t = 6$. The threshold $t = 6$ is chosen to obtain a true volume of excursion of around 3%, thus rendering the excursion set a moderately small set.

A GRF model is built with a Gaussian prior Z with a tensor product Matérn covariance kernel ($\nu = 5/2$). The parameters of the covariance kernel are estimated by Maximum Likelihood from $n = 60$ observations of h; the same observations are used to compute the conditional random field. We consider the discretization $G = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \subset D$ with $r = 10.000$ and $\mathbf{u}_1, \ldots, \mathbf{u}_r$ Sobol points in $[0,1]^6$. The conditional field Z is simulated 10.000 times on G and consequently $N = 10.000$ realizations of the trace of Γ over G are obtained.

The distribution of the volume of excursion can be estimated by computing for each realization the proportion of points where the field takes values above the threshold. While this procedure is acceptable for realizations coming from full design simulations, it introduces a bias when it is applied to reconstructed realizations of excursion sets. In fact, the paths of the predicted field are always smoother than the paths of full design simulations due to the linear nature of the predictor ([33]). This introduces a systematic bias on the volume of excursion for each realization because the sets of excursion induced by small rougher variations of the true Gaussian field are not intercepted by the predicted field. The effect changes the mean of the distribution, but it does not influence the variance of the distribution. In the present setting, for example, we observed that the mean volume of excursion was consistently underestimated. A modification of the classic estimate of the distribution of the volume is here considered. Given a full simulation design of size r, the distribution of the volume of excursion is obtained with the following steps: firstly the mean volume of excursion is esti-

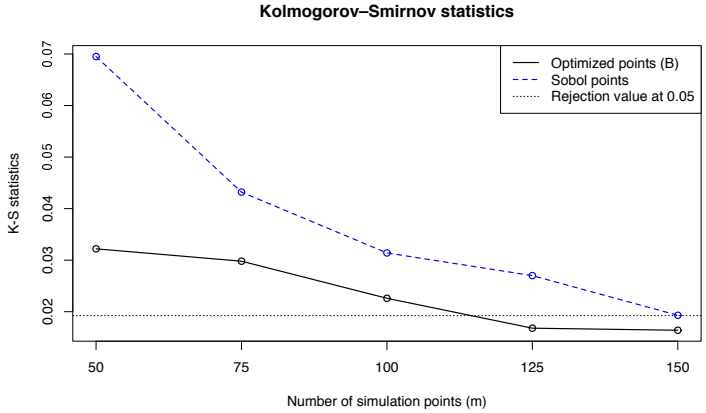


Figure 5: Kolmogorov-Smirnov statistics for testing if the distribution of V_m is equal to the distribution of V_{full}. The values for distributions of volume of reconstructed sets with simulations at points obtained with algorithm B (black full line) are compared with values obtained from simulations at space filling points (blue dashed line). The dotted horizontal line is the rejection value at level 0.05.

mated by integrating the coverage function of Γ over the full design; secondly the distribution is obtained by computing the volume of excursion for each realization of the reconstructed excursion set; finally the distribution is centered in the mean value obtained with the first step. The optimal simulation points are computed with Algorithm B because for a large number of points it achieves very similar results to Algorithm A but at the same time the optimized points are much cheaper to compute, as showed in the previous sections.

Denote with $V_{full} = \mu(\Gamma(E_{10.000}))$ the random variable representing the volume of the excursion set obtained with full design simulations and $V_m = \mu(\Gamma(E_m))$ the recentered random variable representing the volume of the reconstructed set obtained from simulations at m points. We compare the distribution of V_{full} and V_m for different values of m with Kolmogorov-Smirnov two sample tests. Figure 5 shows the values of the Kolmogorov-Smirnov statistics obtained for testing whether the distribution of V_m is equal to the distribution of V_{full}, for $m = 50, 75, 100, 125, 150$. V_m is computed both with simulation points optimized with Algorithm B and with points from a space filling Sobol sequence. In the Figure the horizontal line is the rejection value at level 0.05. With confidence level 0.05, the distribution of V_m is not distinguishable from the distribution of V_{full} if $m \geq 125$ with optimized points and if $m > 150$ with Sobol points.

The estimate of the distribution of the volume of excursion is much faster with reconstructed sets from simulations at few optimal locations. In fact, the computational costs are significantly reduced with by interpolating simulations:

the CPU time needed to simulate on the full 10.000 points design is 60293 seconds while the total time for the optimization of $m = 150$ points, the simulation on those points and the prediction over the full design is 575 seconds. Both times were computed on the cluster of the University of Bern with an Intel Xeon E5649 2.53GHz CPU with 4GB RAM.

7 Conclusions

In the context of excursion set estimation, simulating a conditional random field to obtain realizations of a related excursion set can be useful in many practical situations. However often the random field needs to be simulated at a fine design to obtain meaningful realization of the excursion set. Even in moderate dimensions it is often impractical to simulate at such fine designs, thus rendering good approximations hard to achieve.

In this paper we introduced a new method to simulate realizations from a conditional Gaussian random field that mitigates this problem. While the approach of predicting simulations from few simulation points has already been introduced in the literature, this is the first attempt to define optimal points where to simulate based on a specific distance between random closed sets: the expected distance in measure.

This approach allowed us to study an uncertainty measure that, to the best of our knowledge, has not been proposed before: the distance transform variability. The estimation of the distance transform variability is appealing if it is possible to obtain realizations of the excursion set on fine grids at low computational costs. We showed on a two-dimensional test function that it is possible to reduce the computational costs by at least one order of magnitude, thus making this uncertainty quantification technique appealing. In general the optimal simulation points approach could improve the speed of distance average based methods as for example [22] and [23].

Conditional realizations of the excursion set can also be used to estimate the volumes of excursion. This problem requires Monte Carlo simulations at fine designs in order to attain good approximations of the excursion volumes. We showed on a test case in six dimensions that it is possible to obtain estimates of the distribution with simulations at few optimal points that are indistinguishable from the estimates of the distribution obtained with full design simulations. This study drew our attention to the regularity of the predicted paths because we observed a bias in the estimate of the volume due to different smoothness properties of full design simulation and predicted realizations. In this case the bias was corrected by estimating the mean of the distribution via some fast centering step, however this issue highlights the need for further studies on the biases introduced by our random field reconstruction approach.

We presented two algorithms to compute optimal simulation points. While the heuristic algorithm B is appealing for its computational cost and precision, there are a number of extensions that could lead to even more savings in computational time, for example, the optimization of the points in this work was

carried out with generic black box optimizers but it would be possible to achieve appreciable reductions in optimization time with methods based on analytical gradients.

A Proof of Proposition 4

Let us first assume that $s^2_{n,m} \to 0$ μ-almost everywhere. The expected distance in measure can be rewritten, according to Equation (4), as $d_{\mu,n}(\Gamma, \widetilde{\Gamma}) = \int_D \rho_{n,m}(\mathbf{x})\, \mu(\mathrm{d}\mathbf{x})$. Since μ is a finite measure on D and $\rho_{n,m}(x) \leq 1$, it is sufficient by the dominated convergence theorem to prove that $\rho_{n,m} \to 0$ μ-almost everywhere.

Pick any $x \in D$ such that $s^2_n(\mathbf{x}) > 0$ and $s^2_{n,m}(\mathbf{x}) \to 0$. Then, for any $w > 0$,

$$\begin{aligned}\rho_{n,m}(\mathbf{x}) &\leq \mathbb{P}_n\big(\big|Z(\mathbf{x}) - t\big| \leq w\big) + \mathbb{P}_n\big(\big|\widetilde{Z}(\mathbf{x}) - Z(\mathbf{x})\big| \geq w\big) \\ &\leq \frac{2w}{\sqrt{2\pi s^2_n(\mathbf{x})}} + \frac{s^2_{n,m}(\mathbf{x})}{w^2}.\end{aligned}$$

With $w = \sqrt{s_{n,m}(\mathbf{x})}$, it follows that

$$\rho_{n,m}(\mathbf{x}) \leq \frac{2\sqrt{s_{n,m}(\mathbf{x})}}{\sqrt{2\pi s^2_n(\mathbf{x})}} + s_{n,m}(\mathbf{x}) \to 0. \tag{10}$$

Since $s^2_{n,m} \to 0$ μ-almost everywhere and $\rho_{n,m}(\mathbf{x}) = 0$ wherever $s^2_n(\mathbf{x}) = 0$, Equation (10) proves the sufficiency part of Proposition 4.

Conversely, assume that $d_{\mu,n}(\Gamma, \widetilde{\Gamma}) \to 0$ when $m \to +\infty$, or equivalently that $(\rho_{n,m})_{m\geq 0}$ converges to zero in $L^1(D, \mu)$. Then $(\rho_{n,m})_{m\geq 0}$ also converges to zero in measure:

$$\forall \varepsilon > 0,\ \mu\left(A^{\varepsilon}_{n,m}\right) \xrightarrow[m\to+\infty]{} 0, \qquad \text{where } A^{\varepsilon}_{n,m} = \{x \in D : \rho_{n,m}(x) \geq \varepsilon\}.$$

For any $M > 1$, consider the following sets:

$$\begin{aligned}D_{n,M} &= \big\{x \in D :\ 0 < \tfrac{1}{M} s_n(x) \leq \big|t - m_n(x)\big| \leq M s_n(x)\big\}, \\ A^{M,\varepsilon}_{n,m} &= D_{n,M} \cap A^{\varepsilon}_{n,m}, \\ B^{M,\varepsilon}_{n,m} &= D_{n,M} \cap \{s_{n,m} \geq \varepsilon s_n\}.\end{aligned}$$

Then we have the following technical result.

Lemma 6. *For all $M > 1$ and $\varepsilon > 0$, there exists $\varepsilon' > 0$ (that does not depend on n, m or t) such that $\forall x \in B^{M,\varepsilon}_{n,m}$, $\rho_{n,m}(x) \geq \varepsilon'$, and therefore $B^{M,\varepsilon}_{n,m} \subset A^{M,\varepsilon'}_{n,m}$.*

Using Lemma 6, for any $M > 1$ and $\varepsilon > 0$, we have

$$\mu\left(B^{M,\varepsilon}_{n,m}\right) \leq \mu\left(A^{M,\varepsilon'}_{n,m}\right) \leq \mu\left(A^{\varepsilon'}_{n,m}\right) \xrightarrow[m\to+\infty]{} 0.$$

In other words, $(s_{n,m}/s_n)_{m\geq 0}$ converges to zero in measure on $D_{n,M}$. As a consequence, since this is a decreasing sequence, $(s_{n,m}/s_n)_{m\geq 0}$ converges to zero μ-almost everywhere on $D_{n,M}$, and therefore μ-almost everywhere on $\cup_{M>1} D_{n,M} = \{x \in D : s_n(x) > 0\}$. Convergence also trivially holds where $s_n(x) = 0$.

A.1 Proof of Lemma 6

Let $M > 1$, $\varepsilon > 0$, $\mathbf{x} \in B_{n,m}^{M,\varepsilon}$ and set $\zeta = m_n(\mathbf{x})$. Because $B_{n,m}^{M,\varepsilon} \subset D_{n,M}$, $|t-\zeta| \geq \varepsilon s_n(\mathbf{x}) > 0$ and in particular $t \neq \zeta$. Assume, without loss of generality, that $\zeta < t$ and $\varepsilon \leq \frac{1}{\sqrt{2}}$. Recall that $\rho_{n,m}(\mathbf{x}) = \mathbb{P}_n\left(E_{n,m}(\mathbf{x})\right)$ where $E_{n,m}(\mathbf{x})$ is the event defined by

$$
\begin{aligned}
E_{n,m}(\mathbf{x}) &= \left\{\mathbf{x} \in \Gamma \Delta \widetilde{\Gamma}\right\} = E_{n,m}^{+}(\mathbf{x}) \cup E_{n,m}^{-}(\mathbf{x}),\\
E_{n,m}^{+}(\mathbf{x}) &= \left\{Z(\mathbf{x}) < t,\ \widetilde{Z} \geq t\right\},\\
E_{n,m}^{-}(\mathbf{x}) &= \left\{Z(\mathbf{x}) \geq t,\ \widetilde{Z} < t\right\}.
\end{aligned}
$$

Recall also that, since Z is a Gaussian process, $\widetilde{Z}$ and $\epsilon_{n,m}(\mathbf{x}) = Z(\mathbf{x}) - \widetilde{Z}$ are independent Gaussian variables under $\mathbb{P}_n$, with $\kappa_{n,m}^2(\mathbf{x}) \coloneqq \mathrm{Var}_n\left(\epsilon_{n,m}(\mathbf{x})\right) = s_n^2(\mathbf{x}) - s_{n,m}^2(\mathbf{x})$.

Let us first assume that $s_{n,m}(\mathbf{x}) \geq \frac{1}{\sqrt{2}} s_n(\mathbf{x}) \geq \varepsilon s_n(\mathbf{x})$. As a consequence, $\kappa_{n,m}(\mathbf{x}) \leq \frac{1}{\sqrt{2}} s_n(\mathbf{x})$ and therefore

$$
\frac{t-\zeta}{\kappa_{n,m}(\mathbf{x})} \geq \frac{1}{M}\frac{s_n(\mathbf{x})}{\kappa_{n,m}(\mathbf{x})} \geq \frac{\sqrt{2}}{M} \qquad \text{and} \quad \frac{t-\zeta}{s_{n,m}(\mathbf{x})} \leq M\frac{s_n(\mathbf{x})}{s_{n,m}(\mathbf{x})} \leq M\sqrt{2}.
$$

For any $w > 0$, the following inclusions hold:

$$
\begin{aligned}
E_{n,m}(\mathbf{x}) \supset E_{n,m}^{-}(\mathbf{x}) &= \left\{\widetilde{Z} < t \text{ and } \widetilde{Z} + \epsilon_{n,m}(\mathbf{x}) \geq t\right\}\\
&\supset \left\{\zeta - w \leq \widetilde{Z} \leq t \text{ and } \zeta - w + \epsilon_{n,m}(\mathbf{x}) \geq t\right\}\\
&\supset \left\{-w \leq \widetilde{Z} - \zeta \leq t - \zeta \text{ and } \epsilon_{n,m}(\mathbf{x}) \geq t - \zeta + w\right\}.
\end{aligned}
$$

With $w = t - \zeta$, using the independence of $\widetilde{Z}$ and $\epsilon_{n,m}(\mathbf{x})$, we get

$$
\begin{aligned}
\rho_{n,m}(\mathbf{x}) &\geq \mathbb{P}_n\left(E_{n,m}^{-}(\mathbf{x})\right)\\
&\geq \mathbb{P}_n\big(\big|\widetilde{Z} - \zeta\big| \leq t - \zeta\big)\,\mathbb{P}_n\big(\epsilon_{n,m}(\mathbf{x}) \geq 2(t-\zeta)\big)\\
&\geq \left(1 - 2\Phi\left(-\sqrt{2}/M\right)\right)\,\Phi\left(-2M\sqrt{2}\right). \qquad (11)
\end{aligned}
$$

Let us now assume that $\frac{1}{\sqrt{2}} s_n(\mathbf{x}) \geq s_{n,m}(\mathbf{x}) \geq \varepsilon s_n(\mathbf{x})$. As a consequence, $\kappa_{n,m}(\mathbf{x}) \geq \frac{1}{\sqrt{2}} s_n(\mathbf{x})$ and therefore

$$
\frac{t-\zeta}{s_{n,m}(\mathbf{x})} \leq M\frac{s_n(\mathbf{x})}{s_{n,m}(\mathbf{x})} \leq \frac{M}{\varepsilon} \qquad \text{and} \quad \frac{1}{M} \leq \frac{t-\zeta}{\kappa_{n,m}(\mathbf{x})} \leq M\sqrt{2}.
$$

For any $w > 0$, the following inclusions hold:

$$\begin{aligned} E_{n,m}(\mathbf{x}) \supset E^+_{n,m}(\mathbf{x}) &= \left\{ \widetilde{Z} \geq t \text{ and } \widetilde{Z} + \epsilon_{n,m}(\mathbf{x}) < t \right\} \\ &\supset \left\{ t \leq \widetilde{Z} \leq t + w \text{ and } \epsilon_{n,m}(\mathbf{x}) < -w \right\}. \end{aligned}$$

Using again $w = t - \zeta$ and the independence of $\widetilde{Z}$ and $\epsilon_{n,m}(\mathbf{x})$, we get

$$\begin{aligned} \rho_{n,m}(\mathbf{x}) &\geq \mathbb{P}_n\left(E^+_{n,m}(\mathbf{x})\right) \\ &\geq \mathbb{P}_n\big(t - \zeta \leq \widetilde{Z} - \zeta \leq 2(t-\zeta)\big)\, \mathbb{P}_n\big(\epsilon_{n,m}(\mathbf{x}) \geq -(t-\zeta)\big) \\ &\geq \frac{1}{M}\, \varphi\left(2\sqrt{2}M\right)\, \Phi\left(-\frac{M}{\varepsilon}\right), \end{aligned} \tag{12}$$

where φ denotes the probability density function of the standard Gaussian distribution.

Finally, $\rho_{n,m}(\mathbf{x}) \geq \varepsilon'$, where ε' denotes the minimum of the lower bounds obtained in (11) and (12).

Acknowledgment

The authors wish to thank Ilya Molchanov for helpful discussions and insight.

References

[1] R. Adler, *On excursion sets, tube formulas and maxima of random fields*, Ann. Appl. Probab., 10 (2000), pp. 1–74.

[2] A. J. Baddeley and I. S. Molchanov, *Averaging of random sets based on their distance functions*, J. Math. Imaging Vision, 8 (1998), pp. 79–92.

[3] M. Bayarri, J. O. Berger, E. S. Calder, K. Dalbey, S. Lunagomez, A. K. Patra, E. B. Pitman, E. T. Spiller, and R. L. Wolpert, *Using statistical and computer models to quantify volcanic hazards*, Technometrics, 51 (2009), pp. 402–413.

[4] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez, *Sequential design of computer experiments for the estimation of a probability of failure*, Stat. Comput., 22 (3) (2012), pp. 773–793.

[5] M. Binois, D. Ginsbourger, and O. Roustant, *Quantifying uncertainty on Pareto fronts with Gaussian process conditional simulations*, European J. Oper. Res., In Press (2014).

[6] P. Bratley and B. L. Fox, *Algorithm 659: Implementing Sobol's quasirandom sequence generator*, ACM Trans. Math. Software, 14 (1988), pp. 88–100.

[7] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, *A limited memory algorithm for bound constrained optimization*, SIAM J. Sci. Comput., 16 (1995), pp. 1190–1208.

[8] C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, and Y. Richet, *Fast kriging-based stepwise uncertainty reduction with application to the identification of an excursion set*, Technometrics, 56 (2014), pp. 455–465.

[9] C. Chevalier, D. Ginsbourger, J. Bect, and I. Molchanov, *Estimating and quantifying uncertainties on level sets using the Vorob'ev expectation and deviation with Gaussian process models*, in mODa 10 Advances in Model-Oriented Design and Analysis, D. Uciński, A. Atkinson, and C. Patan, eds., Physica-Verlag HD, 2013.

[10] C. Chevalier, D. Ginsbourger, and X. Emery, *Corrected kriging update formulae for batch-sequential data assimilation*, in Mathematics of Planet Earth, Lecture Notes in Earth System Sciences, Springer Berlin Heidelberg, 2014, pp. 119–122.

[11] C. Chevalier, V. Picheny, and D. Ginsbourger, *The KrigInv package: An efficient and user-friendly R implementation of kriging-based inversion algorithms*, Comput. Statist. Data Anal., 71 (2014), pp. 1021–1034.

[12] J.-P. Chilès and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty, Second Edition*, Wiley, New York, 2012.

[13] N. Cressie, *Statistics for spatial data*, Wiley, New York, 1993.

[14] A. Cuevas and R. Fraiman, *Set estimation*, in New Perspectives in Stochastic Geometry, W. S. Kendall and I. Molchanov, eds., Oxford Univ. Press, Oxford, 2010, pp. 374–397.

[15] V. Dubourg, B. Sudret, and J.-M. Bourinet, *Reliability-based design optimization using kriging surrogates and subset simulation*, Struct. Multidiscip. Optim., 44 (2011), pp. 673–690.

[16] P. Felzenszwalb and D. Huttenlocher, *Distance transforms of sampled functions*, tech. rep., Cornell University, 2004.

[17] J. P. French and S. R. Sain, *Spatio-temporal exceedance locations and confidence regions*, Ann. Appl. Stat., 7 (2013), pp. 1421–1449.

[18] A. Genz, *Numerical computation of multivariate normal probabilities*, J. Comput. Graph. Statist., 1 (1992), pp. 141–149.

[19] R. B. Gramacy and H. K. H. Lee, *Adaptive design and analysis of supercomputer experiments*, Technometrics, 51 (2009), pp. 130–145.

[20] P. Hall and I. Molchanov, *Sequential methods for design-adaptive estimation of discontinuities in regression curves and surfaces*, Ann. Statist., 31 (2003), pp. 921–941.

[21] M. Hamada, H. Martz, C. Reese, and A. Wilson, *Finding near-optimal bayesian experimental designs via genetic algorithms*, Amer. Statist., 55 (2001), pp. 175–181.

[22] H. Jankowski and L. Stanberry, *Expectations of random sets and their boundaries using oriented distance functions*, J. Math. Imaging Vision, 36 (3) (2010), pp. 291–303.

[23] ———, *Confidence regions for means of random sets using oriented distance functions*, Scand. J. Stat., 39 (2) (2012), pp. 340–357.

[24] D. R. Jones, M. Schonlau, and W. J. Welch, *Efficient global optimization of expensive black-box functions*, J. Global Optim., 13 (1998), pp. 455–492.

[25] C. Lantuéjoul, *Geostatistical simulation: models and algorithms*, Springer, Berlin, 2002.

[26] W. J. Mebane and J. Sekhon, *Genetic optimization using derivatives: The rgenoud package for R*, Journal of Statistical Software, 42 (11) (2011), pp. 1–26.

[27] I. Molchanov, *Theory of Random Sets*, Springer, London, 2005.

[28] J. Oakley, *Bayesian Uncertainty Analysis for Complex Computer Codes*, PhD thesis, University of Sheffield, 1999.

[29] M. Reitzner, E. Spodarev, D. Zaporozhets, et al., *Set reconstruction by Voronoi cells*, Adv. in Appl. Probab., 44 (2012), pp. 938–953.

[30] O. Roustant, D. Ginsbourger, and Y. Deville, *DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodelling and optimization*, Journal of Statistical Software, 51 (2012), pp. 1–55.

[31] J. Sacks and S. Schiller, *Spatial designs*, Statistical decision theory and related topics IV, 2 (1988), pp. 385–399.

[32] J. Sacks, W. Welch, T. Mitchell, and H. Wynn, *Design and analysis of computer experiments*, Statist. Sci., 4 (1989), pp. 409–435.

[33] M. Scheuerer, *A comparison of models and methods for spatial interpolation in statistics and numerical analysis*, PhD thesis, University of Göttingen, 2009.

[34] M. Stein, *Interpolation of spatial data, some theory for kriging*, Springer, New York, 1999.

[35] M. L. Stein, *Large sample properties of simulations using latin hypercube sampling*, Technometrics, 29 (1987), pp. 143–151.

[36] E. Vazquez and J. Bect, *Convergence properties of the expected improvement algorithm with fixed mean and covariance functions*, J. Statist. Plann. Inference, 140:11 (2010), pp. 3088–3095.

[37] E. Vazquez and M. P. Martinez, *Estimation of the volume of an excursion set of a Gaussian process using intrinsic kriging*, arXiv preprint math/0611273, (2006).

[38] J. Villemonteix, E. Vazquez, and E. Walter, *An informational approach to the global optimization of expensive-to-evaluate functions*, J. Global Optim., 44 (2009), pp. 509–534.