Contributions to Bayesian set estimation relying on random field priors

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Dario Filippo Azzimonti
von Italien

Leiter der Arbeit:
PD Dr. David Ginsbourger
Institut de Recherche Idiap und Institut für Mathematische
Statistik und Versicherungslehre der Universität Bern
und
Prof. Dr. Ilya Molchanov
Institut für Mathematische Statistik und Versicherungslehre
der Universität Bern

Contributions to Bayesian set estimation relying on random field priors

vorgelegt von

Dario Filippo Azzimonti
von Italien

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, 17. November 2016

Der Dekan
Prof. Dr. Gilberto Colangelo

ii

Ai miei genitori.

iv

Acknowledgements

It is still not easy to grasp, but the time of my Ph.D. studies is over. These last three years flew by very quickly and I can quite safely say that this has been the most stimulating period of my life.

Although many people helped me during this time, the work presented here would have not been possible without the constant motivation, support and help of my supervisors PD Dr. David Ginsbourger and Prof. Ilya Molchanov. I would like to warmly thank Prof. Chien-Fu Jeff Wu for serving as external referee for this manuscript and Prof. Lutz Dümbgen for being the president of the defence committee.

Among the many people who helped me in this journey I first wish to thank David Ginsbourger who, already at our first meeting during the interview, impressed me for his wealth of ideas. Since the beginning of my Ph.D. he helped me to grow both personally and from a research point of view. I owe it in large part to David if my understanding of statistics and of the research world increased so fast in the last years. Working with him has been demanding, stimulating and always exciting. He was there during all "emergencies" occurred in my Ph.D. I will always remember our last minute Skype talks and revisions over restaurant tables.

I wish to thank my co-authors, Clément Chevalier (UniNe), Julien Bect (L2S, CentraleSupelec) and Yann Richet (IRSN) for their help and collaboration in the development of the research presented here. Clément is an extremely valuable co-author and co-worker. All discussions with him have always been very enriching and productive: his particular mix of theoretical depth and practical sense has been invaluable for this work. He has also been an easy-going colleague and a friend. I wish to thank Julien for being a very dynamic and precise colleague. His visit in Bern in August 2014 was short but very productive. Likewise, during all our subsequent meetings at conferences or workshops our discussions were fruitful and constructive. My sincere thanks

During my Ph.D. I had the opportunity to attend and present at many meetings and workshops. The discussions I had at all those gatherings were very enriching and contributed to a better understanding of the current work. I am very thankful to Prof. Ethan Anderes and Prof. Wolfgang Polonik for hosting me at UC Davis and making me part of their institute life for a few days. I am very grateful to Prof. Evgeny Spodarev for hosting me in Ulm and allowing me to present my work there. I would like to extend my gratitude to all members of the Ensemble research group, who gave me the opportunity to participate to their meetings and to their workshop on MCMC. I would also like to thank the ReDice community for welcoming me at their meetings and giving me the opportunity to present my work. I am grateful to the travel support fund for young researchers of the Faculty of Science of the University of Bern for financially supporting my presentation at the 8th International Congress on Industrial and Applied Mathematics (ICIAM 2015).

During my time in Bern, I was fortunate enough to meet very nice people who would then become very good friends. Thank you all for the nice evenings, the wonderful hikes and our fun Aare trips, for introducing me to the art of Jassen and for all your attempts at teaching me German. In December 2013, few months after the beginning of my Ph.D., I met Anna and since then I have been so fortunate to share my life with her. Even if we have been living in different cities, she has supported me through good and bad times. Thank you Anna for bearing with me in the (several) moments of discomfort, emergency and utter panic that I lived in those three years. I can only imagine how difficult I can be sometimes. Finally I wish to thank my family for always being close to me even from a distance. Thank you Oscar for your irony that always makes me think and laugh. Thank you mom and dad for "forcing" me to take back Italian food supplies every time I came back. With the amount of food you put in my suitcase I was always sure to survive in the unlikely event of a war in Switzerland. More importantly, thank you for teaching me to always look forward and not to be afraid of new experiences.

viii

Abstract

In this work we study the problem of estimating excursions sets of expensive to evaluate functions under a limited evaluation budget. We consider a Bayesian approach and we use Gaussian process modelling. The posterior distribution of the process induces a distribution on the excursion set, i.e. it defines a random closed set. By summarizing the posterior distribution of the random closed set we can provide set estimates and quantify their uncertainty. The first contribution of this work provides a method to generate quasi-realizations of the posterior field from a finite dimensional multivariate Gaussian distribution. In particular the proposed technique generates optimal quasi-realizations of the posterior excursion set. The method allowed preliminary studies on a new uncertainty measure for the set estimate and more in general it could facilitate the computation of empirical quantities related to the set. The other two major contributions are linked to the concept of conservative estimates. In fact, current implementations of conservative estimates require many evaluations of a high dimensional multivariate Gaussian orthant probability. We propose an estimator for this quantity which proved to be more efficient than current state-of-the-art methods on the analysed test cases. The last contribution extends the Stepwise Uncertainty Reduction framework to conservative estimates. We propose different strategies to sequentially select the next evaluation(s) that optimally reduces the uncertainty on conservative estimates. The methods proposed in this work have been tested and benchmarked against analytical test functions and reliability engineering test cases.

x

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Set estimation is a problem found in many statistical, engineering and scientific disciplines. In this work, we focus on the problem of estimating excursion sets of expensive to evaluate functions.

We consider the setup where a deterministic function can only be evaluated at few points and we are interested in recovering the set of inputs that leads to a certain range of output values. This framework can be found both in computer (Sacks et al., 1989) and physical experiments (see, e.g. Bayarri et al., 2009). One typical example is found in reliability engineering, where the safety of a system is often modelled with a numerical simulator that takes as input a set of physical measurements and returns the safety level of the system. In this setting, a quantity of interest is the set of all physical parameters that leads to a safe level. These numerical codes are often very expensive to evaluate and the computation of this value on a grid or on a high resolution space filling design is often impractical, if not impossible. Such numerical simulators are used in nuclear criticality (see, e.g., Chevalier et al., 2014a), earth science (see, e.g., Rohmer and Idier, 2012; Bayarri et al., 2009), cosmology (see, e.g., Seikel et al., 2012), climate science (see, e.g., French and Sain, 2013) and many others. In those situation practitioners often rely on surrogate models (Santner et al., 2003) that approximate the response surface of interest.

In this work we consider a Bayesian approach and we assume that the function is a realization of a Gaussian process (GP). We proceed then to construct a Gaussian process model from the available function evaluations. The GP model provides a posterior field and, by taking the excursion set of this posterior field, we obtain a random set. It is then possible to provide estimates for the set of interest and to quantify the uncertainties on these estimates by summarizing the posterior distribution of the random set. Here we consider only closed excursion sets, therefore we can exploit the the-

ory of random closed sets (RACS). The random closed set set-up provides several definitions of expectation. In this work we review the Vorob'ev expectation (see, e.g. Vorob'ev, 1984; Molchanov, 2005), revisited in the GP framework in Chevalier et al. (2013), we revise the distance average expectation (Baddeley and Molchanov, 1998) in the GP framework and we study conservative estimates introduced in Bolin and Lindgren (2015). The latter two estimates present computational challenges that could restrict their use in practice. In this work we analyse these challenges and we provide algorithms to reduce the computational cost of these techniques. The algorithms introduced here make distance average expectation and its related notion of distance average variability feasible at least for problems where the input domain is small. We also introduce a new algorithm that allows to compute conservative estimates in a general GP framework.

The remainder of this work is divided in 5 chapters. There is no strict reading order for the remaining chapters, nonetheless we suggest to start with Chapter 2, as it provides a brief introduction to the notions used throughout the manuscript. Chapters 3 and Chapters 4, 5 present relatively independent contributions. In particular, both Chapters 4 and 5 present results on conservative estimates and could thus be read sequentially. A brief summary of the contribution of each chapter closes this section.

- Chapter 2 contains a brief introduction to Gaussian process models and links these models with Bayesian set estimation. In the first part we mainly highlight the aspects of the literature on Gaussian processes that are useful for our work. We also briefly touch some aspects that are not directly related to this work, but that could provide future extensions. In the second part of this chapter we briefly review the theory of random closed sets with a specific focus on the concept of expectation of a random closed set. This notion is, in fact, later used to provide set estimates. Finally, the chapter introduces the link between Gaussian process modelling and the theory of random closed sets in the Bayesian set estimation framework. This chapter introduces the issues tackled in the main contributions of this work.

- Chapter 3 introduces an approach for computing quasi-realizations of the set from approximate realizations of the underlying field. Instead of simulating the field over high resolution grids, the field is simulated only at few *simulation points* optimally selected in order to obtain quasi-realizations of the set that are as close as possible to the true realizations. This method allows us to compute a new type of set estimate along with its uncertainty quantification. The method relies

on the distance average expectation, a concept introduced in the theory of random closed sets in Baddeley and Molchanov (1998), see also Molchanov (2005, Chapter 2). These types of set estimates require realizations of the random closed set discretized over a high resolution designs. Therefore they can be computationally very expensive for high resolutions. We show numerically that the quasi-realization technique brings significant computational savings for estimating distance average variability.

The simulation points technique can also be generally used when many samples of a set are required on high resolution designs. We show that this approach can also be adopted to obtain realizations of contour lines for a set and to estimate the distribution of an excursion set's volume.

- Chapter 4 starts by describing the computational challenges of conservative estimates under Gaussian random field priors. In particular we recall how these estimates are closely related with orthant probabilities of high dimensional Gaussian vectors. The main contribution of this chapter is an algorithm (GMC/GanMC) to approximate such probabilities. The method relies on a decomposition of the probability in an easy to compute biased part and in a remainder estimated with Monte Carlo methods. Two Monte Carlo methods are presented for this task: a standard MC estimator (GMC) and an asymmetric nested MC (GanMC) method. The latter is a newly introduced technique that exploits the simulations' computational costs to obtain more efficient estimators. In this chapter we introduce the anMC algorithm in a more generic fashion and we prove that, if the cost functions satisfy certain hypothesis, anMC achieves lower variance than MC for a fixed computational budget. Moreover we show that GMC/GanMC provides more efficient estimates than state-of-the-art algorithms for the estimation of high dimensional Gaussian orthant probabilities. Finally we close this chapter by showing how the GanMC method allows to efficiently compute conservative estimates of excursion sets under Gaussian process models.

- Chapter 5 focuses on sequential uncertainty reduction strategies for conservative estimates. In this chapter, we focus on Stepwise Uncertainty Reduction (SUR) strategies (Bect et al., 2012) which select function evaluations in order to minimize an uncertainty function. In particular, we introduce new uncertainty functions related to conservative estimates and we develop criteria to select points that minimize the future expected uncertainties. Three numerical test cases are presented

where the criteria are compared with state-of-the-art methods.

Chapters 3, 4 and 5 are either reproductions or extensions of the following papers:

- Azzimonti, D., Bect, J., Chevalier, C., and Ginsbourger, D. (2016). Quantifying uncertainties on excursion sets under a Gaussian random field prior. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1): 850–874. DOI:10.1137/141000749.

- Azzimonti, D. and Ginsbourger, D. (2016). Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation. *Under revision. Preprint available at hal-01289126.*

- Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J. and Richet, Y. (2016). Adaptive design of experiments for conservative estimation of excursion sets. *Preprint available at hal-01379642.*

In particular, the work on adaptive design of experiments for conservative estimates follows the work presented at conferences Azzimonti et al. (2015) and Azzimonti et al. (2016b).

The algorithms presented in Chapter 4 are implemented in the **R** package `ConservativeEstimates` available on GitHub.

Notation

In this work we use the following notation

$(\Omega, \mathfrak{F}, P)$ is a complete probability space with σ-algebra $\mathfrak{F}$ and probability measure P.

$\mathbb{R}$ denotes the real line and $\mathcal{B}_{\mathbb{R}}$ denotes the Borel σ-algebra on $\mathbb{R}$.

$\mathbb{E}$ denotes the expectation with respect to the underlying probability measure.

Random variables are denoted with non bold capital letters, e.g. $X \sim N(0,1)$.

$\mathbb{R}^d$ is the d-dimensional Euclidean space and $\mathcal{B}_{\mathbb{R}^d}$ denotes the Borel σ-algebra on $\mathbb{R}^d$.

$\mathbb{N}$ denotes the set of natural numbers not including zero.

Random vectors are denoted with bold capital letters, e.g. $\mathbf{Z} \sim N_d(\mathbf{m}, \Sigma)$.

$|M|$ denotes the determinant of a square matrix M.

Stochastic processes are denoted with non bold capital letter and an index set, e.g. $(Z_i)_{i \in I}$.

$\mathbb{R}^+$ denotes the set of non-negative real values.

$\mathfrak{m}$ is the mean function of a stochastic process, when it exists.

$\mathfrak{K}$ denotes a covariance kernel.

$\mathbb{X}$ is often a locally compact Hausdorff second countable topological space.

$\mathcal{F}$ is the family of closed sets in $\mathbb{X}$.

$\mathcal{K}$ is the family of compact sets in $\mathbb{X}$.

A^C is the complement of the set A.

$\mathbb{F}$ is a Banach space with norm $\|\cdot\|_{\mathbb{F}}$.

μ denotes a (usually σ-finite) measure on $\mathbb{X}$.

$\mathfrak{d}$ is a (pseudo-)distance between elements in $\mathbb{F}$.

$\mathbf{1}_A(\cdot)$ is the indicator function of the set A.

$\mathbf{X}_n$ usually denotes a design of experiments containing n points from the input space.

Γ denotes a random closed set. In most of this work $\Gamma = \{x \in \mathbb{X} : Z_x \in T\}$, where Z is a real valued stochastic process defined on $\mathbb{X}$ and $T \subset \mathbb{R}$.

$\mathrm{CE}_\alpha(\Gamma)$ denotes the conservative estimate at level α of Γ.

$\Gamma^\star$ denotes the set of values in $\mathbb{X}$ for which a deterministic function f takes values in $T \in \mathbb{R}$, where T and f are specified in each case.

$\mathcal{A}_n$ is the σ-algebra generated by the couples $(X_1, Z_{X_1}), \ldots, (X_n, Z_{X_n})$, where $X_1, \ldots, X_n$ are random elements in $\mathbb{X}$, the index space of the stochastic process Z.

$\mathcal{A}_n(\mathbf{X}_n)$ is the σ-algebra generated by the couples $(x_1, Z_{x_1}), \ldots, (x_n, Z_{x_n})$, where $\mathbf{X}_n = \{x_1, \ldots, x_n\} \in \mathbb{X}^n$ is a fixed, deterministic design of experiments.

$\mathbf{E}_m$ denotes a set of m points in the input space used as simulation points.

$\pi(t)$ denotes the orthant probability $P(\mathbf{X} \leq (t, \ldots, t))$, where $\mathbf{X} \sim N_d(\mathbf{m}, \Sigma)$ and $(t, \ldots, t) \in \mathbb{R}^d$.

Chapter 2

Gaussian processes and the theory of random closed sets

2.1 Context

The problem of estimating a set from a sample of points can be found in many parts of the statistical literature. Non parametric set estimation techniques, see, e.g., Devroye and Wise (1980); Baíllo et al. (2000) and references therein, exploit assumptions on the set to provide consistent estimates. If the set is a level set of a probability density function, a wealth of techniques has been developed in the literature for its inference relying on realizations sampled from this density, see, e.g., Hartigan (1987); Polonik (1995); Cuevas et al. (2006); Gayraud and Rousseau (2007) and references therein. If the probability density is replaced by a generic function and what is known are the values of this function at the given points, a direct technique involves the use of plug-in estimators (Molchanov, 1998; Cuevas et al., 2006). In this work we focus on sets of excursion for deterministic, expensive to evaluate functions. In this setup the points are locations in the input space where the deterministic function is evaluated. The set of interest is a subset of the input space that contains all locations that produce a particular range of response values. In this framework a plug-in approach approximates the function with an appropriate estimator and returns as set estimate the excursion set of the estimator (Ranjan et al., 2008). While this setup could be effective for predictions in some situations, it does not provide a quantification of the uncertainties introduced by the approximation. In this work we take a Bayesian approach and we use Gaussian process techniques to approximate the response function. From the posterior Gaussian process we derive the posterior distribution of the excursion set. We then exploit random set

techniques to summarize this distribution and quantify its variability.

In this chapter we review relationships between the theory of random closed sets and the problem of excursion set estimation under Gaussian processes priors. In Section 2.2 we introduce the basic definitions in Gaussian process modelling and we summarize how this technique provides efficient tools to approximate deterministic functions from few evaluations. We conclude this section by recalling the links between Gaussian process modelling and kriging and we briefly survey some useful extensions of the standard techniques. In Section 2.3 we present basic concepts from the theory of random closed sets (RACS) with a specific focus on the different definitions of expectation available for such objects. In this section we also review the concept of conservative estimates from the RACS point of view. Finally in Section 2.4 we provide a short introduction to how these two concepts are linked in the setup underlying the present work and we briefly recall some state-of-the-art methods in sequential uncertainty reduction for Gaussian process based excursion set estimation.

2.2   Gaussian processes: preliminaries

In this section we review the basic definitions and properties of Gaussian processes and their use for Bayesian set estimation.

2.2.1   Basic definitions

In this work, for simplicity, we always consider a complete probability space $(\Omega, \mathfrak{F}, P)$. Moreover we denote with $\mathcal{B}_{\mathbb{R}}$ the Borel σ-algebra on $\mathbb{R}$.

A real-valued random variable is a measurable function $Z : (\Omega, \mathfrak{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $Z^{-1}(B) \in \mathfrak{F}$, for each $B \in \mathcal{B}_{\mathbb{R}}$. Most of the following results still hold if we consider the field of complex numbers $\mathbb{C}$ with its Borel σ-algebra in place of the couple $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, however in what follows we mostly consider real valued variables, vectors and processes.

Let us start by recalling the definition of Gaussian random variable.

Definition 1 (Gaussian random variable). *A random variable* $Z : (\Omega, \mathfrak{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *has a Gaussian distribution if and only if it has a characteristic function of the form*

$$\phi_Z(t) = \mathbb{E}[e^{itZ}] = e^{itm - \frac{1}{2}\sigma^2 t^2}, \ t \in \mathbb{R} \tag{2.1}$$

for some $m \in \mathbb{R}$ *and* $\sigma^2 \geq 0$.

If Z is a Gaussian random variable then its moments $\mathbb{E}[|Z|^p]$ exist and are finite for all $p \in \mathbb{N}$. We denote the mean and variance with $m = \mathbb{E}[Z]$ and $\sigma^2 = \operatorname{Var}(Z)$ and we say that $Z \sim N(m, \sigma^2)$.

The cumulative distribution function (c.d.f.) of a Gaussian random variable Z is denoted with $\Phi(\cdot; m, \sigma^2)$. If $\sigma^2 > 0$, then Z has a probability density function (p.d.f.) defined as

$$\varphi(z; m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-m)^2}{2\sigma^2}}, \ z \in \mathbb{R}.$$

A Gaussian random variable has many important properties and numerous characterizations, we refer here to the book Patel and Read (1996, Chapter 4) for an extensive list. Here we only recall that a Gaussian distribution is uniquely determined by its first two moments and thus also by its mean m and variance σ^2. This is a direct consequence of Equation (2.1).

The notion of random variables can be naturally generalized to vectors. A random vector is a map $\mathbf{Z} : (\Omega, \mathfrak{F}) \to (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, where $\mathcal{B}_{\mathbb{R}^d}$ is the Borel σ-algebra on $\mathbb{R}^d$, $d \in \mathbb{N}$. We denote the vector as $\mathbf{Z} := (Z_1, \ldots, Z_d)$, where Z_i, $i = 1, \ldots, n$ are random variables. The moments of $\mathbf{Z}$ are defined by extending the univariate definitions. In particular, if vector $\mathbf{Z}$ has finite first and second moments, its expectation is defined as $\mathbb{E}[\mathbf{Z}] := (\mathbb{E}[Z_1], \ldots, \mathbb{E}[Z_d])$ and its covariance matrix as $\Sigma := [\sigma_{i,j}]_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}$, where

$$\sigma_{i,j} = \operatorname{Cov}(Z_i, Z_j) \text{ for all } i, j = 1, \ldots, d.$$

A covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$ is a positive semi-definite matrix, that is, it satisfies the following

$$\mathbf{z}^T \Sigma \mathbf{z} \geq 0 \text{ for all } \mathbf{z} \in \mathbb{R}^d.$$

Definition 2 (Gaussian vector). *We say that a random vector $\mathbf{Z} = (Z_1, \ldots, Z_d)$ is Gaussian if the linear combinations*

$$a_1 Z_1 + \cdots + a_d Z_d,$$

have univariate Gaussian distributions for any choice of coefficients $a_i \in \mathbb{R}$.

All moments of a Gaussian vector $\mathbf{Z}$ are finite, in particular we denote with $\mathbf{m} \in \mathbb{R}^d$ its expectation and with $\Sigma \in \mathbb{R}^{d \times d}$ its covariance matrix. We denote with $\Phi_d(\cdot; \mathbf{m}, \Sigma)$ the c.d.f. of a d-dimensional Gaussian vector with mean $\mathbf{m}$ and covariance matrix Σ and we say that $\mathbf{Z} \sim N_d(\mathbf{m}, \Sigma)$.

If the covariance matrix Σ is positive definite, i.e. it is non-singular, then $\mathbf{Z} \sim N_d(\mathbf{m}, \Sigma)$ possesses a p.d.f. with respect to the Lebesgue measure on $\mathbb{R}^d$ defined as

$$\varphi_d(\mathbf{z}; \mathbf{m}, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{z}-\mathbf{m})^T \Sigma^{-1} (\mathbf{z}-\mathbf{m})}, \ \mathbf{z} \in \mathbb{R}^d,$$

where $|\Sigma|$ denotes the determinant of the matrix Σ. We refer to the book Tong (2012) for a systematic exposition of Gaussian random vectors' properties. Here we only recall that $\mathbf{Z}$ is characterized by its first two moments: the mean $\mathbf{m} \in \mathbb{R}^d$ and the covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$, see, e.g., Tong (2012, Chapter 3).

Random vectors can be further generalized by taking infinite dimensional sequences of random variables, denoted as stochastic processes.

A real valued *stochastic process* is a collection of random variables $Z_i : (\Omega, \mathfrak{F}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$ with $i \in I$. Note that no assumptions are needed on I, the index space. In most of the present work, however, we restrict ourselves to $I = D \subset \mathbb{R}^d$, with D a compact subset. We say that a process is first order marginally integrable if $\mathbb{E}[|Z_i|] < \infty$ for all $i \in I$. In this case the process defines a mean function

$$\mathfrak{m} : I \to \mathbb{R},$$

where for each $i \in I$, $\mathfrak{m}(i) := \mathbb{E}[Z_i]$. If the process is also marginally square integrable, i.e. $\mathbb{E}[Z_i^2] < \infty$, for all $i \in I$, then it is possible to define the covariance kernel of Z as the function

$$\mathfrak{K} : I \times I \to \mathbb{R},$$

such that, for each $x, y \in I$, $\mathfrak{K}(x, y) := \mathbb{E}[(Z_x - \mathfrak{m}(x))(Z_y - \mathfrak{m}(y))]$.

No particular assumptions are required on the mean function, however a function $\mathfrak{K}$ is a covariance kernel of a stochastic process if and only if it is a symmetric positive definite function, i.e. if for any $n \in \mathbb{N}$, for each $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and for each choice of points $x_1, \ldots x_n \in I$ the following inequality holds

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathfrak{K}(x_i, x_j) \geq 0.$$

This is equivalent to the condition that all matrices $\Sigma = [\mathfrak{K}(x_i, x_j)]_{i,j=1,\ldots,n}$ are positive semi-definite for any choice of points $\{x_1, \ldots, x_n\} \subset I$.

First and second order properties are of key importance in the study of stochastic processes and are fundamental for Gaussian processes.

Definition 3 (Gaussian process)**.** *Consider a collection of real valued random variables $Z_i : (\Omega, \mathfrak{F}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$ with $i \in I$. We say that $(Z_i)_{i \in I}$ is a Gaussian process if any finite collection of those variables has a multivariate Gaussian distribution, i.e.*

$$(Z_{i_1}, \ldots, Z_{i_k}) \sim N_k(\mathbf{m}, \Sigma),$$

where $\mathbf{m} = (\mathbb{E}[Z_{i_1}], \ldots, \mathbb{E}[Z_{i_k}])$ and $\Sigma = [\sigma_{j,j'}]_{j,j'=1,\ldots,k}$, with $\sigma_{j,j'} = \mathrm{Cov}(Z_{i_j}, Z_{i_{j'}})$, for any $k \geq 1$ and for any finite subset $(i_1, \ldots, i_k) \subset I$.

Gaussian processes (GPs) defined on subsets of $\mathbb{R}^d$ with $d \geq 2$ are also called Gaussian random fields (GRFs), a notation mainly used in spatial statistics, see e.g. Chilès and Delfiner (2012). In this work those terms will be used interchangeably, with a preference for process when talking about the object defined on a generic space and for random field when the object is specifically defined on $\mathbb{R}^d, d \geq 2$.

From Definition 3, a Gaussian process is a stochastic process with all finite moments. We denote a Gaussian process $(Z_x)_{x \in I}$ with $Z \sim GP(\mathfrak{m}, \mathfrak{K})$ as it is uniquely determined in distribution[1] by its mean function $\mathfrak{m}$ and covariance kernel $\mathfrak{K}$. This can be shown by noticing that all finite dimensional distributions of a Gaussian process are Gaussian and for every choice of $(i_1, \ldots, i_k) \subset I$ the mean and covariance kernel define a mean vector and a covariance matrix, see, e.g., Rasmussen and Williams (2006, Chapter 2) and Adler and Taylor (2007, Chapter 1).

The covariance kernel $\mathfrak{K}$ plays a crucial role in the definition of a Gaussian process as it is closely linked with many of its properties. For example, in the centred case, the regularity of the kernel defines both the mean square regularity of the process and its path regularity, see e.g. Adler and Taylor (2007, Chapter 1) and Scheuerer (2009, Chapter 5) for more details. It is also possible to link invariances of the covariance kernel with path-wise invariances, see e.g. Ginsbourger et al. (2012, 2016) and references therein. In Chapter 3 this aspect is briefly commented on when discussing the regularity of Gaussian process quasi-realizations.

We close this section by reviewing the important notion of stationary process.

Definition 4 (Stationarity)**.** *Consider a real valued random field $(Z_x)_{x \in \mathbb{R}^d}$, defined on $\mathbb{R}^d$. We say that the field Z is* (strictly) stationary *if*

$$P(Z_{x_1} \leq t_1, \ldots, Z_{x_n} \leq t_n) = P(Z_{x_1+h} \leq t_1, \ldots, Z_{x_n+h} \leq t_n)$$

for any choice of $\{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, $\{t_1, \ldots, t_n\} \subset \mathbb{R}$ and $h \in \mathbb{R}^d$.

---

[1]with respect to the cylindrical σ-algebra, see, e.g., Dudley (2002, Chapter 12) for a proof in the general setup.

Moreover consider a random field $(Z_x)_{x \in \mathbb{R}^d}$ such that $\mathbb{E}[Z_x^2] < +\infty$, for all $x \in \mathbb{R}^d$, and denote with $\mathfrak{K}$ its covariance kernel. We say that Z is second order stationary *if*

$$\mathbb{E}[Z_x] = m \in \mathbb{R}, \ for \ all \ x \in \mathbb{R}^d$$
$$\mathfrak{K}(x + h, x) = \phi(h),$$

for some function $\phi : \mathbb{R}^d \to \mathbb{R}^+$.

Remark 1. *The definitions of strict stationarity and second order stationarity are equivalent for a Gaussian process, see, e.g., Chilès and Delfiner (2012, Chapter 1).*

A first example of stationary process defined on $\mathbb{R}$ can be built with a trigonometric polynomial with random amplitudes.

Example 1 (Trigonometric processes)**.** *Let us fix a frequency $\omega \in [0, \pi]$ and consider two random variables A, B uncorrelated, with zero mean and variance equal to $\lambda < +\infty$. The process $(Z_t)_{t \in \mathbb{R}}$ defined as*

$$Z_t = A\cos(\omega t) + B\sin(\omega t), \ t \in \mathbb{R}, \tag{2.2}$$

is a trigonometric process. This process has covariance kernel $\mathfrak{K}(t, t') = \lambda\cos(\omega(t - t'))$ and it is a second order stationary process. If A, B are Gaussian then Z is a Gaussian process and it is also strictly stationary.

Consider now n frequencies $\omega_k \in [0, \pi]$ and n random variables A_k, B_k uncorrelated, with zero mean with variances λ_k, for $k = 1, \dots, n$. The straightforward extension of Equation (2.2) is

$$Z_t = \sum_{k=1}^{n} \left(A_k \cos(\omega_k t) + B_k \cos(\omega_k t) \right). \tag{2.3}$$

This process has covariance kernel $\mathfrak{K}(t, t') = \sum_{k=1}^{n} \lambda_k \cos(\omega_k(t - t'))$ and it is also stationary.

The covariance kernel of the trigonometric process in Equation (2.3) is a function of $h = t - t'$ and can be represented as

$$\mathfrak{K}(t, t') = \phi(h) = \sum_{k=1}^{n} \lambda_k \cos(\omega_k h) = \Re\left(\sum_{k=1}^{n} \lambda_k \exp(i\omega_k h) \right),$$

where $\Re$ denotes the real part of an complex number.

The representation above can be seen as a Fourier transform of a discrete measure. This is actually a generic property of stationary fields. Bochner's theorem, stated below in the multivariate setting, makes this relationship explicit. See, e.g., Stein (1999, Chapter 2), Adler and Taylor (2007, Chapter 5) for a more detailed exposition.

Theorem 1 (Bochner's theorem). *A continuous function $k : \mathbb{R}^d \to \mathbb{R}$ is positive definite, i.e. it is a covariance kernel, if and only if there exists a finite Borel measure λ such that*

$$k(h) = \int_{\mathbb{R}^d} e^{2\pi i h^T \omega} d\lambda(\omega),$$

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

The measure λ is called *spectral measure* for k and we denote with $F(\omega) := \lambda\left(\prod_{i=1}^d (-\infty, \omega_i]\right)$ for $\omega = (\omega_1, \dots, \omega_d) \in \mathbb{R}^d$ the spectral distribution function. If F is absolutely continuous with respect to the Lebesgue measure then λ has spectral density.

The following is an example of stationary covariance kernel that has a spectral density.

Example 2 (Matérn family). *The covariance kernels with spectral density*

$$\phi(\omega) = \sigma^2 \frac{2\Gamma(\nu + 1/2)(2\nu)^\nu}{\pi^{-1/2}\theta^{2\nu}\Gamma(\nu)} \left(\frac{2\nu}{\theta^2} + 4\pi^2\omega^2\right)^{-(\nu+1/2)},$$

with $\nu, \sigma, \theta \geq 0$ are the Matérn *family of covariance kernels. This family was introduced in Matérn (1960) (see also Matérn, 1986), here we follow the parametrization in Rasmussen and Williams (2006). The covariance kernels defined by this family can be written as*

$$\mathfrak{K}_\nu(x, y) = \mathfrak{K}(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{\theta}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}r}{\theta}\right) \tag{2.4}$$

where $r = |x - y|$ denotes the Euclidean distance, $\nu, \sigma, \theta \in \mathbb{R}^+$, $\Gamma(\cdot)$ denotes the Gamma function and $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν. The parameter ν controls the field's mean square regularity and its sample path regularity. In fact, Z is k times mean square differentiable if and only if $\nu > k$ (Rasmussen and Williams, 2006). Moreover the sample paths of a Gaussian process Z with Matérn covariance kernel have continuous $\lceil \nu - 1 \rceil$ order derivatives almost surely if $\nu > \lceil \nu - 1/2 \rceil$, where $\lceil \cdot \rceil$ is the integer ceiling function (Handcock and Stein, 1993; Stein, 1999; Scheuerer, 2009). The parameter $\sigma \in \mathbb{R}^+$ is the standard deviation of the process and θ is called range or scale parameter.

The choice $\nu = p + \frac{1}{2}$, where p is a non-negative integer, makes the representation of the covariance kernel simpler. For $\nu = \frac{1}{2}$ we recover the exponential kernel,

$$\mathfrak{K}_{\nu=\frac{1}{2}}(x, y) = \mathfrak{K}(r) = \sigma^2 \exp\left(-\frac{r}{\theta}\right).$$

This covariance kernel defines the Ornstein-Uhlenbeck process. A Gaussian process Z with this covariance kernel is m.s. continuous but not differentiable; moreover its sample paths are a.s. continuous, see, e.g., Scheuerer (2009, Chapter 5).

For $\nu \to \infty$ we recover the Gaussian (or squared exponential) kernel

$$\mathfrak{K}(x,y)_{\nu \to +\infty} = \mathfrak{K}(r) = \sigma^2 \exp\left(-\left(\frac{r}{\theta}\right)^2 \right),$$

where $r = |x-y|$. It is a stationary and isotropic kernel. A Gaussian process Z with this covariance kernel is infinitely differentiable in a m.s. sense and has a.s. infinitely differentiable paths, see, e.g., Scheuerer (2009, Chapter 5).

In the following chapters we will often use Matérn covariance kernels with $\nu = \frac{3}{2}$ and $\nu = \frac{5}{2}$. These covariance kernels have the following representation

$$\mathfrak{K}_{\nu=\frac{3}{2}}(r) = \sigma^2 \left(1 + \frac{\sqrt{3}r}{\theta} \right) \exp\left(-\frac{\sqrt{3}r}{\theta} \right)$$

$$\mathfrak{K}_{\nu=\frac{5}{2}}(r) = \sigma^2 \left(1 + \frac{\sqrt{5}r}{\theta} + \frac{5r^2}{3\theta^2} \right) \exp\left(-\frac{\sqrt{5}r}{\theta} \right)$$

where $r = |x-y|$ denotes the Euclidean distance and $\sigma, \theta \in \mathbb{R}^+$. These covariance structures are mostly used in machine learning and in the computer experiments literature as they introduce smoothness assumptions met in many practical examples.

In practice stationary processes are important for Gaussian process modelling as we will see in Section 2.2.3, however there exists also non-stationary processes, below we report one notable example.

Example 3 (Wiener process). *A Wiener process $(W_x)_{x \geq 0}$ is a non-stationary Gaussian process with mean $\mathfrak{m}(x) = \mathbb{E}[W_x] = 0$, for each $x \in [0, +\infty)$, and covariance kernel $\mathfrak{K}(x,y) = \mathrm{Cov}(W_x, W_y) = \min(x,y), \ x, y \in [0, +\infty)$.*

The previous examples of covariance kernels, introduced in $\mathbb{R}$, can be generalized to $\mathbb{R}^d$. The Matérn family, for example, can be defined on $\mathbb{R}^d$ by considering the multivariate Euclidean distance r and by replacing the scalar products with vector products. This leads to an isotropic multivariate covariance kernel. It is also possible to extend these definitions to fields defined on $\mathbb{R}^d$ with other techniques, see Rasmussen and Williams (2006, Chapter 4.1.2) and Durrande (2011, Chapter 2). The applications presented

in following chapters make use of tensor product covariance kernels, where a kernel in d dimensions is obtained from the univariate kernel $\mathfrak{K}$ as

$$\mathfrak{K}^d(x,y) = \prod_{i=1}^{d} \mathfrak{K}(x_i, y_i),$$

where $x, y \in \mathbb{R}^d$ and x_i is the i-th coordinate of x. In the following chapters the superscript d is often removed when the dimension is obvious from the context. Note that the results presented in the following chapters are not restricted to this type of covariance kernels. In the applications we choose tensor product covariance kernels as they are readily implemented in the **R** package `DiceKriging` (Roustant et al., 2012). On the other hand, the developed approaches would be applicable to all kind of stationary and non-stationary covariance kernels.

2.2.2 Conditional expectations and distributions

The conditional expectation is a fundamental notion in probability and it is at the basis of Gaussian process modelling. In fact, conditional expectations provide a way of approximating the field knowing its value only at few observations. In this section we review the notions of conditional expectation and conditional distribution for Gaussian vectors and we show the straightforward extension to the case of Gaussian processes.

Proposition 1. *Consider the random vector* $\mathbf{Z} = (Z^1, Z^2)$, *where* $\mathbf{Z}, Z^1, Z^2$ *are vectors of dimension* d, d_1, d_2 *respectively, with* $d_1 + d_2 = d$. *Assume*

$$\mathbf{Z} \sim N_d \left(\begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix} \right)$$

and that the matrix $\Sigma_{2,2}$ *is invertible. For each* $z_2 \in \mathbb{R}^{d_2}$, *the conditional distribution of* $Z^1 \mid Z^2 = z_2$ *is a Gaussian vector with mean* $\mathbf{m}_{1|2}$ *and covariance matrix* $\Sigma_{1|2}$, *where*

$$\mathbf{m}_{1|2} = \mathbf{m}_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(z_2 - \mathbf{m}_2) \tag{2.5}$$

$$\Sigma_{1|2} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1} \tag{2.6}$$

Proof. Consider the following decomposition

$$Z_1 = \text{Proj}_{\text{span}(Z_2)}(Z_1) + R, \tag{2.7}$$

where $R := Z_1 - \text{Proj}_{\text{span}(Z_2)}(Z_1)$ and $\text{Proj}_{\text{span}(Z_2)}(Z_1)$ is the projection of Z_1 onto the linear subspace generated by Z_2 in the $L^2(\Omega, \mathfrak{F}, P)$ sense. The

projection can be developed as $\text{Proj}_{\text{span}(Z_2)}(Z_1) = \mathbf{m}_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(Z_2 - \mathbf{m}_2)$. Moreover the projection theorem guarantees that $\text{Proj}_{\text{span}(Z_2)}(Z_1)$ and R are orthogonal in $L^2(\Omega, \mathfrak{F}, P)$, i.e. uncorrelated. Observe that R is Gaussian, as it is a linear combination of Gaussian vectors. We also have $\mathbb{E}[R] = 0$ and

$$\text{Var}(R) = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}.$$

The projection has a Gaussian distribution and $(\text{Proj}_{\text{span}(Z_2)}(Z_1), R)$ are also jointly Gaussian, therefore we have that $\text{Proj}_{\text{span}(Z_2)}(Z_1)$ and R are independent. This also implies that R is independent from Z_2. By fixing $Z_2 = z_2$ in Equation (2.7) we obtain

$$Z_1 \mid Z_2 = z_2 \sim \mathbf{m}_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(z_2 - \mathbf{m}_2) + R,$$

which shows that $Z_1 \mid Z_2 = z_2$ is Gaussian and its mean and covariance matrix are described in Equation (2.5) and (2.6).

$\square$

Consider now a Gaussian process $(Z_x)_{x \in I}$ with mean $\mathfrak{m}$ and covariance kernel $\mathfrak{K}$ and denote with $\mathbf{X}_n = \{x_1, \ldots, x_n\} \subset I$ a set of n points in I and with $\mathbf{z}_n = (z_1, \ldots, z_n) \in \mathbb{R}^n$ a vector of real values. The conditional process $Z \mid Z_{\mathbf{X}_n} = \mathbf{z}_n$, where $Z_{\mathbf{X}_n} = (Z_{x_1}, \ldots, Z_{x_n})^T$, is a Gaussian process because all finite dimensional distributions are Gaussian. Moreover, by exploiting Equations (2.5) and (2.6) we can write the conditional mean at $x \in I$ and covariance kernel at $x, x' \in I$ as

$$\mathfrak{m}_n(x) = \mathfrak{m}(x) + \mathfrak{K}(x, \mathbf{X}_n)\mathfrak{K}(\mathbf{X}_n, \mathbf{X}_n)^{-1}(\mathbf{z}_n - \mathfrak{m}(\mathbf{X}_n)) \qquad (2.8)$$
$$\mathfrak{K}_n(x, x') = \mathfrak{K}(x, x') - \mathfrak{K}(x, \mathbf{X}_n)\mathfrak{K}(\mathbf{X}_n, \mathbf{X}_n)^{-1}\mathfrak{K}(\mathbf{X}_n, x'), \qquad (2.9)$$

where $\mathfrak{K}(x, \mathbf{X}_n) = (\mathfrak{K}(x, x_1), \ldots, \mathfrak{K}(x, x_n))$, $\mathfrak{K}(\mathbf{X}_n, x) = \mathfrak{K}(x, \mathbf{X}_n)^T$ and $\mathfrak{K}(\mathbf{X}_n, \mathbf{X}_n) = [\mathfrak{K}(x_i, x_j)]_{i,j=1,\ldots,n}$.

Denote with $\boldsymbol{\lambda}_n(x)$ the vector of weights $\boldsymbol{\lambda}_n(x) = \mathfrak{K}(\mathbf{X}_n, \mathbf{X}_n)^{-1}\mathfrak{K}(\mathbf{X}_n, x)$. The vector $\boldsymbol{\lambda}_n(x)$ is a function of $x \in I$ and depends on $\mathbf{X}_n$. If we fix the values of the field $Z_{\mathbf{X}_n} = \mathbf{z}_n$, then $Z_{x_0} \mid Z_{\mathbf{X}_n}$, the conditional field at a point $x_0 \in I$, is distributed as $\mathfrak{m}(x_0) + \boldsymbol{\lambda}_n(x)(Z_{\mathbf{X}_n} - \mathfrak{m}(\mathbf{X}_n))$. This notation allows us to see that $Z_{x_0} \mid Z_{\mathbf{X}_n}$ is an unbiased estimator for Z_{x_0} and an affine function of the field values at $\mathbf{X}_n$. Finally the projection theorem tells us that it is the linear unbiased predictor that minimizes the mean square error $\mathbb{E}[(Z_{x_0} - h(Z_{\mathbf{X}_n}))^2]$ among all affine functions of $Z_{\mathbf{X}_n}$.

The conditional field can be employed to approximate a response surface from few values. Consider the case where the response is known at points $\mathbf{X}_n \subset I$ with values $\mathbf{z}_n \in \mathbb{R}^n$. If we assume that the response is a realization

of a Gaussian process then we can predict the response at a new value $x_0 \in I$ with the conditional expectation. In the next section we briefly review the relationships between kriging and Gaussian processes models. This is further discussed in Section 2.4 where the process is used as a tool to estimate excursion sets of deterministic expensive to evaluate functions.

2.2.3 Kriging and Gaussian process modelling

In many scientific fields, such as geophysics, hydrology, climatology among many others, some deterministic phenomena can be modelled as deterministic functions

$$f : D \subset \mathbb{R}^d \to \mathbb{R}$$

where $f(x)$ is the response of a system given a set of inputs $x \in D$. Here we are interested in the setting where this function is known, or can be evaluated, only at a finite number of points $\mathbf{X}_n = \{x_1, \ldots, x_n\} \subset D$ and we are interested in giving an approximation of the function f at any point $x \in D$.

Many methods have been proposed to approximate a function f starting from the vector of evaluations $\mathbf{f}_n = (f(x_1), \ldots, f(x_n)) \in \mathbb{R}^n$, see, e.g. Simpson et al. (2001); Wang and Shan (2007) for extensive reviews. Here we restrict ourselves to Bayesian models with a Gaussian process prior (O'Hagan, 1978; Sacks et al., 1989).

The methods used here were first explored in the geostatistics literature (Krige, 1951; Matheron, 1963) with the technique currently known as kriging. In this framework the unknown function f is usually a spatial field in low dimensions observed only at few points. In second order geostatistics, the function is assumed to be a realization of a square-integrable random field $(Z_x)_{x \in D}$, not necessarily Gaussian, with mean $\mathfrak{m}$ and covariance kernel $\mathfrak{K}$. We denote with $Z_{\mathbf{X}_n} \in \mathbb{R}^n$ the vector of field values at $\mathbf{X}_n$ and we predict the field Z at a point $x \in \mathbb{R}^d$ with

$$\widehat{Z}_x = \boldsymbol{\lambda}(x)^T Z_{\mathbf{X}_n},$$

where $\boldsymbol{\lambda}(x)$ is a vector of weights that depend on the point $x \in D$. Since the function f is assumed to be a realization of the field Z, the predictor $\widehat{Z}_{x_0}$ is a random variable as the values $Z_{\mathbf{X}_n}$ are random. For a point $x_0 \in D$ we choose the weights $\boldsymbol{\lambda}(x_0)$ such that $\widehat{Z}_{x_0}$ is an unbiased estimator of Z_{x_0} and

$$\epsilon = \text{Var}[(Z_{x_0} - \widehat{Z}_{x_0})] \tag{2.10}$$

is minimized. The quantity ϵ is called variance of the residual or kriging error variance.

We first assume that the mean of the field Z is a known function $\mathfrak{m}$ and we denote with $\mathfrak{K}$ the covariance kernel of the field. In this setting the minimization procedure leads to the Simple Kriging (SK) weights $\boldsymbol{\lambda}_{SK}(x) = \mathfrak{K}(\mathbf{X}_n, \mathbf{X}_n)^{-1}\mathfrak{K}(\mathbf{X}_n, x)$ for each $x \in D$, where we follow the notation introduced in Equation (2.8) for $\mathfrak{K}(x, \mathbf{X}_n)$ and $\mathfrak{K}((\mathbf{X}_n, \mathbf{X}_n))$. If we fix the evaluations of the function, i.e. $Z_{\mathbf{X}_n} = \mathbf{f}_n$, then the field $\widehat{Z}$ has mean and covariance kernels given by the following *kriging equations*

$$\mathfrak{m}_n(x) = \mathfrak{m}(x) + \boldsymbol{\lambda}_{SK}^T(x)(\mathbf{f}_n - \mathfrak{m}(\mathbf{X}_n))$$
$$\mathfrak{K}_n(x, x') = \mathfrak{K}(x, x') - \boldsymbol{\lambda}_{SK}^T(x)\mathfrak{K}(\mathbf{X}_n, x').$$

The kriging equations have many useful properties such as interpolation of the observations and lead to an unbiased estimator, see Chilès and Delfiner (2012, Chapter 3) for more details. If the field Z is Gaussian then the simple kriging equations are the conditional mean and covariance of $Z \mid Z_{\mathbf{X}_n} = \mathbf{f}_n$.

In many practical cases, however the mean of the field is not known and needs to be estimated from the observations. A common assumption is that the mean can be modelled as

$$\mathfrak{m}(x) = \sum_{j=1}^{k} \beta_j g_j(x),$$

where $\beta_1, \ldots, \beta_k \in \mathbb{R}$ are unknown coefficients and $g_1, \ldots, g_k$ are known basis functions. In this setting the minimization of ϵ as introduced in Equation (2.10) leads to the Universal Kriging (UK) equations, see Chilès and Delfiner (2012, Chapter 3) for a detailed exposition on the subject.

In practice here we will often use universal kriging equations where an unknown constant mean function is assumed. This setting is also called Ordinary Kriging (OK), see, e.g. Chilès and Delfiner (2012, Chapter 3).

Recently, in the context of computer experiments, further forms of kriging, such as blind kriging (see, e.g. Joseph et al., 2008) or models including qualitative factors (see, e.g. Qian et al., 2008) were introduced.

The classic kriging procedures can be seen in a Bayesian way, see, e.g. Handcock and Stein (1993); Neal (1998); Gramacy and Polson (2011) and references therein. In a Bayesian framework we define an appropriate prior distribution over the space of possible functions and we study the posterior distribution given the function evaluations. The prior distribution encodes the information on f known before any evaluation takes place, as for example its regularity, symmetries or invariances. By selecting a Gaussian process as prior distribution on functions it is possible to easily encode this information.

As it was briefly sketched in the previous sections, Gaussian processes are particularly convenient because by choosing an appropriate covariance

kernel $\mathfrak{K}$ we obtain a prior that puts mass only on specific types of functions. For example it is possible to select a prior that puts mass only on functions with a certain regularity by appropriately choosing the kernel $\mathfrak{K}$. Often the covariance kernel is chosen from a parametric family, such as the Matérn family for example. A parametric family of kernels allows much flexibility on the regularity of the prior function space, however it needs some procedure to select the hyper-parameters. In a fully Bayesian framework (see, e.g. Kennedy and O'Hagan, 2001; Qian and Wu, 2008, and references therein) the hyper-parameters are not fixed but hyperprior distributions are fixed. This often leads to full posterior distribution that do not have Gaussian distribution and require Monte Carlo methods to be approximated. Good treatments of this subject can be found in Gibbs (1998); Neal (1998) and references therein.

In this work we mostly follow an "empirical Bayes" approach where an improper uniform prior is selected for each trend coefficient and the covariance parameters are estimated from the data with maximum likelihood estimators. Given the evaluations $\mathbf{f}_n$ the conditional field represents a distribution over the space of functions. Since, under these assumptions, the conditional process is still Gaussian, it is possible to compute analytically its mean and covariance kernel, see Equations (2.8) and (2.9). We could summarize the full posterior distribution with the posterior mean which gives an estimate for f. In this work we focus on estimating excursion sets of f and on quantifying the uncertainty on those estimates, thus the main object of interest is the posterior distribution itself. In fact the posterior distribution naturally defines a distribution on random closed sets. In Section 2.4 we explain this concept in more details and we show how random closed set expectations can be used to obtain estimates for the actual set of interest. First however let us review the notion of random closed set and some of its properties in the next section.

2.3 Random closed sets (RACS) and their expectations

In probability and statistics the main theoretical framework where the problem of estimating sets is studied is the theory of random closed sets. In this section we give a brief review of the concepts from this theory that are needed in this work.

Let us start by introducing a formal definition for a random set. Consider a space $\mathbb{X}$ which for our purposes will be a locally compact Hausdorff second

countable topological space[2]. The definitions introduced in this section hold in this general setting, however in the rest of this work we mostly choose $\mathbb{X} = D \subset \mathbb{R}^d$, where D is a compact subset. We denote with $\mathcal{F}$ the family of closed sets in $\mathbb{X}$ and with $(\Omega, \mathfrak{F}, P)$ a complete probability space.

Definition 5. *We say that a map $\Gamma : \Omega \to \mathcal{F}$ is a random closed set if, for any compact set $K \in \mathcal{K} \subset \mathcal{F}$*

$$\{\omega : \Gamma(\omega) \cap K \neq \emptyset\} \in \mathfrak{F}, \tag{2.11}$$

where $\mathcal{K}$ is the family of compact sets in $\mathbb{X}$.

The condition in Equation (2.11) implies that Γ is a measurable map between the measurable space $(\Omega, \mathfrak{F})$ and the space of closed sets $\mathcal{F}$ endowed with the σ-algebra $\mathcal{B}_{\mathcal{F}}$ generated by sets of the form $\{\Gamma \cap K \neq \emptyset\}$, with $K \in \mathcal{K}$.

Example 4. *Consider a random element Z defined on $\mathbb{X}$, measurable with respect to the Borel σ-algebra on $\mathbb{X}$, then the set $\Gamma = \{Z\} \subset \mathbb{X}$ is a random closed set. This set is called a* singleton.

Example 5. *Consider the random set $\Gamma \subset \mathbb{R}$ defined as*

$$\Gamma(\omega) = \begin{cases} [0,1] & \omega \in \Omega_1 \\ \{0,1\} & \omega \in \Omega_2 \end{cases},$$

where $P(\Omega_1) = P(\Omega_2) = 1/2$. Γ is a random closed set, in fact for each compact set $K \subset \mathbb{R}$ we have the following: $\{\Gamma \cap K \neq \emptyset\} = \Omega_1$ if $K \subset (0,1)$, $\{\Gamma \cap K \neq \emptyset\} = \Omega_1 \cup \Omega_2$ if K contains a neighbourhood of 0 or 1, $\{\Gamma \cap K \neq \emptyset\} = \Omega_2$ if $K = \{0\}$ or $K = \{1\}$ and $\{\Gamma \cap K \neq \emptyset\} = \emptyset$ if $K \subset [0,1]^C = \mathbb{R} \setminus [0,1]$.

Example 6. *Consider a real valued random process $(Z_x)_{x \in \mathbb{X}}$ with continuous sample paths almost surely and the set*

$$\Gamma = \{x \in \mathbb{X} : Z_x \in T\}, \tag{2.12}$$

where $T \subset \mathbb{R}$ is a closed set. Since the field Z is almost surely continuous, Γ is almost surely closed as it is the preimage of a closed set under a continuous function. Moreover, let us denote with $\Omega' = \Omega \setminus \Omega_0$ the set of events where the paths of Z are continuous in D, then the set $\Gamma : \Omega' \to \mathcal{F}$ is a properly defined random closed set. In fact, $\{\Gamma \cap K \neq \emptyset\} = \{\exists x \in K : Z_x \in T\}$ is measurable

---

[2]See Singer and Thorpe (2015) for a good introduction to these concepts, in particular Chapter 2 for the definition of this notion.

as Z is a measurable stochastic process. In the following chapters the sets as in Equation (2.12) will implicitly be defined on Ω', unless otherwise noted. Furthermore, we will focus mainly on the particular cases obtained for $T = T_\ell = (-\infty, t]$ or $T = T_u = [t, +\infty)$, with $t \in \mathbb{R}$ a fixed threshold. In the first case we have the set of excursion below or sojourn set $\Gamma_l = \{x \in \mathbb{X} : Z_x \leq t\}$ while in the second case we have the excursion above or simply excursion set $\Gamma_u = \{x \in \mathbb{X} : Z_x \geq t\}$.

2.3.1 Choquet theorem and distribution of a random closed set

The probability distribution of a random closed set Γ is determined by the probabilities of the form $P(\Gamma \in A)$, for all $A \in \mathcal{B}_\mathcal{F}$. In general the σ-algebra $\mathcal{B}_\mathcal{F}$ is too large to obtain a meaningful representation. However, as for random variables, we can restrict to a particular family of generators. The σ-algebra $\mathcal{B}_\mathcal{F}$ is generated by sets of the form $\{F \cap K \neq \emptyset\}$, for $K \in \mathcal{K}$ and $F \in \mathcal{F}$, thus we can consider the probabilities $P(\{\Gamma \cap K \neq \emptyset\})$.

Definition 6 (Capacity functional for RACS)**.** *The functional $T_\Gamma : \mathcal{K} \longrightarrow [0, 1]$ defined as*

$$T_\Gamma(K) := P(\{\Gamma \cap K \neq \emptyset\})$$

is called capacity functional *of the random closed set Γ.*

For a general random closed set Γ, the capacity functional T_Γ is not a probability measure as it is subadditive and not additive, i.e.

$$T_\Gamma(K_1 \cup K_2) \leq T_\Gamma(K_1) + T_\Gamma(K_2), \qquad K_1, K_2 \in \mathcal{K}.$$

If Γ is a singleton $\Gamma = \{Z\}$, where Z is a random variable defined on $\mathbb{X}$, then $T_\Gamma(K) = P(Z \in K)$ is a probability measure. In fact this is a necessary and sufficient condition, see Molchanov (2005, Chapter 1.1), for a proof of this statement and more examples.

The concept of capacity functional is more general than Definition 6. In fact, consider a real-valued functional T defined on $\mathcal{K}$, we say that T is a capacity functional if

1. $T(\emptyset) = 0$;

2. $0 \leq T(K) \leq 1$, for all $K \in \mathcal{K}$;

3. $T(K_n) \downarrow T(K)$ if $K_n \downarrow K$ in $\mathcal{K}$;

4. $\Delta_{K_n} \dots \Delta_{K_1} T(K) \leq 0$, with $n \geq 1$, $K_1, \dots, K_n \in \mathcal{K}$, where

$$\Delta_{K_1} T(K) = T(K) - T(K \cup K_1) \qquad \text{and}$$
$$\Delta_{K_n} \dots \Delta_{K_1} T(K) = \Delta_{K_{n-1}} \dots \Delta_{K_1} T(K) - \Delta_{K_{n-1}} \dots \Delta_{K_1} T(K \cup K_n).$$

If a functional satisfies property 3 it is *upper semi-continuous* and if it satisfies property 4 it is *completely alternating*.

A capacity functional defined on $\mathcal{K}$ satisfying the properties above uniquely determines the distribution of a random closed set. This result is the Choquet theorem, reported but not proven here. For more details and proofs see Molchanov (2005, Chapter 1.3).

Theorem 2 (Choquet). *A capacity functional $T : \mathcal{K} \to [0,1]$ such that $T(\emptyset) = 0$ is the capacity functional of a unique random closed set on $\mathbb{X}$ if and only if it is completely alternating and upper semi-continuous.*

In the following chapters we will focus on random sets generated as excursion sets of almost surely continuous stochastic processes or random fields. In particular we will be interested in the sets introduced in Example 6, i.e. $\Gamma = \{x \in \mathbb{X} : Z_x \in T\}$, where $(Z_x)_{x \in \mathbb{X}} \sim GP(\mathfrak{m}, \mathfrak{K})$ with a.s. continuous paths and $T \subset \mathbb{R}$ is a closed set. This set has capacity functional $T_\Gamma(K) = P(\{\Gamma \cap K \neq \emptyset\}) = P(\exists x \in K : Z_x \in T)$.

Particular cases of the set Γ in Equation (2.12) are obtained for $T_\ell = (-\infty, t]$ or $T_u = [t, +\infty)$ with $t \in \mathbb{R}$ a fixed threshold. In the first case the excursion below t, or "sojourn set", has Choquet capacity $T_{\Gamma_l}(K) = P(\{\Gamma_l \cap K \neq \emptyset\}) = P(\inf_{x \in K} Z_x \leq t)$. In the second case we have the excursion above t, or simply "excursion set", which has Choquet capacity $T_{\Gamma_u}(K) = P(\{\Gamma_u \cap K \neq \emptyset\}) = P(\sup_{x \in K} Z_x \geq t)$.

2.3.2 Expectations of random closed sets

A random closed set Γ defines a distribution over the space of closed sets, which is uniquely described by its capacity functional. In this work we are interested in estimating Γ and quantifying uncertainties on this estimate, thus we focus on ways to summarize this distribution. As for the case of real-valued or vector-valued random elements we would like to make use of the moments of the distribution. In particular here we focus on the expected value, a rough but practical summary for a distribution. The definition of this object however poses several problems. First of all we are interested in averaging a random object which takes values in the space of closed sets $\mathcal{F}$. Since this space is not a vector space, the definition of an integral over

$\mathcal{F}$ is not straightforward. Moreover, as recalled above, the capacity of a random closed set is a probability measure if and only if the set is a singleton (Molchanov, 2005). These properties make it hard to define an expectation in the usual sense, however it is still possible to define expectations by resorting to other properties of random sets.

The most commonly used set expectation in probability is the Aumann expectation (Kudo, 1954; Aumann, 1965; Artstein and Vitale, 1975) which is based on the concept of selections. We refer to Molchanov (2005, Chapter 2.1) for an extensive treatment of the subject. Here instead we focus on different definitions of expectation coming from the linearisation approach (Molchanov, 2005, Chapter 2.2).

The idea behind linearisation is to transform the problem of expectation in the space of closed sets $\mathcal{F}$ into a problem in a Banach space, where the definition of expectation is more straightforward. Let us consider a Banach space $\mathbb{F}$ with a norm $\|\cdot\|_{\mathbb{F}}$ and denote with

$$\xi : \mathcal{F} \to \mathbb{F}$$

a functional that associates each closed set $F \in \mathcal{F}$ with a measurable element $\xi_F \in \mathbb{F}$ with respect to the Borel σ-algebra generated by the topology of the norm in $\mathbb{F}$. In particular given a random closed set Γ we associate ξ_Γ which is a measurable random element in $\mathbb{F}$. If the expectation of ξ_Γ is properly defined, then $\mathbb{E}[\xi_\Gamma] = \zeta$ for some $\zeta \in \mathbb{F}$. This element, however does not necessarily correspond to ξ_F for some $F \in \mathcal{F}$. We can nonetheless approximate the expectation of Γ with the set F that has the "closest" functional ξ_F to ζ. The notion of "closeness" is fixed by defining a distance $\mathfrak{d} : \mathbb{F} \times \mathbb{F} \to \mathbb{R}^+$, that associates to each couple $(\xi, \zeta) \in \mathbb{F} \times \mathbb{F}$ a non-negative value. Often $\mathfrak{d}$ is only a pseudo-distance as $\mathfrak{d}(\xi, \zeta) = 0$ also for some $\xi \neq \zeta$. In what follows we refer to $\mathfrak{d}$ as a distance and we specify when it is only a pseudo-distance.

The notion of "closeness" introduced by $\mathfrak{d}$ is used to find an element in ξ_F, with $F \in \mathcal{F}$, that achieves the minimum distance $\mathfrak{d}(\xi_F, \mathbb{E}[\xi_\Gamma])$. This optimization problem is usually not easy as the search space $\mathcal{F}$ does not enjoy convexity properties, therefore we need to restrict the search to a parametric family of sets $\mathcal{F}_\rho \subset \mathcal{F}$, for some ρ in a parameter space.

A linearisation approach is thus defined by the following elements

- a Banach space $\mathbb{F}$ with norm $\|\cdot\|_{\mathbb{F}}$;

- a functional $\xi : \mathcal{F} \to \mathbb{F}$;

- a pseudo-distance $\mathfrak{d}(\xi, \zeta)$ between elements in $\mathbb{F}$;

- a parametric family of sets $\mathcal{F}_\rho$.

Given a random closed set Γ, then we define a candidate expectation for Γ as

$$\mathbb{E}_{\xi,\mathbb{F},\mathfrak{d}}[\Gamma] \in \arg\min_{F\in\mathcal{F}_\rho} \mathfrak{d}(\mathbb{E}[\xi_\Gamma], \xi_F),$$

where the minimizer $\mathbb{E}_{\xi,\mathbb{F},\mathfrak{d}}[\Gamma]$ is either unique or chosen according to an ad hoc criterion.

By properly selecting these four elements we obtain different definitions of expectation for a random closed set.

The Vorob'ev expectation

A first application of the linearisation approach is the Vorob'ev expectation (Vorob'ev, 1984; Stoyan and Stoyan, 1994; Molchanov, 2005). Throughout this section we denote with μ a $\sigma-$finite measure defined on $\mathbb{X}$; let us define the functional ξ that associates to each set its indicator function.

$$\xi_F := \mathbf{1}_F.$$

In particular consider the random closed set Γ and assume that $\mathbb{E}[\mu(\Gamma)] < \infty$. Examples where this assumption is verified are when $\mathbb{X} = D$ is a compact subset of $\mathbb{R}^d$ with μ the usual Lebesgue measure, or when $\mathbb{X}$ is a measurable space with μ a probability measure on $\mathbb{X}$. The expectation of the indicator function $\mathbf{1}_\Gamma$ defines

$$p_\Gamma : \mathbb{X} \to [0,1],$$

where $p_\Gamma(x) := \mathbb{E}[\mathbf{1}_\Gamma(x)] = P(x \in \Gamma)$. This function is called the *coverage function* of Γ. A random closed set Γ is almost surely equal to a deterministic closed set $F \in \mathcal{F}$, if and only if the coverage function of Γ is an indicator function. This property shows that it is typically not possible to find $F \in \mathcal{F}$ such that $\mathbb{E}[\xi_\Gamma] = \xi_F$. In order to find the "closest" indicator function to p_Γ we need to define a pseudo-distance. Consider the space $L^1(\mu)$ and $\xi_F, \xi_G \in L^1(\mu)$, let us define the pseudo-distance

$$\mathfrak{d}(\xi_F, \xi_G) = \left| \int_{\mathbb{X}} (\xi_F(x) - \xi_G(x)) d\mu(x) \right| \tag{2.13}$$

$$= \left| \int_{\mathbb{X}} (\mathbf{1}_F(x) - \mathbf{1}_G(x)) d\mu(x) \right|.$$

In order to minimize $\mathfrak{d}$ we restrict ourselves to the parametric family of sets $\mathcal{F}_\rho = \{Q_\rho : \rho \in [0,1]\}$ where

$$Q_\rho = \{x \in \mathbb{X} : p_\Gamma(x) \geq \rho\}. \tag{2.14}$$

The sets Q_ρ are closed because p_Γ is upper semi-continuous (Molchanov, 2005). Consider now a random closed set Γ, we have that

$$\arg \min_{F \in \mathcal{F}_\rho} \mathfrak{d}(\mathbb{E}[\mathbf{1}_\Gamma], \mathbf{1}_F) = \arg \min_{F \in \mathcal{F}_\rho} \left| \int_{\mathbb{X}} (\mathbb{E}[\mathbf{1}_\Gamma(x)] - \mathbf{1}_F(x)) \, d\mu(x) \right|.$$

Since μ is $\sigma-$finite and $\mathbf{1}_\Gamma(x)$ is positive we can exchange expectation and integral obtaining

$$\arg \min_{F \in \mathcal{F}_\rho} |\mathbb{E}[\mu(\Gamma)] - \mu(F)| = \arg \min_{\rho \in [0,1]} |\mathbb{E}[\mu(\Gamma)] - \mu(Q_\rho)|. \qquad (2.15)$$

Equation (2.15) leads to the following definition (Vorob'ev, 1984) of Vorob'ev expectation.

Definition 7 (Vorob'ev expectation)**.** *The Vorob'ev expectation of a random closed set Γ is the set*

$$\mathbb{E}_V[\Gamma] = \mathbb{E}_{\mathbf{1}_{(\cdot)}, L^1(\mu), \mathfrak{d}}[\Gamma] = Q_{\rho_V},$$

where ρ_V is either the unique minimizer of $\arg \min_{\rho \in [0,1]} |\mathbb{E}[\mu(\Gamma)] - \mu(Q_\rho)|$ or the solution to

$$\mu(Q_\rho) \leq \mathbb{E}[\mu(\Gamma)] \leq \mu(Q_{\rho_V}), \qquad (2.16)$$

for all $\rho > \rho_V$.

The computation of the Vorob'ev expectation is greatly simplified in practice because the expected volume of Γ is often simple to compute. In fact, as a consequence of Robbins' theorem (see, e.g. Robbins, 1945; Molchanov, 2005), it is possible to write the expected volume of Γ as the integral of the coverage function. Since $p_\Gamma \geq 0$ we can exchange the order of integration and we obtain

$$\mathbb{E}[\mu(\Gamma)] = \int_\Omega \int_{\mathbb{X}} \mathbf{1}_{\Gamma(\omega)}(x) d\mu(x) P(d\omega)$$

$$= \int_{\mathbb{X}} \mathbb{E}[\mathbf{1}_\Gamma(x)] d\mu(x) = \int_{\mathbb{X}} p_\Gamma(x) d\mu(x).$$

In the following chapters we will extensively use the coverage function because it can be computed analytically for RACS Γ coming from excursions of Gaussian processes, as shown in Example 8.

The parametric sets Q_ρ as defined in Equation (2.14) are also called *Vorob'ev quantiles.* They are an important family for the problem of set estimation under Gaussian priors because, due to existence of an analytic

formula for p_Γ, those sets are both conceptually simple and very fast to compute.

Let us compute the Vorob'ev expectations for the random closed sets introduced in examples 5 and 6. The second example in particular is central for the remainder of this work.

Example 7. *Consider the random closed set Γ defined as in Example 5, consider a compact subset D in $\mathbb{R}$ such that $[0,1] \subsetneq D$ and define the measure μ as the usual Lebesgue measure on $\mathbb{R}$ restricted to D. The coverage function for Γ is*

$$p_\Gamma(x) = \begin{cases} 1 & \text{if } x = 0 \text{ or } x = 1 \\ \frac{1}{2} & \text{if } x \in (0,1) \\ 0 & \text{if } x \in [0,1]^C = D \setminus [0,1] \end{cases}.$$

Moreover we have that $\mathbb{E}[\mu(\Gamma)] = \frac{1}{2}$. The Vorob'ev quantile with the closest volume is $Q_{1/2} = \{x \in D : p_\Gamma(x) \geq 1/2\} = [0,1]$. Thus $\rho_V = 1/2$ and the Vorob'ev expectation of Γ is $Q_{\rho_V} = [0,1]$.

Example 8. *Consider a real valued Gaussian process $(Z_x)_{x \in \mathbb{X}} \sim GP(\mathfrak{m}, \mathfrak{K})$ with almost surely continuous paths and assume that $\mu(\mathbb{X}) < \infty$. For a fixed $t \in \mathbb{R}$ denote with Γ the set*

$$\Gamma = \{x \in \mathbb{X} : Z_x \geq t\}.$$

The coverage function of this set is

$$p_\Gamma(x) = P(x \in \Gamma) = P(Z_x \geq t) = \Phi\left(\frac{\mathfrak{m}(x) - t}{\sqrt{\mathfrak{K}(x,x)}}\right),$$

where $\Phi(\cdot)$ is the c.d.f. of a standard Gaussian distribution. This coverage function is properly defined for all $x \in \mathbb{X}$ and generates the following Vorob'ev quantiles

$$Q_\rho = \left\{x \in \mathbb{X} : \Phi\left(\frac{\mathfrak{m}(x) - t}{\sqrt{\mathfrak{K}(x,x)}}\right) \geq \rho\right\}.$$

In this case the coverage function has a closed form expression, therefore we can compute numerically the integral

$$\mathbb{E}[\mu(\Gamma)] = \int_{\mathbb{X}} p_\Gamma(x) d\mu(x),$$

which is finite as $\mu(\mathbb{X}) < \infty$. We can select the quantile ρ_V with a numerical optimization procedure. Such quantile is either chosen such that $\mu(Q_{\rho_V}) = \mathbb{E}[\mu(\Gamma)]$ or it is the smallest ρ_V that satisfies the inequality in Equation (2.16).

We now introduce a flexible notion of distance between two random closed sets. While this concept is here introduced for random closed sets, it can also be used to quantify the distance between a Vorob'ev quantile and the corresponding random closed set.

Definition 8 (Expected distance in measure). *Consider two random closed sets* $\Gamma_1, \Gamma_2 \subset \mathbb{X}$. *The* expected distance in measure *between* Γ_1 *and* Γ_2 *with respect to the measure* μ *is*

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

where $\Gamma_1 \Delta \Gamma_2 = (\Gamma_1 \setminus \Gamma_2) \cup (\Gamma_2 \setminus \Gamma_1)$.

This distance quantifies how far apart are two sets by measuring the non overlapping parts. In particular if two sets Γ_1, Γ_2 are equal almost surely then $d_\mu(\Gamma_1, \Gamma_2) = 0$. Conversely if $P(\Gamma_1 \cap \Gamma_2 = \emptyset) = 1$ then $d_\mu(\Gamma_1, \Gamma_2) = \mathbb{E}[\mu(\Gamma_1)] + \mathbb{E}[\mu(\Gamma_2)]$.

The expected distance in measure is useful in many set estimation problems. In Chapter 3 it is used to quantify the distance between quasi-realizations of Γ and Γ itself under a Gaussian process model. However this distance can also be used to measure the variability associated with an estimator of Γ. In Chapter 5 we show that the Vorob'ev quantiles achieve the minimum distance with respect to Γ among all sets of equal measure. This properties gives a stronger justification for the use of this family when computing conservative estimates, see Section 2.3.3. Moreover in Chapter 5 we use the variability notion induced by the expected distance in measure to adaptively reduce the uncertainty on conservative estimates.

The Vorob'ev expectation is a linearised expectation based on the metric defined in Equation (2.13). Changing the metric gives different results, for example, if we choose the uniform metric on the space of indicator functions then the expectation becomes $Q_{\rho_{\mathrm{unif}}}$, where ρ_{unif} is chosen as

$$\arg \min_{\rho \in [0,1]} \mathfrak{d}_{\mathrm{unif}}(\mathbb{E}[\mathbf{1}_\Gamma], \mathbf{1}_{Q_\rho}) = \arg \min_{\rho \in [0,1]} \sup_{x \in \mathbb{X}} \left| p_\Gamma(x) - \mathbf{1}_{Q_\rho}(x) \right|.$$

The minimum is achieved for $\rho = 1/2$, thus this procedure always selects the "median" quantile $Q_{1/2}$.

Other possible choices are the $L^1(\mu)$ or the $L^2(\mu)$ metric on the indicator functions. With the $L^1(\mu)$ metric the Vorob'ev procedure selects a level ρ

that minimizes the function

$$\mathfrak{d}_{L^1(\mu)}(\mathbb{E}[\mathbf{1}_\Gamma], \mathbf{1}_{Q_\rho}) = \big\|\mathbb{E}[\mathbf{1}_\Gamma] - \mathbf{1}_{Q_\rho}\big\|_1$$
$$= \left(\int_{Q_\rho}(1 - p_\Gamma(x))d\mu(x) + \int_{Q_\rho^C} p_\Gamma(x)d\mu(x)\right)$$
$$= \left(\mathbb{E}[\mu(\Gamma)] + \int_{Q_\rho}(1 - 2p_\Gamma(x))d\mu(x)\right).$$

The $L^2(\mu)$ metric instead selects ρ, a minimizer of

$$\mathfrak{d}_{L^2(\mu)}(\mathbb{E}[\mathbf{1}_\Gamma], \mathbf{1}_{Q_\rho}) = \big\|\mathbb{E}[\mathbf{1}_\Gamma] - \mathbf{1}_{Q_\rho}\big\|_2$$
$$= \left(\int_{\mathbb{X}} p_\Gamma^2(x)d\mu(x) + \int_{Q_\rho}(1 - 2p_\Gamma(x))d\mu(x)\right).$$

When a minimum exists, the two procedures select the same minimizer ρ^*.

The distance average expectation

The linearisation approach is more general than the Vorob'ev expectation. One generalization of this expectation is the distance average expectation (Baddeley and Molchanov, 1998; Molchanov, 2005). While the Vorob'ev approach selects the indicator function of a set as the functional ξ, the distance average approach uses more general distance functions.

Let us denote with ϱ a metric on the space $\mathbb{X}$. Here we follow Baddeley and Molchanov (1998) and we define a distance function as follows

Definition 9. *Consider the space of closed non empty sets $\mathcal{F}' = \mathcal{F} \setminus \{\emptyset\}$, the* distance function *of a point $x \in \mathbb{X}$ to a set $F \in \mathcal{F}$ is defined as*

$$d : \mathbb{X} \times \mathcal{F}' \to \mathbb{R}^+,$$

such that $d(x, F) := \inf_{y \in F} \varrho(x, y)$.

This type of distance function is also called *metric distance function*. In particular if $\mathbb{X} = \mathbb{R}^d$ and ϱ is the Euclidean metric we have the Euclidean distance function. The properties introduced in this section can be extended to more general distance functions, see Baddeley and Molchanov (1998) for an extensive list. Since our contributions use only metric distance functions we restrict ourselves to Definition 9.

For each $F \in \mathcal{F}$, we consider the functional $\xi_F : \mathcal{F} \to \mathbb{F}$ defined as

$$\xi_F = d : (\cdot, F) \in \mathbb{X} \times \mathcal{F} \to \mathbb{R}^+,$$

where $d(x, F)$ is the distance between a point $x \in \mathbb{X}$ and the set $F \in \mathcal{F}'$.

Given a random set Γ we have now that the distance function $d(x, \Gamma)$ is a random variable for each $x \in \mathbb{X}$. We can compute its expectation for each $x \in \mathbb{X}$, which leads to the following definition.

Definition 10 (Mean distance function). *Consider a distance function d and a random closed set Γ. The function*

$$\bar{d} : \mathbb{X} \to \mathbb{R}^+,$$

defined as $\bar{d}(x) := \mathbb{E}[d(x, \Gamma)]$ is called the mean distance function *of Γ.*

The mean distance function in general is not a distance function of a specific set. Moreover, it can be proven (Baddeley and Molchanov, 1998) that for a given set Γ, $\bar{d}(x) = \mathbb{E}[d(x, \Gamma)]$ is a distance function if and only if Γ is deterministic. The following examples show two cases where $\bar{d}$ is not a distance function.

Example 9. (a) *Consider a random set Γ and the distance function $\mathbf{1}_{\Gamma^c}$. This is not a metric distance function, as in Definition 9, however it can be defined as a distance function in a wider sense, see Baddeley and Molchanov (1998). The mean distance function is equal to $\bar{d}(x) = 1 - p_\Gamma(x)$, where p_Γ is the coverage function. As seen in the previous section this function is an indicator if and only if Γ is deterministic.*

(b) *Consider the random set Γ defined in Example 5 and the Euclidean distance function on $\mathbb{R}$. Its mean distance function is*

$$\bar{d}_\Gamma(x) = \begin{cases} -x & \text{if } x < 0 \\ \frac{1}{2}x & \text{if } x \in [0, \frac{1}{2}) \\ \frac{1}{2}(1 - x) & \text{if } x \in [\frac{1}{2}, 1) \\ x - 1 & \text{if } x \geq 1 \end{cases} \tag{2.17}$$

which is not a distance function to a set.

Following the general linearisation approach we need to define a pseudo-metric over the space of distance functions to compute the distance average expectation.

Let us denote with $\mathfrak{d} : \mathbb{F} \times \mathbb{F} \to \mathbb{R}^+$ a pseudo-metric between two distance functions. Here we will restrict ourselves to the case $\mathbb{X} = D \subset \mathbb{R}^d$, where D

is a compact subset and we review two choices for $\mathfrak{d}$. The first is the uniform metric on the space of continuous functions defined on a compact space

$$\mathfrak{d}_{\text{unif}}(d(\cdot, F), d(\cdot, G)) := \sup_{x \in \mathbb{X}} |d(x, F) - d(x, G)|. \qquad (2.18)$$

With the metric $\mathfrak{d}_{\text{unif}}$ then we select the space $\mathbb{F}$ as the space of continuous functions on $\mathbb{X}$.

The second choice is $\mathfrak{d}_{L^2}$, the L^2 distance with the standard Lebesgue measure on $\mathbb{X}$:

$$\mathfrak{d}_{L^2}(d(\cdot, F), d(\cdot, G)) := \left(\int_{\mathbb{X}} (d(x, F) - d(x, G))^2 dx \right)^{1/2}.$$

In this case the space $\mathbb{F}$ is $L^2(\mathbb{X})$. This is also the metric chosen in Chapter 3 where a new uncertainty quantification measure using the distance average expectation is introduced.

For a generic distance function d and a metric $\mathfrak{d}$, the distance average expectation is defined as follows.

Definition 11 (Distance average expectation)**.** *Consider a random closed set Γ, denote with $\bar{d}$ its mean distance function. Further consider the family of sets*

$$D_u = \{x \in \mathbb{X} : \bar{d}(x) \leq u\}, \ u \in \mathbb{R}.$$

The distance average expectation of Γ is the set

$$\mathbb{E}_{DA}[\Gamma] = D_{u_{DA}} = \{x \in \mathbb{X} : \bar{d}(x) \leq u_{DA}\}$$

where $u_{DA} \in \arg\min_{u \in \mathbb{R}} \mathfrak{d}(\bar{d}, D_u)$ is either the unique minimizer or the infimum of the minimizers.

Example 10. *Consider the random closed set Γ introduced in Example 5, its mean distance function was computed in Equation (2.17). Let us fix d as the Euclidean distance. The distance average expectation with respect to the metric $\mathfrak{d}_{unif}$ as defined in Equation (2.18) is equal to $[-\frac{1}{12}, \frac{1}{6}] \cup [\frac{5}{6}, \frac{13}{12}]$.*

By resorting on the distance functions it is possible to define a notion of variability (see Baddeley and Molchanov, 1998) for the distance average expectation.

Definition 12 (Distance average variability)**.** *Consider a random closed set Γ and its distance average expectation $\mathbb{E}_{DA}[\Gamma]$ with respect to the distance function d and the metric $\mathfrak{d}$. The distance average variability of $\mathbb{E}_{DA}[\Gamma]$ is equal to*

$$\text{DAV}(\Gamma) = \mathbb{E}\left[\mathfrak{d}(\bar{d}, d(\cdot, \Gamma))^2 \right].$$

Example 11. *Consider the random closed set Γ defined in Example 5. In Example 10 we showed that the distance average expectation with respect to the Euclidean distance d and the uniform metric $\mathfrak{d}_{unif}$ is the set $\mathbb{E}_{DA}[\Gamma] = [-\frac{1}{12}, \frac{1}{6}] \cup [\frac{5}{6}, \frac{13}{12}]$. The distance average variability is equal to*

$$\text{DAV}(\Gamma) = \mathbb{E}\left[\mathfrak{d}_{unif}(\bar{d}, d(\cdot, \Gamma))\right] = \mathbb{E}\left[\left(\sup_{x \in \mathbb{X}} |\bar{d}(x) - d(x, \Gamma)|\right)^2\right].$$

By using Equation (2.17) for $\bar{d}$ and exploiting the fact that Γ is a random set we obtain $\text{DAV}(\Gamma) = \frac{1}{16}$.

In Chapter 3 the distance average variability is used to define a new way to quantify the uncertainty for excursion set estimation with Gaussian process priors.

2.3.3 Other types of estimates: conservative approach

The random set expectations defined in the previous section provide a wide class of set estimates. Moreover for each expectation there is a variability notion that quantifies their uncertainties. However, they are not directly equipped with probabilistic statements on the estimate. For example in some situations we are interested in estimates that with a certain probability are included in the random set Γ. These types of problem are often found in reliability engineering, where the quantity of interest might be a set of safe configurations for a system. In this case we are not only interested in estimating the set, but we would like to control in a probabilistic sense the overestimation of the set, in order to select only safe configurations. The Vorob'ev expectation and the distance average do not provide this type of probabilistic statement. The concept of *conservative estimates* was developed to address this issue. In this section we review this concept in the general framework of random closed sets. This idea was developed for the particular case of excursion sets of latent models in the papers by Bolin and Lindgren (2015) and French and Sain (2013).

First of all let us introduce the notion of *inclusion functional* that allows this type of probabilistic evaluations on the set.

Definition 13 (Inclusion functional). *Consider a random closed set Γ and denote with $\mathcal{K}$ the family of compact sets in $\mathbb{X}$. The* inclusion functional *of Γ is*

$$I_\Gamma : \mathcal{K} \to [0, 1],$$

defined, for each $K \in \mathcal{K}$, as $I_\Gamma(K) := P(K \subset \Gamma)$.

This functional is related to the Choquet capacity functional of the complement of Γ by the following equality

$$I_\Gamma(K) = 1 - T_{\Gamma^C}(K).$$

For the remainder of this section we denote with μ a σ−finite measure on $\mathbb{X}$. We can now define the notion of conservative estimate for Γ.

Definition 14 (Conservative estimate). *Consider the random compact set Γ and a family of compact sets $\mathfrak{C}$. The conservative estimate at level α of Γ is the set $\mathrm{CE}_\alpha(\Gamma)$, defined as*

$$\mathrm{CE}_\alpha(\Gamma) \in \arg\max_{K \in \mathfrak{C}}\{\mu(K) : I_\Gamma(K) \geq \alpha\}.$$

In general, the family $\mathfrak{C}$ might not enjoy properties that guarantee a unique solution $\mathrm{CE}_\alpha(\Gamma)$. Often, the optimization procedure in Definition 14 is not feasible unless we parametrize the family $\mathfrak{C}$. One rather natural choice is the family of Vorob'ev quantiles, which we recall are defined as $Q_\rho = \{x \in \mathbb{X} : p_\Gamma(x) \geq \rho\}$, for $\rho \in [0, 1]$. This family was also chosen in French and Sain (2013) and Bolin and Lindgren (2015) for estimates of excursion sets with latent Gaussian models. In Chapter 4 we use this parametric family in the application section and in Chapter 5 we provide elements for a justification of this choice.

Example 12. *Consider the random set Γ introduced in Example 5. Let us compute the conservative estimate at level α for Γ with the Vorob'ev quantiles Q_ρ as family $\mathfrak{C}$ and with the Lebesgue measure on D as μ. Recall that the coverage function of Γ was computed in Example 7. Then the family $\mathfrak{C}$ only contains the sets $Q_0 = D$, $Q_{1/2} = [0, 1]$ and $Q_1 = \{0, 1\}$. We have then*

$$
\begin{aligned}
I_\Gamma(Q_0) &= P(D \subset \Gamma) = 0 & \mu(Q_0) &= \mu(D) \\
I_\Gamma(Q_{1/2}) &= P([0, 1] \subset \Gamma) = 1/2 & \mu(Q_{1/2}) &= 1 \\
I_\Gamma(Q_1) &= P(\{0, 1\} \subset \Gamma) = 1 & \mu(Q_1) &= 0.
\end{aligned}
$$

If we fix $\alpha = 0.95$, then $\mathrm{CE}_\alpha(\Gamma) = Q_1 = \{0, 1\}$.

If uniqueness is guaranteed, the set $\mathrm{CE}_\alpha(\Gamma)$ is the set of higher measure among the ones that have probability of being inside Γ at least equal to α. For example, if α is chosen relatively high, e.g. $\alpha = 0.95, 0.99$, then $\mathrm{CE}_\alpha(\Gamma)$ will result in a set that is highly likely to be inside the random set.

A particular linearised expectation

The conservative estimate does not involve the notion of expectation and therefore it is not in general a linearised expectation. The usual linearised expectation however can be modified to include the conservative notion. These two approaches in general result in different estimates, however we show in this section that in a particular case they are equivalent.

Let us first introduce the conservative linearised expectation. Consider a random compact set Γ and denote with

$$\xi : \mathcal{K} \to \mathbb{F}$$

a functional that associates to each compact set $K \in \mathcal{K}$ an element in a Banach space $\mathbb{F}$.

In the linearised expectation approach we choose a pseudo-metric $\mathfrak{d}$ on $\mathbb{F}$ and a parametric family where the metric is optimized. Let us fix a generic metric $\mathfrak{d}$ and the parametric family

$$\mathcal{K}_{\rho,\alpha} = \{K_\rho \in \mathcal{K} : I_\Gamma(K_\rho) \geq \alpha\}. \tag{2.19}$$

In this framework the linearised expectation becomes the set K_{ρ^*}, where

$$K_{\rho^*} \in \arg \min_{K_\rho \in \mathcal{K}_{\rho,\alpha}} \mathfrak{d}(\mathbb{E}[\xi_\Gamma], \xi_{K_\rho}). \tag{2.20}$$

This estimate encodes the conservative aspect by restricting the search space to the sets that have probability at least α of being contained in the random set of interest.

If we choose $\xi_K = \mathbf{1}_K$, $\mathfrak{d}$ as the metric introduced in Equation (2.13) we obtain a modified version of the Vorob'ev expectation where the minimization is conducted on the subset of sets such that $I_\Gamma(K) \geq \alpha$.

A specific set of choices for the functional ξ, the metric $\mathfrak{d}$ and the family $\mathcal{K}_{\rho,\alpha}$ leads to an expectation K_{ρ^*} which is equivalent to the conservative estimate $\mathrm{CE}_\alpha(\Gamma)$.

Lemma 1. *Assume that K_{ρ^*} is the unique minimizer of the linearised expectation defined in Equation* (2.20), *with the following choices*

- *$\xi_K = \mathbf{1}_K$;*

- *$\mathfrak{d}$ is the pseudo-metric introduced in Equation* (2.13)*;*

- *$\mathcal{K}_{\rho,\alpha}$ is the family defined in Equation* (2.19), *with K_ρ the Vorob'ev quantiles defined in Equation* (2.14).

If $\alpha > \rho_V$, where ρ_V is the Vorob'ev level from Definition 7, then $\mathrm{CE}_\alpha = K_{\rho*}$.

Proof. First of all notice that $\alpha > \rho_V$ implies $P(K_{\rho_V} \subset \Gamma) < \alpha$. In fact

$$P(K_{\rho_V} \subset \Gamma) = 1 - T_{\Gamma^c}(K_{\rho_V})$$
$$\leq 1 - T_{\Gamma^c}(\{x\}) = p_\Gamma(x),$$

for any $x \in K_{\rho_V}$, where the inequality comes from the monotonicity of the capacity functional T_{Γ^c}. Moreover Definition 7 implies that there exists $x^* \in K_{\rho_V}$ such that $p_\Gamma(x^*) = \rho_V$. We then have $P(K_{\rho_V} \subset \Gamma) \leq p_\Gamma(x^*) = \rho_V < \alpha$. This property implies that $K_{\rho_V} \notin \mathcal{K}_{\rho,\alpha}$. Since the family of Vorob'ev quantiles is nested we have $K_{\rho^*} \subsetneq K_{\rho_V}$. Finally the conditions

$$|\mathbb{E}[\mu(\Gamma)] - \mu(K_{\rho^*})| \geq |\mathbb{E}[\mu(\Gamma)] - \mu(K_{\rho_V})|$$

and $\mu(K_{\rho^*}) \leq \mu(K_{\rho_V})$ imply that the minimizer of the distance $\mathfrak{d}(\mathbb{E}[\xi_\Gamma], \xi_{K_\rho})$ is the maximizer of the measure $\mu(K_\rho)$ in the search space $\mathcal{K}_{\rho,\alpha}$. $\qquad\square$

In Chapter 5 the notions introduced in this section, in particular the concept of conservative estimates, are used to estimate excursion sets with Gaussian process modelling. In that chapter we also introduce methods to sequentially reduce the uncertainty on conservative estimates.

In the next section we close the current chapter by revisiting the random set estimates in the framework of Gaussian process modelling and excursion set estimation.

2.4   State of the art in Bayesian set estimation with Gaussian Process priors

In this section we introduce the Bayesian set estimation procedures analysed in this work. Let us consider a deterministic continuous function

$$f : D \subset \mathbb{R}^d \to \mathbb{R}$$

defined on D, a compact subset of $\mathbb{R}^d$ and let us assume that the function is known only at the points $\mathbf{X}_n = \{x_1, \ldots, x_n\} \subset D$ where it takes values $\mathbf{f}_n = (f(x_1), \ldots, f(x_n))$. Let us fix a closed set $T \subset D$, the main objective of this work is to study estimation procedures for the preimage of T under f, i.e. the set

$$\Gamma^\star = \{x \in D : f(x) \in T\}.$$

A traditional set estimation approach to this problem is to consider the subset of points $\{x_{i_1}, \ldots, x_{i_k}\} \subset \mathbf{X}_n$ such that $f(x_{i_j}) \in T$. In some settings, this set of points can be regarded as a sample of points randomly selected in $\Gamma^\star$. It is then possible to use traditional estimators such as the convex hull or Devroye Wise type (Devroye and Wise, 1980) estimators. These methods are well studied in the literature (see, e.g., Baíllo et al., 2000, and references therein), however, they usually require a quite large number of points in the interior of the excursion to obtain reasonable estimates. In our framework we often work with a very limited number of points and, while we can often select the points $\mathbf{X}_n$, we cannot control the subset $\{x_{i_1}, \ldots, x_{i_k}\}$, thus it is not possible to make distributional assumptions on those points. These two properties make traditional estimators not well suited for our problem.

The set of interest $\Gamma^\star$ is completely defined by the function f, thus it is reasonable to exploit the information available on f to estimate $\Gamma^\star$. In Section 2.2.3 we reviewed a Bayesian approach to approximate functions from few evaluations. Here we follow this framework and select a prior Gaussian process $(Z_x)_{x \in D} \sim GP(\mathfrak{m}, \mathfrak{K})$. Since the function f is assumed continuous, we select a prior distribution that puts mass only on functions that are continuous. As briefly introduced in Section 2.2, the regularity can be controlled with an appropriate choice of the covariance kernel. For example if $d = 1$, we can choose a Matérn covariance kernel with $\nu = p + 1/2$, $p \geq 2$ and we obtain a prior Gaussian process with sample paths at least a.s. continuous. Figure 2.1a shows an example of Gaussian Process model. The function f is evaluated at $n = 8$ points, the prior Gaussian process has a constant mean and a Matérn covariance kernel with $\nu = 3/2$, $\theta = 0.2$ and $\sigma = 1$. The figure shows the posterior mean function $\mathfrak{m}_n$, the 95% point-wise posterior prediction intervals $\mathfrak{m}_n(x) \pm 1.96\sqrt{\mathfrak{K}(x,x)}$ for $x \in D$ and 20 realizations of the posterior field. The prior Gaussian process naturally defines a prior distribution on excursion sets, in fact the set

$$\Gamma = \{x \in D : Z_x \in T\}$$

is a random closed set, as shown in Example 6. In the example of Figure 2.1 the set of interest $\Gamma^\star$ is the excursion above $t = 0.5$.

Let us consider now the process Z conditioned on the observations $Z_{\mathbf{X}_n} = \mathbf{f}_n$. Since the prior is Gaussian we have that posterior distribution of the conditional field is $GP(\mathfrak{m}_n, \mathfrak{K}_n)$, where $\mathfrak{m}_n$ and $\mathfrak{K}_n$ are the conditional mean and covariance kernel defined in Equations (2.8) and (2.9).

A simple plug-in approach for estimating $\Gamma^\star$ uses the mean of the posterior process $\mathfrak{m}_n$ and approximates $\Gamma^\star$ with

$$\Gamma_{\text{plug-in}} = \{x \in D : \mathfrak{m}_n(x) \in T\}.$$
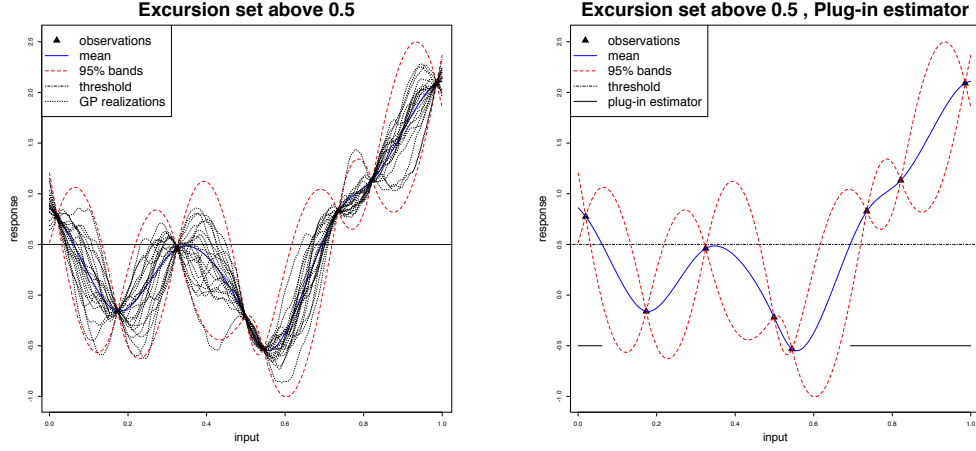
(a) Gaussian process model with $n = 8$ observations and posterior GP realizations.

(b) Gaussian process model and plug-in estimator $\Gamma_{\text{plug-in}}$.

Figure 2.1: Example of Gaussian process regression. The prior field has a constant mean function and a Matérn covariance kernel, see Equation (2.4), Section 2.2.1, with $\nu = 3/2$, $\theta = 0.2$, $\sigma = 1$.

Figure 2.1b shows the plug-in estimator for $\Gamma^\star$ in the example introduced previously. The estimator $\Gamma_{\text{plug-in}}$ is straightforward to implement, however it does not provide a reliable quantification of uncertainty.

Instead here we notice that the prior field Z induces a prior distribution on sets, described by the random closed set

$$\Gamma = \{x \in D : Z_x \in T\}, \tag{2.21}$$

as shown in Example 6. By exploiting the Bayesian assumption we have then that the posterior process induces a posterior distribution for Γ given $\mathbf{f}_n$. Under the previous assumptions on the prior process, the posterior field has a.s. continuous paths. This implies that the posterior set is a properly defined random closed set. We can now use the tools introduced in Sections 2.3.2 and 2.3.3 to obtain estimates for $\Gamma^\star$.

2.4.1   Vorob'ev quantiles and expectation for set estimation

Let us start by reviewing the Vorob'ev quantiles and the related expectation (Vorob'ev, 1984; Molchanov, 2005) in the Gaussian process setting. The article Chevalier et al. (2013) revised the Vorob'ev expectation in the Gaussian

process setup. The Ph.D. thesis Chevalier (2013) also introduces this notion in Chapter 4.2. The form of the posterior excursion set Γ in Equation (2.21) greatly simplifies the computations for this expectation.

The notion of coverage function in this setting becomes

$$
\begin{aligned}
p_{\Gamma,n} &= P_n(x \in \Gamma) \\
&= P_n(Z_x \in T) = \int_T \varphi_{m=\mathfrak{m}_n(x), \sigma^2=\mathfrak{K}_n(x,x)}(s)ds,
\end{aligned}
\tag{2.22}
$$

where $\varphi_{m,\sigma^2}(s)$ is the Gaussian density with mean m and variance σ^2 and $P_n(\cdot) = P(\cdot \mid Z_{\mathbf{X}_n} = \mathbf{f}_n)$ is the posterior probability given the function evaluations $\mathbf{f}_n$.

In particular if $T = (-\infty, t]$, with $t \in \mathbb{R}$ the coverage function becomes

$$
\begin{aligned}
p_{\Gamma,n}(x) &= P_n(Z_x \leq t) \\
&= P_n\left(\frac{Z_x - \mathfrak{m}_n(x)}{\sqrt{\mathfrak{K}_n(x,x)}} \leq \frac{t - \mathfrak{m}_n(x)}{\sqrt{\mathfrak{K}_n(x,x)}} \right) = \Phi\left(\frac{t - \mathfrak{m}_n(x)}{\sqrt{\mathfrak{K}_n(x,x)}} \right),
\end{aligned}
\tag{2.23}
$$

where $\Phi(\cdot)$ is the c.d.f. of the standard Gaussian distribution.

The representations of $p_{\Gamma,n}$ in Equations (2.22) and (2.23) make the computation of the Vorob'ev quantiles Q_ρ, defined in Equation (2.14), numerically very fast. In practice, the Vorob'ev quantiles are discretized over a design $G = \{u_1, \ldots, u_r\} \subset D$, with r much larger than n. The set of points G can be chosen, for example, as a space filling design such as a Latin Hypercube Sample (LHS), a low discrepancy sequence such as the Sobol' or Halton sequences, a lattice or a grid, see, e.g. Franco (2008); Pronzato and Müller (2012) and references therein. Given a design G, we evaluate the posterior mean $\mathfrak{m}_n$ and the posterior kernel $\mathfrak{K}_n$ at G and, by evaluating the Gaussian c.d.f., we can approximate Q_ρ on G with

$$
Q_\rho^G = \left\{ u \in G : \Phi\left(\frac{t - \mathfrak{m}_n(u)}{\sqrt{\mathfrak{K}_n(u,u)}} \right) \geq \rho \right\}.
\tag{2.24}
$$

The definition of the Vorob'ev expectation requires a $\sigma-$finite measure μ on the space D. Common choices are probability measures on the space D or the usual Lebesgue measure, if the space D is compact. The first choice is particularly common in reliability engineering, see, e.g. Dubourg et al. (2011); Bect et al. (2012). In most of the examples in this work, the measure μ is the Lebesgue measure on the compact subset D.

As introduced in Definition 7, the procedure to obtain the Vorob'ev expectation requires the computation of the posterior expected measure of Γ
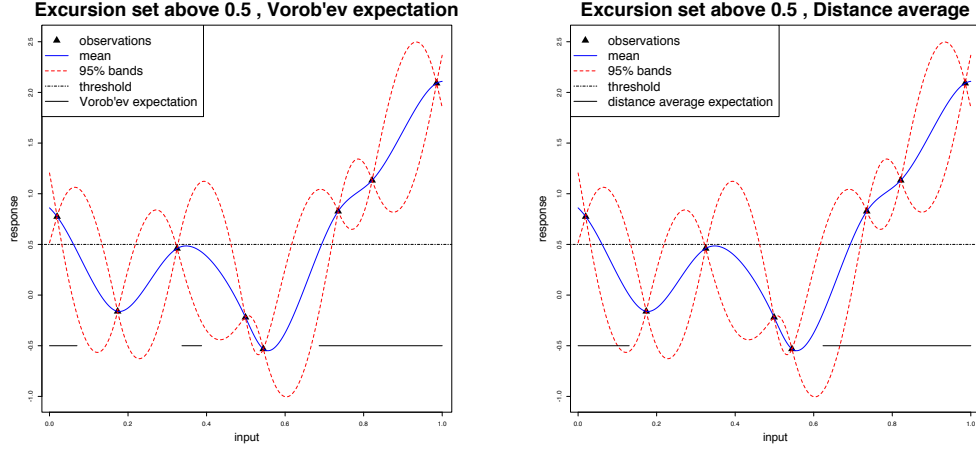
(a)   Vorob'ev   expectation   with μ (b) Distance average expectation with
Lebesgue measure on $\mathbb{R}$.   Euclidean distance function and L^2 metric.

Figure 2.2: Example of set estimation with Gaussian process regression. The posterior process defines a posterior distribution of excursion sets, here summarized with two different expectations.

given $\mathbf{f}_n$. Here we denote with $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \mid Z_{\mathbf{X}_n} = \mathbf{f}_n]$ and we notice that the coverage function as defined in Equation (2.22) greatly simplifies the computation of this quantity. In fact we have

$$\mathbb{E}_n[\mu(\Gamma)] = \int_D p_{\Gamma,n}(x) d\mu(x).$$

This quantity can be approximated numerically with quadrature methods. The numerical computation is efficient as the evaluation of the coverage function at one point relies on fast approximations of the Gaussian distribution function. In the **R** computing language, for example, fast approximations of the integrand are obtained with the function `pnorm`, which is based on an efficient Fortran implementation, see Cody (1993).

The Vorob'ev expectation can then be computed with a simple optimization over the parameter $\rho \in [0, 1]$. In **R**, the package `KrigInv` (Chevalier et al., 2014c) implements this procedure with the function `vorob_threshold`. Figure 2.2a shows the Vorob'ev expectation computed with μ as the Lebesgue measure for the example introduced in Figure 2.1.

2.4.2 Distance average expectation

In Section 2.3.2 we reviewed the concept of distance average expectation for random closed sets. This notion has been used in image analysis (Lewis et al., 1999; Jankowski and Stanberry, 2010, 2012), in medical (Ayala et al., 2005) and environmental (Zhao et al., 2011) applications among others. In Zhang et al. (2008) this notion was introduced to predict spatial quantiles, using techniques coming from spatial statistics community. To the best of our knowledge, however, the notion of distance average expectation has not been used for Bayesian set estimation under Gaussian process priors. In Chapter 3 we introduce this concept and we propose a new uncertainty quantification procedure exploiting the distance average variability, reviewed in Definition 12.

Figure 2.2b shows an example of distance average expectation on the problem described in the previous section. As opposed to the Vorob'ev expectation, the distance average quantities are estimated from posterior realizations of the excursion set Γ given the observations $\mathbf{f}_n$. This procedure can be very costly as it involves conditional simulations of a Gaussian process. In Chapter 3 we present a faster method to obtain approximate realizations of the process also called quasi-realizations. These quasi-realizations are optimal in the sense that they minimize the posterior expected distance in measure with Γ. This contribution reduces the computational cost of sample based methods for random sets in the Gaussian process framework.

2.4.3 Conservative estimates

In this work we consider the conservative estimates CE_α as defined in Definition 14. This type of set estimates were introduced in the Bayesian excursion set estimation framework in Bolin and Lindgren (2015). In this paper $\mathfrak{C}$ is the family of Vorob'ev quantiles $\mathfrak{C}_\rho = \{Q_\rho : \rho \in [0, 1]\}$, where Q_ρ is the posterior Vorob'ev quantile of Γ at level ρ and the measure μ is chosen as the Lebesgue measure over D, thus leading to the conservative estimate that selects the set with the largest volume such that $I_{\Gamma,n}(Q_\rho) = P_n(Q_\rho \subset \Gamma) \geq \alpha$.

The computational procedure to obtain CE_α implicit in Definition 14 requires an optimization over $\rho \in [0, 1]$ where at each step we need to evaluate $P_n(Q_\rho \subset \Gamma)$. Since the family of Vorob'ev quantiles is a nested family of sets a dichotomy algorithm is guaranteed to converge in this setting. Figure 2.3 shows the conservative estimate at 95% for the example introduced in Figure 2.1.

In Bolin and Lindgren (2015) the computational costs are eased by the assumption that the underlying process Z is a Gaussian Markov random
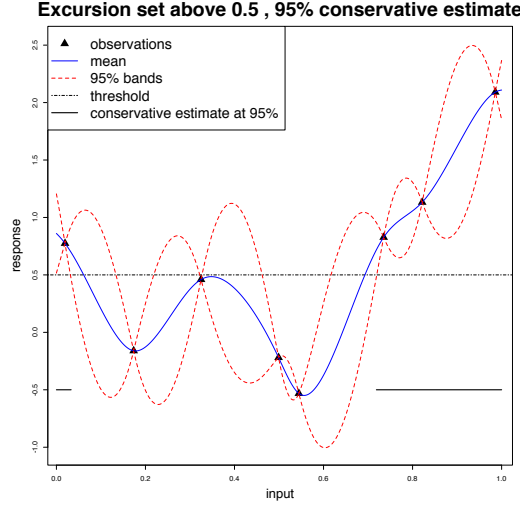
Figure 2.3: Conservative estimate at 95% for the excursion set above $t = 0.5$ with the Gaussian model introduced in Figure 2.1.

field. In our work the process Z does not necessarily satisfy this assumption, thus a new method is developed to reduce the conservative estimate's computational burden. The main computational bottleneck of the procedure is the evaluation of $P_n(Q_\rho \subset \Gamma)$ for each candidate set Q_ρ. Consider now the case where

$$\Gamma = \{x \in D : Z_x \leq t\},$$

the set of excursion below t. In this framework we have

$$P_n(Q_\rho \subset \Gamma) = P_n(Q_\rho \subset \{Z_x \leq t\}).$$

The Vorob'ev quantiles are usually discretized over fine designs, e.g. regular grids for small d, as described in Equation (2.24). If we consider the discretization $G = \{u_1, \ldots, u_r\} \subset D$, we have

$$P_n(Q_\rho \subset \{Z_x \leq t\}) = P_n(Z_{u_{i_1}} \leq t, \ldots, Z_{u_{i_l}} \leq t), \qquad (2.25)$$

where $\{u_{i_1}, \ldots, u_{i_l}\}$ is the subset of points in G belonging to K_ρ. Equation (2.25) shows the link between conservative estimates of excursion sets and orthant probabilities for Gaussian vectors. In Chapter 4 we introduce a new asymmetric nested Monte Carlo approach for the computation of such probabilities. This algorithm was originally developed to compute the exceedance probability of the maximum of a Gaussian vector. The estimator implements a technique that exploits the different computational costs at stake, thus maximizing the efficiency of the Monte Carlo estimator.

2.4.4 Sequential methods for set estimation

In the previous sections we have introduced the main methods used in this work for excursion set estimation. The techniques are based on a Bayesian model relying on n evaluations of f at the set of points $\mathbf{X}_n$. This set of points is also called design of experiments and plays an important role in the modelling. In fact if the design selects points that are non informative the resulting model might be very inaccurate. As an illustrative example, let us fix $|\epsilon| < 1$ and imagine that we are interested in $\Gamma^\star$, the excursion above $t = \frac{1}{2}\exp(-1/\epsilon)$ of a function $f : [-1, 1] \to \mathbb{R}$ defined as

$$f(x) := \begin{cases} 0 & \text{if } x \in [-1, -\epsilon) \cup (\epsilon, 1] \\ e^{-1/(\epsilon - |x|^2)} & \text{if } x \in [-\epsilon, \epsilon] \end{cases}.$$

If we choose a design $\mathbf{X}_n$ that selects only points in $[-1, -\epsilon) \cup (\epsilon, 1]$, then a Gaussian process posterior will not be able to estimate $\Gamma^\star$.

In the community of computer experiments many methods for choosing appropriate design of experiments have been proposed in the recent years, see, e.g., the reviews Chen et al. (2006); Franco (2008); Pronzato and Müller (2012) and references therein. Experimental designs can be roughly divided into two types: model-free and model-based designs. The former type consists of designs that are not related to the chosen model and are often fixed before any evaluation of the function. The most common designs of this type are space filling, where the geometric properties of the function domain are exploited to select points that cover the space in an optimal sense. Examples are minimax or maximin distance designs (Johnson et al., 1990), Latin hypercube samples (McKay et al., 1979) or low discrepancy sequences such as the Halton sequence and the Sobol' sequence (Bratley and Fox, 1988), see, e.g. Pronzato and Müller (2012) for a review.

Another type of designs are model-based, where the information available from the meta-model is exploited to select the next evaluations of the function. Model-based designs are often implemented sequentially: starting from a small initial design $\mathbf{X}_n$, chosen either ad hoc or with a space filling method, new evaluations of the objective function are added by choosing points that minimize a specific criterion. Two examples of model-based criteria were introduced in Sacks et al. (1989): the Integrated Mean Squared Error (IMSE) and the Maximum Mean Squared Error (MMSE) criteria. These two criteria add new evaluations at locations that minimize the prediction error of the model. Model-based designs, proved to be particularly useful for global optimization of a continuous functions $f : D \subset \mathbb{R}^d \to \mathbb{R}$ under a limited evaluation budget. A notable example is the Expected Improvement (EI)

algorithm, introduced in Mockus et al. (1978) and popularized in Jones et al. (1998) under the Efficient Global Optimization (EGO) terminology. This approach exploits meta-models to surrogate the objective function f from few evaluations and approximates the global optimum by sequentially updating the meta-model. Each update consists of a new evaluation or a batch of new evaluations of the function f at locations that maximize the "expected improvement" of the model. As an example consider the problem of finding the global minimum of a function f. Here we assume that it is a realization of a Gaussian process and, as in Section 2.2.3, we fix a prior $(Z_x)_{x \in D} \sim GP(\mathfrak{m}, \mathfrak{K})$. Let us denote the initial design with $\mathbf{X}_n$, the respective evaluations of f with the vector $\mathbf{f}_n = (f(x_1), \ldots, f(x_n))$ and their current minimum with $f_{\min}$. The meta-model can be updated with new evaluations of f at points maximizing the expected improvement criterion

$$\mathrm{EI}(x) := \mathbb{E}[\max(f_{\min} - Z_x, 0) \mid Z_{\mathbf{X}_n} = \mathbf{f}_n].$$

In the Gaussian case this criterion can be computed in closed form as detailed in Jones et al. (1998). Further, more recent works developed the expected improvement criterion in the batch sequential case and allowed a fast computation of the criterion, see, e.g., Schonlau (1997); Chevalier and Ginsbourger (2014); Marmin et al. (2016) and references therein.

   Sequential strategies have been also used in recent years for other purposes, such as recovering contour lines (Ranjan et al., 2008), target regions (Picheny et al., 2010) and excursion sets (Chevalier, 2013). These strategies can be cast as *Stepwise Uncertainty Reduction* (SUR) strategies, introduced in this framework by Bect et al. (2012). The idea behind SUR strategies is to select the next evaluation of the function f by minimizing an uncertainty function. The specific definition of such uncertainties depend on the problem at hand. For example, in Bect et al. (2012), the object of interest is the probability of failure $\alpha = \mu(\Gamma^\star)$, where μ is a probability measure defined on D and $\Gamma^\star = \{x \in D : f(x) \geq t\}$. An uncertainty function related to this problem is the variance of the estimator $\hat{\alpha} := \int_D \mathbf{1}_{Z_x \geq t} d\mu(x)$ for α. Chapter 5 reviews the SUR strategies in the context of excursion set estimation and introduces new strategies to reduce the uncertainty on conservative estimates.

Chapter 3

Quantifying uncertainties on excursion sets under a Gaussian random field prior

This chapter reproduces the paper Azzimonti et al. (2016a), co-authored with Julien Bect, Clément Chevalier and David Ginsbourger published in SIAM/ASA Journal on Uncertainty Quantification (DOI:10.1137/141000749).

3.1   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 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 is 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 Dubourg et al. (2011), Bect et al. (2012), nuclear criticality Chevalier et al. (2014a), etc.), but also in natural sciences, where conditions leading to dangerous phenomena in climatological (French and Sain, 2013) or geophysical (Bayarri et al., 2009) settings are of crucial interest.

In this paper we consider a setup where the forward model is a function

$f : D \subset \mathbb{R}^d \to \mathbb{R}$ and we are interested in the inverse problem of reconstructing the set $\Gamma^\star = f^{-1}(T) = \{\mathbf{x} \in D : f(\mathbf{x}) \in T\}$, where $T \subset \mathbb{R}$ denotes the range of values of interest for the output. Often the forward model f is costly to evaluate and a systematic exploration of the input space D, e.g., on a fine grid, is out of reach, even in small dimensions. Therefore reconstructions of $\Gamma^\star$ have to be performed based on a small number of evaluations, thereby 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 networks, and more. Here we focus on the Gaussian random field (GRF) modelling approach (also known as *Gaussian process modelling*, Rasmussen and Williams (2006)). 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., Jones et al., 1998; Villemonteix et al., 2009; Ranjan et al., 2008; Bect et al., 2012; Roustant et al., 2012; Binois et al., 2015). In this framework we assume that f is a realization of a random field $Z = (Z_\mathbf{x})_{\mathbf{x} \in D}$, which throughout the paper, unless otherwise noted, is assumed to be Gaussian with continuous paths almost surely. A major advantage of GRF models over deterministic approximation models is that, given a few observations of the function f at the points $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, they deliver a posterior probability distribution on functions, enabling not only predictions of the objective function at any point but also a quantification of the associated uncertainties.

The mean of the posterior field Z gives a plug-in estimate of the set $\Gamma^\star$ (see, e.g., Ranjan et al., 2008, and references therein), however here we focus on estimates based on 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., Lantuéjoul, 2002; Chilès and Delfiner, 2012; Bolin and Lindgren, 2015). Instead of having a single estimate of the excursion set like in most *set estimation* approaches (see, e.g., Cuevas and Fraiman, 2010; Hall and Molchanov, 2003; Reitzner et al., 2012), it is possible to get a distribution of sets. For example, Figure 3.1 shows some realizations of an excursion set obtained by simulating a GRF Z conditional on few observations of the function f at locations $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ ($n = 6$, black triangles). A natural question arising in practice is how to summarize this distribution by appealing to simple concepts, analogous 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 (Chevalier et al., 2013) in the context of excursion set estimation and uncertainty quantification with GRF models. In Sections 3.2 and 3.5 we review another random set expectation,

the distance average expectation (see, e.g., Baddeley and Molchanov, 1998). This expectation provides a different uncertainty quantification estimate in the context of GRF modeling, the distance average variability. Since the distance average variability heavily relies on conditional simulations, to the best of our knowledge, it has not been used before as an uncertainty quantification technique.

One of the key contributions of the present paper is a method to approximate conditional realizations of the random excursion set based on simulations of the underlying GRF at few points. By contrast, in the literature, Monte Carlo simulations of excursion sets are often obtained by simulating the underlying field at space filling designs, as shown in Figure 3.1b. 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 $\mathbf{E}_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ where the field is simulated. The field's values are then approximated on the full design with a suitable affine predictor. We call a *quasi-realization* of the excursion set the excursion region of a simulation of the approximate field. Coming back to the example introduced in Figures 3.1 and 3.1c shows quasi-realizations of the excursion set Γ based on simulations of the field at $m = 30$ points predicted at the fine design with the best linear unbiased predictor. 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 of how close those approximations are to the true random set distribution.

The paper is divided into six sections. In Section 3.2 we introduce the framework and the fundamental definitions needed for our method. In Section 3.3 we give an explicit formula for 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 proofs are in Appendix 3.8. Section 3.4 explains the computational aspects and introduces two algorithms to calculate the optimized points. In this section we also discuss the advantages and limitations of these algorithms. Section 3.5 presents the implementation of the distance average variability as an uncertainty quantification measure. We show that this uncertainty measure can be computed accurately with the use of quasi-realizations. In Section 3.6 we show how the simulation method allows us to compute estimates of the level sets in a two-dimensional test case from nuclear safety engineering. The proposed method to generate accurate quasi-realizations of the excursion set from few simulations of
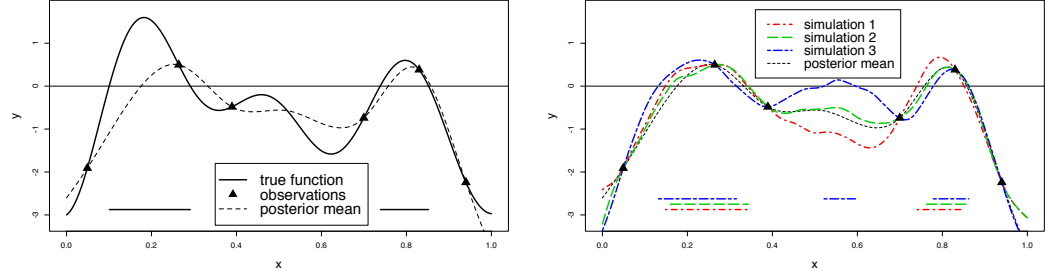
(a) True function, posterior GRF mean and true excursion set $\Gamma^\star = \{x \in [0,1] : f(x) \geq t\}$ with $t = 0$ (horizontal lines at $y = -3$).

(b) 3 realizations of the conditional GRF and the associated excursion set (horizontal lines at $y = -3$), obtained with simulations at 1000 points in $[0,1]$.

(c) 3 quasi-realizations of the conditional GRF and the associated random set (horizontal lines at $y = -3$) generated by simulating at 30 optimally-chosen points (black diamonds, shown at $y = 1.7$) and predicting the field at the 1000 points design.
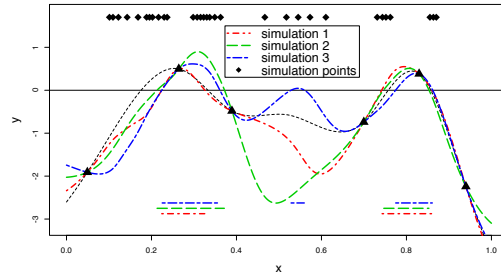
Covariance function	Matérn ($\nu = 5/2$)
Number of observations	$n = 6$ (▲)
Simulation points optimized	Algorithm B
Number of simulation points	$m = 30$ (♦)
Threshold	$t = 0$

Figure 3.1: GRF model based on few evaluations of a deterministic function.

the underlying field is pivotal in this test case as it allows us to operate on high-resolution grids thus obtaining good linear approximations of the level set curve. Another six-dimensional application is presented in Appendix 3.9, where the distribution of the excursion volume is estimated with approximate conditional simulations generated using the proposed simulation method.

3.2 Preliminaries

In this section we recall two concepts coming from the theory of random closed sets. The first one gives us the distance between the reconstructed set and the true random set, while the second one leads to the definition of an uncertainty quantification measure for the excursion set estimate. See, e.g., Molchanov (2005, Chapter 2), for a detailed overview on the subject.

Throughout the paper $f : D \subset \mathbb{R}^d \longrightarrow \mathbb{R}$, $d \geq 1$, is considered an unknown real-valued continuous objective function and D is a compact subset of $\mathbb{R}^d$. We model f with $Z = (Z_{\mathbf{x}})_{\mathbf{x} \in D}$, a GRF with continuous paths, whose mean function and covariance kernel are denoted by $\mathfrak{m}$ and $\mathfrak{K}$. The range of critical responses 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 $\Gamma^{\star} = f^{-1}(T) = \{\mathbf{x} \in D : f(\mathbf{x}) \in T\}$, respectively. In most applications, T is a closed set of the form $[t, \infty)$ for some $t \in \mathbb{R}$. Here we solely need to assume that T is closed in $\mathbb{R}$; however, we restrict ourselves to $T = [t, \infty)$ for simplicity. Generalizations to unions of intervals are straightforward. The excursion set $\Gamma^{\star}$ is closed in D because it is the preimage of a closed set by a continuous function. Similarly, $\Gamma = \{\mathbf{x} \in D : Z(\mathbf{x}) \in T\}$ defines a random closed set.

3.2.1 Vorob'ev approach

A key element for the proposed simulation method is the notion of *distance in measure*. Let μ be a 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 closed sets Γ_1 and Γ_2, one can define a distance as follows.

Definition 15 (Expected distance in measure)**.** *The expected distance in measure between two random closed sets Γ_1, Γ_2 with respect to a Borel measure μ is the function defined by*

$$d_{\mu}(\Gamma_1, \Gamma_2) = \mathbb{E}[\mu(\Gamma_1 \Delta \Gamma_2)].$$
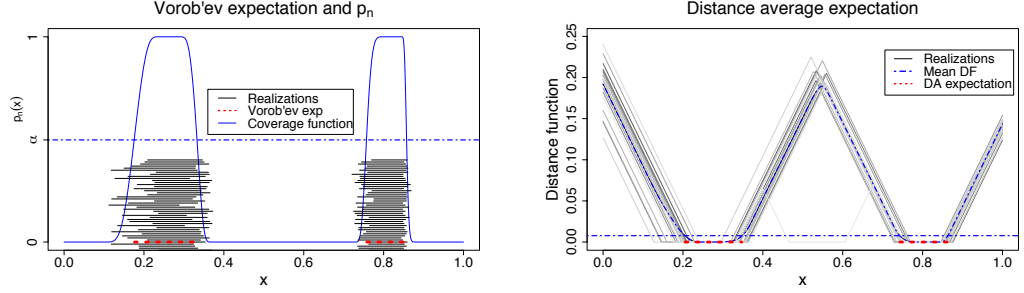
(a) Excursion set realizations, coverage function (blue), selected α-level (0.498, dashed blue), Vorob'ev expectation (red dashed line at $y = 0$, length= 0.257).

(b) Distance function realizations, average distance function (blue) and distance average expectation (red dashed line at $y = 0$, length= 0.264) obtained with the Euclidean distance function.

Figure 3.2: Realizations Γ obtained from the GRF presented in Figure 3.1 and two random set expectations for this excursion set.

Several notions of expectation have been proposed for random closed sets, in particular, the Vorob'ev expectation is related to the expected distance in measure. Consider the coverage function of a random closed set Γ, $p_\Gamma :$ $D \longrightarrow [0, 1]$ defined as $p_\Gamma(\mathbf{x}) := P(\mathbf{x} \in \Gamma)$. The Vorob'ev expectation Q_α of Γ is defined as the α level set of its coverage function, i.e. $Q_\alpha = \{\mathbf{x} \in D :$ $p_\Gamma(\mathbf{x}) \geq \alpha\}$ (Vorob'ev, 1984), where the level α satisfies $\mu(Q_\beta) < \mathbb{E}[\mu(\Gamma)] \leq$ $\mu(Q_\alpha)$ for all $\beta > \alpha$. It is a well-known fact (Molchanov, 2005) 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 $\mu(M) = \mathbb{E}[\mu(\Gamma)]$. Figure 3.2a shows the Vorob'ev expectation computed for the excursion set of the GRF in the example of Figure 3.1. While the Vorob'ev expectation is used for its conceptual simplicity and its tractability, there exist other definitions of random closed set expectation and variability. In the following we review another notion of expectation for a random closed set: the distance average and its related notion of variability.

3.2.2   Distance average approach

The *distance function* of a point $\mathbf{x}$ to a set S is defined as the function $d : D \times \mathcal{F}' \longrightarrow \mathbb{R}$ that returns the distance between $\mathbf{x} \in D$ and $S \in \mathcal{F}'$, where $\mathcal{F}'$ is the space of all nonempty closed sets in D (see Molchanov, 2005, pp. 179–180 for details). In general, such distance functions can take

any value in $\mathbb{R}$ (see Baddeley and Molchanov, 1998; Molchanov, 2005, for examples), however here we restrict ourselves to non-negative distances. In what follows, we use the distance function $d(\mathbf{x}, S) = \inf\{\rho(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in D, \mathbf{y} \in S\}$, where ρ is the Euclidean distance in $\mathbb{R}^d$.

Consider $S = \Gamma$ and assume that $d(\mathbf{x}, \Gamma)$ has finite expectation for all $\mathbf{x} \in D$, the *mean distance function* is $\bar{d} : D \to \mathbb{R}^+$, defined as $\bar{d}(\mathbf{x}) := \mathbb{E}[d(\mathbf{x}, \Gamma)]$. Recall that, after a restriction to D, it is possible to embed the space of Euclidean distance functions in $L^2(D)$. Let us further denote with $\mathfrak{d}(f, g)$ the L^2 metric, defined as $\mathfrak{d}(f, g) := \left(\int_D (f - g)^2 d\mu\right)^{1/2}$. The *distance average* of Γ (Molchanov, 2005) is defined as the set that has the closest distance function to $\bar{d}$ with respect to the metric $\mathfrak{d}$.

Definition 16 (Distance average and distance average variability). *Let $\bar{u}$ be the value of $u \in \mathbb{R}$ that minimizes the $\mathfrak{d}$-distance $\mathfrak{d}(d(\cdot, \{\bar{d} \leq u\}), \bar{d})$ between the distance function of $\{\bar{d} \leq u\}$ and the mean distance function of Γ. If $\mathfrak{d}(d(\cdot, \{\bar{d} \leq u\}), \bar{d})$ achieves its minimum in several points we assume $\bar{u}$ to be their infimum. The set*

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

is called the distance average of Γ with respect to $\mathfrak{d}$. In addition, we define the distance average variability of Γ with respect to $\mathfrak{d}$ as $\mathrm{DAV}(\Gamma) = \mathbb{E}[\mathfrak{d}^2(\bar{d}, d(\cdot, \Gamma))]$.

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 $\mathrm{DAV}(\Gamma)$ relying on approximate GRF simulations. In general, distance average and distance average variability can be estimated only with Monte Carlo techniques, therefore we need to be able to generate realizations of Γ. By taking a standard matrix decomposition approach for GRF simulations, a straightforward way to obtain realizations of 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 with its discrete approximation on the design G, $\Gamma_G = \{\mathbf{u} \in G : Z_{\mathbf{u}} \in T\}$. 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. Figure 3.2b shows the distance average computed with Monte Carlo simulations for the excursion set of the example in Figure 3.1. In the example the distance average expectation has a slightly bigger Lebesgue measure than the Vorob'ev expectation. In general the two random set expectations yield different estimates, sometimes even resulting in a different number of connected components, as in the example introduced in Section 3.5.

3.3   Main results

In this section we assume that Z has been evaluated at locations $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset D$; thus we consider the GRF conditional on the values $Z(\mathbf{X}_n) := (Z_{\mathbf{x}_1}, \ldots, Z_{\mathbf{x}_n})$. Following the notation for the moments of Z introduced in Section 3.2, we denote the mean and covariance kernel of Z conditional on $Z(\mathbf{X}_n) := (Z_{\mathbf{x}_1}, \ldots, Z_{\mathbf{x}_n})$ with $\mathfrak{m}_n$ and $\mathfrak{K}_n$, respectively. The proposed approach consists in replacing conditional GRF simulations at the finer design G with approximate simulations that rely on a smaller simulation design $\mathbf{E}_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$, with $m \ll r$. The quasi-realizations generated with this method can be used as basis for quantifying uncertainties on Γ, for example, with the distance average variability. Even though such an 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.

3.3.1   A Monte Carlo approach with predicted conditional simulations

We propose to replace Z by a simpler random field denoted by $\widetilde{Z}$, whose 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.

Consider a set $\mathbf{E}_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ of m points in D, $1 \leq m \leq r$, and denote by $Z(\mathbf{E}_m) = (Z_{\mathbf{e}_1}, \ldots, Z_{\mathbf{e}_m})^T$ the random vector of values of Z at $\mathbf{E}_m$. Conditional on $Z(\mathbf{X}_n)$, this vector is multivariate Gaussian with mean $\mathfrak{m}_n(\mathbf{E}_m) = (\mathfrak{m}_n(\mathbf{e}_1), \ldots, \mathfrak{m}_n(\mathbf{e}_m))^T$ and covariance matrix $\mathfrak{K}_n(\mathbf{E}_m, \mathbf{E}_m) = [\mathfrak{K}_n(\mathbf{e}_i, \mathbf{e}_j)]_{i,j=1,\ldots,m}$. The essence of the proposed approach is to appeal to affine predictors of Z, i.e. to consider $\widetilde{Z}$ of the form

$$\widetilde{Z}(\mathbf{x}) = a(\mathbf{x}) + \mathbf{b}^T(\mathbf{x})Z(\mathbf{E}_m) \qquad (\mathbf{x} \in D), \tag{3.1}$$

where $a : D \longrightarrow \mathbb{R}$ is a continuous trend function and $\mathbf{b} : D \longrightarrow \mathbb{R}^m$ is a continuous vector-valued function of deterministic weights. Note that usual kriging predictors are particular cases of Equation (3.1) with adequate choices of the functions a and $\mathbf{b}$; see, for example, Cressie (1993) for an extensive review. Re-interpolating conditional simulations by kriging is an idea that has already been proposed in different contexts, notably by Oakley (1999) in the context of Bayesian uncertainty analysis for complex computer codes. However, while the problem of selecting the evaluation points $\mathbf{X}_n$ has been addressed in many works (see,e.g., Sacks et al., 1989; Jones et al., 1998;

Gramacy and Lee, 2009; Ranjan et al., 2008; Chevalier et al., 2014a, and references therein), to the best of our knowledge the derivation of optimal criteria for choosing the simulation points $\mathbf{E}_m$ has not been addressed until now, be it for excursion set estimation or for further purposes. Computational savings for simulation procedures are hinted at by the computational complexity of simulating the two fields. Simulating Z at a design with r points with standard matrix decomposition approaches has a computational complexity $O(r^3)$, while simulating $\widetilde{Z}$ has a complexity $O(rm^2 + m^3)$. Thus if $m \ll r$ simulating $\widetilde{Z}$ might bring substantial savings.

In Figure 3.3 we present an example of work flow that outputs a quantification of uncertainty over the estimate Γ for $\Gamma^\star$ based on the proposed approach. In the following sections we provide an equivalent formulation of the expected distance in measure between Γ and $\widetilde{\Gamma}$ introduced in Definition 15 and we provide methods to select optimal simulation points $\mathbf{E}_m$.

3.3.2 Expected distance in measure between $\widetilde{\Gamma}$ and Γ

In the next proposition we show an alternative formulation of the expected distance in measure between $\widetilde{\Gamma}$ and Γ that exploits the assumptions on the field Z.

Proposition 2 (Distance in measure between Γ and $\widetilde{\Gamma}$)**.** *Under the previously introduced assumptions* $(Z, \widetilde{Z})$ *is a bivariate GRF and* Γ *and* $\widetilde{\Gamma}$ *are random closed sets.*
a) Assume that $D \subset \mathbb{R}^d$ *and* μ *is a finite Borel measure on* D, *then we have*

$$d_{\mu,n}(\Gamma, \widetilde{\Gamma}) = \int \rho_{n,m}(\mathbf{x}) \mu(\mathrm{d}\mathbf{x}) \tag{3.2}$$

with

$$\rho_{n,m}(\mathbf{x}) = P_n\big(\mathbf{x} \in \Gamma \Delta \widetilde{\Gamma}\big)$$
$$= P_n(Z(\mathbf{x}) \geq t, \widetilde{Z}(\mathbf{x}) < t) + P_n(Z(\mathbf{x}) < t, \widetilde{Z}(\mathbf{x}) \geq t).$$

where P_n *denotes the conditional probability* $P\big(\cdot \mid Z(\mathbf{X}_n)\big)$.
b) Moreover, using the notation introduced in Section 3.3, we get

$$P_n(Z(\mathbf{x}) \geq t, \widetilde{Z}(\mathbf{x}) < t) = \Phi_2\left(\mathbf{c}_n(\mathbf{x}, \mathbf{E}_m), \Sigma_n(\mathbf{x}, \mathbf{E}_m)\right), \tag{3.3}$$

where $\Phi_2(\,\cdot\,, \Sigma)$ *is the c.d.f. of a centred bivariate Gaussian with covariance* Σ, *where*

$$\mathbf{c}_n(\mathbf{x}, \mathbf{E}_m) = \begin{pmatrix} \mathfrak{m}_n(\mathbf{x}) - t \\ t - a(\mathbf{x}) - \mathbf{b}(\mathbf{x})^T \mathfrak{m}_n(\mathbf{E}_m) \end{pmatrix}$$

Input:
- **Prior** $Z \sim GRF(\mathfrak{m}, \mathfrak{K})$;
- **Data** $\mathbf{X}_n, f(\mathbf{X}_n) = Z(\mathbf{X}_n)$;
- **Fine simulation design** G.

Posterior GRF $Z \mid Z(\mathbf{X}_n)$
with mean $\mathfrak{m}_n$ and covariance kernel $\mathfrak{K}_n$.

Approximation step

Obtain simulation points $\mathbf{E}_m$ with

- **Algorithm A:** see Section 3.4.1; or
- **Algorithm B:** see Section 3.4.2.

Simulation step

Simulate $\widetilde{Z} \mid Z(\mathbf{X}_n)$ at G, where
$\widetilde{Z}(\mathbf{x}) = a(\mathbf{x}) + \mathbf{b}^T(\mathbf{x}) Z(\mathbf{E}_m), \ \mathbf{x} \in G$
(see Section 3.3)

Obtain quasi-realizations of $\Gamma \mid Z(\mathbf{X}_n)$
$\widetilde{\Gamma} = \{\mathbf{x} \in G : \widetilde{Z}(\mathbf{x}) \in T\}$.

Output:

- Quasi-realizations of $\Gamma \mid Z(\mathbf{X}_n)$;
- Uncertainty quantification on $\Gamma \mid Z(\mathbf{X}_n)$
  (e.g. DAV(Γ), $\tilde{l}(\partial\Gamma), \mu(\Gamma)$,
  Sections 3.5, 3.6, 3.9).

Figure 3.3: Flow chart of proposed operations to quantify the posterior uncertainty on Γ.

and

$$\Sigma_n(\mathbf{x}, \mathbf{E}_m) = \begin{pmatrix} \mathfrak{K}_n(\mathbf{x}, \mathbf{x}) & -\mathbf{b}(\mathbf{x})^T \mathfrak{K}_n(\mathbf{E}_m, \mathbf{x}) \\ -\mathbf{b}(\mathbf{x})^T \mathfrak{K}_n(\mathbf{E}_m, \mathbf{x}) & \mathbf{b}(\mathbf{x})^T \mathfrak{K}_n(\mathbf{E}_m, \mathbf{E}_m) \mathbf{b}(\mathbf{x}) \end{pmatrix}.$$

c) Particular case: if $\mathbf{b}(\mathbf{x})$ *is chosen as the simple kriging weights* $\mathbf{b}(\mathbf{x}) =$

$\mathfrak{K}_n(\mathbf{E}_m, \mathbf{E}_m)^{-1}\mathfrak{K}_n(\mathbf{E}_m, \mathbf{x})$, *then*

$$\Sigma_n(\mathbf{x}, \mathbf{E}_m) = \begin{pmatrix} \mathfrak{K}_n(\mathbf{x}, \mathbf{x}) & -\gamma_n(\mathbf{x}, \mathbf{E}_m) \\ -\gamma_n(\mathbf{x}, \mathbf{E}_m) & \gamma_n(\mathbf{x}, \mathbf{E}_m) \end{pmatrix}, \tag{3.4}$$

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

Proof. (of Proposition 2)
a) Interchanging integral and expectation by Tonelli's theorem, we get

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

b) Since the random field Z is assumed to be Gaussian, the vector-valued random field $(Z(\mathbf{x}), \widetilde{Z}(\mathbf{x}))$ is also Gaussian conditionally on $Z(\mathbf{X}_n)$, and proving the property boils down to calculating its conditional moments. Now we directly get $\mathbb{E}_n[Z(\mathbf{x})] = \mathfrak{m}_n(\mathbf{x})$ and $\mathbb{E}_n[\widetilde{Z}(\mathbf{x})] = a(\mathbf{x}) + \mathbf{b}(\mathbf{x})^T\mathfrak{m}_n(\mathbf{E}_m)$. Similarly, $\mathrm{Var}_n[Z(\mathbf{x})] = \mathfrak{K}_n(\mathbf{x}, \mathbf{x})$ and $\mathrm{Var}_n[\widetilde{Z}(\mathbf{x})] = \mathbf{b}(\mathbf{x})^T\mathfrak{K}_n(\mathbf{E}_m, \mathbf{E}_m)\mathbf{b}(\mathbf{x})$. Finally, $\mathrm{Cov}_n[Z(\mathbf{x}), \widetilde{Z}(\mathbf{x})] = \mathbf{b}(\mathbf{x})^T\mathfrak{K}_n(\mathbf{E}_m, \mathbf{x})$ and Equation (3.3) follows by Gaussianity.
c) The expression in Equation (3.4) follows immediately by substituting $\mathbf{b}(\mathbf{x})$ into $\Sigma_n(\mathbf{x}, \mathbf{E}_m)$. $\square$

Remark 2. *The Gaussian assumption on the random field Z in Proposition 2 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 $Z(\mathbf{X}_n)$.*

3.3.3 Convergence result

Let $\mathbf{e}_1, \mathbf{e}_2, \dots$ be a given sequence of simulation points in D and set $\mathbf{E}_m = \{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ for all m. Assume that Z is, conditionally on $Z(\mathbf{X}_n)$, a GRF with conditional mean $\mathfrak{m}_n$ and conditional covariance kernel $\mathfrak{K}_n$. Let $\widetilde{Z}(\mathbf{x}) = \mathbb{E}_n\left(Z(\mathbf{x}) \mid Z(\mathbf{E}_m)\right)$ be the best predictor of $Z(\mathbf{x})$ given $Z(\mathbf{X}_n)$ and $Z(\mathbf{E}_m)$. In

particular, $\widetilde{Z}$ is affine in $Z(\mathbf{E}_m)$. Denote by $s^2_{n,m}(\mathbf{x})$ the conditional variance of the prediction error at $\mathbf{x}$:

$$s^2_{n,m}(\mathbf{x}) = \mathrm{Var}_n\left(Z(x) - \widetilde{Z}(x)\right) = \mathrm{Var}_n\left(Z(x) \mid Z(\mathbf{E}_m)\right)$$
$$= \mathfrak{K}_n(\mathbf{x}, \mathbf{x}) - \mathfrak{K}_n(\mathbf{E}_m, \mathbf{x})^T \, \mathfrak{K}_n(\mathbf{E}_m, \mathbf{E}_m)^{-1} \, \mathfrak{K}_n(\mathbf{E}_m, \mathbf{x}).$$

Proposition 3 (Approximation consistency). *Let $\widetilde{\Gamma}(\mathbf{E}_m) = \left\{\mathbf{x} \in D : \widetilde{Z}(\mathbf{x}) \in T\right\}$ be the random excursion set associated to $\widetilde{Z}$. Then, as $m \to \infty$, $d_{\mu,n}(\Gamma, \widetilde{\Gamma}(\mathbf{E}_m)) \to 0$ if and only if $s^2_{n,m} \to 0$ μ-almost everywhere.*

Corollary 1. *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 $d_{\mu,n}(\Gamma, \widetilde{\Gamma}(\mathbf{E}_m)) \to 0$ when $m \to \infty$). b) Assuming further that the covariance function of Z has the NEB property (Vazquez and Bect, 2010), the density condition is also necessary.*

The proof of Proposition 3 is in Appendix 3.8

3.4   Practicalities

In this section we use the results established in Section 3.3 to implement a method that selects appropriate simulation points $\mathbf{E}_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\} \subset D$, for a fixed $m \geq 1$. The conditional field is simulated on $\mathbf{E}_m$ and approximated at the required design with ordinary kriging predictors. We present two algorithms to find a set $\mathbf{E}_m$ that approximately minimizes the expected distance in measure between Γ and $\widetilde{\Gamma}(\mathbf{E}_m)$. The algorithms were implemented in **R** using the packages `KrigInv` (Chevalier et al., 2014c) and `DiceKriging` (Roustant et al., 2012).

3.4.1   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 (3.2) and we assume that the underlying field Z is Gaussian. Under these assumptions, an optimal set of simulation points is a minimizer

of the problem,

$$
\underset{\mathbf{E}_m}{\text{minimize}} \; d_{\mu,n}(\Gamma, \widetilde{\Gamma}) = \int \rho_{n,m}(\mathbf{x})\mu(\mathrm{d}\mathbf{x})
$$

$$
= \int \Phi_2 \left(\mathbf{c}_n(\mathbf{x}, \mathbf{E}_m), \Sigma_n(\mathbf{x}, \mathbf{E}_m) \right) \mu(\mathrm{d}\mathbf{x})
$$

$$
+ \int \Phi_2 \left(-\mathbf{c}_n(\mathbf{x}, \mathbf{E}_m), \Sigma_n(\mathbf{x}, \mathbf{E}_m) \right) \mu(\mathrm{d}\mathbf{x}). \quad (3.5)
$$

Several classic optimization techniques have already been employed to solve similar problems for optimal designs, for example simulated annealing (Sacks and Schiller, 1988), genetic algorithms (Hamada et al., 2001), or treed optimization (Gramacy and Lee, 2009). In our case such global approaches lead to an $m \times d$ 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 Sacks et al. (1989); Chevalier et al. (2014a) and optimize the criterion sequentially: given $E_{i-1}^* = \{\mathbf{e}_1^*, \ldots, \mathbf{e}_{i-1}^*\}$ 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 fixed as parameters and are not modified by the current optimization.

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 numerically cumbersome. Instead we exploit the sequential nature of the algorithm and use kriging update formulas (Chevalier et al., 2014b) to compute the new value of the criterion each time a new point is analysed.

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 (3.5). 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 c.d.f. introduced in Genz (1992). The integral is approximated via a quasi-Monte Carlo method: the integrand is evaluated using points from a space filling sequence (Sobol', Bratley and Fox (1988)) and then approximated with a sample mean of the values.

The criterion is optimized with the function `genoud` (Mebane and Sekhon, 2011), 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. Here, the gradients are numerically approximated.

3.4.2   Algorithm B: maximizing $\rho_{\mathrm{n,m}}(\mathbf{x})$

The evaluation of the criterion in Equation (3.5) can become computationally expensive because it requires a high number of evaluations of the bivariate normal c.d.f. in order to properly estimate the integral. This consideration led us to develop an alternative procedure.

We follow closely the reasoning used in Sacks et al. (1989) and Bect et al. (2012) for the development of a heuristic method to obtain the minimizer of the integrated mean squared error by maximizing the mean squared error. The characterization of the expected distance in measure in Equation (3.2) 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 $E_{i-1}^* = \{\mathbf{e}_1^*, \ldots, \mathbf{e}_{i-1}^*\}$ points previously optimized, the ith point $\mathbf{e}_i$ is the maximizer of the following problem:

$$\underset{\mathbf{x}}{\text{maximize }} \rho_{n,i-1}^*(\mathbf{x}) \text{ with}$$

$$\rho_{n,i-1}^*(\mathbf{x}) = \Phi_2\left(\mathbf{c}_n(\mathbf{x}, E_{i-1}^*), \Sigma_n(\mathbf{x}, E_{i-1}^*)\right) + \Phi_2\left(-\mathbf{c}_n(\mathbf{x}, E_{i-1}^*), \Sigma_n(\mathbf{x}, E_{i-1}^*)\right),$$
for fixed, previously optimized $E_{i-1}^* = \{\mathbf{e}_1^*, \ldots, \mathbf{e}_{i-1}^*\}$.

The evaluation of the objective function in Algorithm B does not require numerical integration in $\mathbb{R}^d$ and thus it requires substantially less evaluations of the bivariate normal c.d.f..

The maximization of the objective function is performed with the L-BFGS-B algorithm (Byrd et al., 1995) implemented in **R** with the function `optim`. The choice of starting points for the optimization is crucial for gradient descent algorithms. In our case the objective function to maximize is strongly related with p_Γ, the coverage function of Γ, in fact all points $\mathbf{x}_s$ where the function $w(\mathbf{x}) := p_\Gamma(\mathbf{x})(1 - p_\Gamma(\mathbf{x}))$ takes high values are reasonable starting points because they are located in regions of high uncertainty for the excursion set, thus simulations around their locations are meaningful. Before starting the maximization, the function $w(\mathbf{x})$ 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.

3.4.3   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 two-dimensional analytical example.

Consider the Branin-Hoo function (see Jones et al., 1998) multiplied by a factor -1 and normalized so that its 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 GRF Z with constant mean function $\mathfrak{m}$ and covariance $\mathfrak{K}$ chosen as a tensor product Matérn kernel ($\nu = 3/2$) (Stein, 1999). The covariance kernel parameters are estimated by maximum likelihood with the package `DiceKriging` (Roustant et al., 2012). By following the GRF modelling 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 (Stein, 1987) and the conditional mean and covariance are computed with ordinary kriging equations.

Discrete quasi-realizations 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 LHS (Stein, 1987) and points from the Sobol' sequence (Bratley and Fox, 1988).

Figure 3.4 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 a much larger number of points is required to fill the space, and hence optimizing the points becomes more advantageous, as shown in Appendix 3.9. Algorithms A and B show 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 later in Figure 3.6.
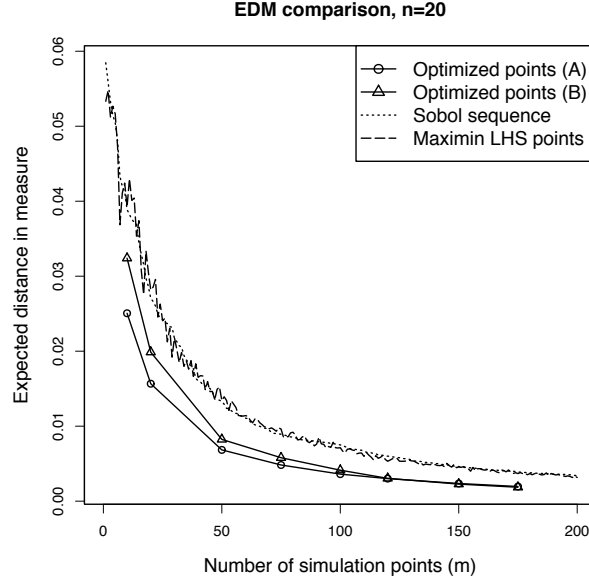
Figure 3.4: Expected distance in measure for different choices of simulation points

3.5 Application: a new variability measure using the distance transform

In this section we deal with the notions of distance average and distance average variability introduced in Section 3.2 and more specifically we present an application where the interpolated simulations are used to efficiently compute the distance average variability.

Let us recall that given $\Gamma_1, \ldots, \Gamma_N$ realizations of the random closed set Γ, we can compute the estimator for $\mathbb{E}_{\mathrm{DA}}(\Gamma)$

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

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 16; see (Baddeley and Molchanov, 1998) for more details. The variability of this estimate is measured with the distance average variability DAV(Γ), which, in the empirical case, is defined as

$$\mathrm{DAV}(\Gamma) = \frac{1}{N} \sum_{i=1}^{N} \mathfrak{d}^2(\bar{d}^*, d(\cdot, \Gamma_i)) = \frac{1}{N} \sum_{i=1}^{N} \int_{\mathbb{R}^d} \left(d(\mathbf{x}, \Gamma_i) - \bar{d}^*(\mathbf{x}) \right)^2 d\mu(\mathbf{x}),$$
$$(3.6)$$

where $\mathfrak{d}(\cdot, \cdot)$ is the $L^2(\mathbb{R}^d)$ distance.

The distance average variability is a measure of uncertainty for the excursion set under the postulated GRF model; this value is high when the distance functions associated with the realizations Γ_i are highly varying, which implies that the distance average estimate of the excursion set is uncertain. This uncertainty quantification method necessitates conditional simulations of the field on a fine grid to obtain a point-wise estimate. Our simulation method generates quasi-realizations in a rather inexpensive fashion even on high-resolution grids, thus making the computation of this uncertainty measure possible.

We consider here the two-dimensional example presented in Section 3.4 and we show that by selecting few well-chosen simulation points $\mathbf{E}_m = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$, with $m \ll r$, and interpolating the results on G, it is possible to achieve very similar estimate 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, the discrete approximation of the distance average, with an adaptation for $\mathbf{R}$ of the fast distance transform algorithm implemented in Felzenszwalb and Huttenlocher (2004). The precision of the estimate $\mathbb{E}_{DA}^*(\Gamma)$ is evaluated with the distance transform variability, denoted here with $\mathrm{DTV}(\Gamma; r)$, an approximation on the grid of the distance average variability, Equation (3.6).

The value of the distance transform variability is estimated with quasi-realizations of Γ obtained from simulations at few points. The conditional GRF is first simulated $10,000$ times at a design $\mathbf{E}_m$ 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 with the affine predictor $\widetilde{Z}$. Three methods to obtain simulation points are compared: Algorithms A and B presented in the previous section and a maximin LHS design. The simulations obtained with points from each of 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 then used to obtain the value of the interpolated field $\widetilde{Z}(\mathbf{u})$ from the simulated values $Z(\mathbf{E}_m)$. This procedure is numerically fast as it only requires algebraic operations.

For comparison a benchmark estimate of $\mathrm{DTV}(\Gamma; r)$ is obtained from realizations of Γ stemming from $10,000$ conditional Gaussian simulations on the same grid of size $r = 50 \times 50$.

Both 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. Figure 3.5 shows a comparison of the distributions of $\mathrm{DTV}(\Gamma; r)$ obtained with full grid simulations and the distributions obtained with the interpolation
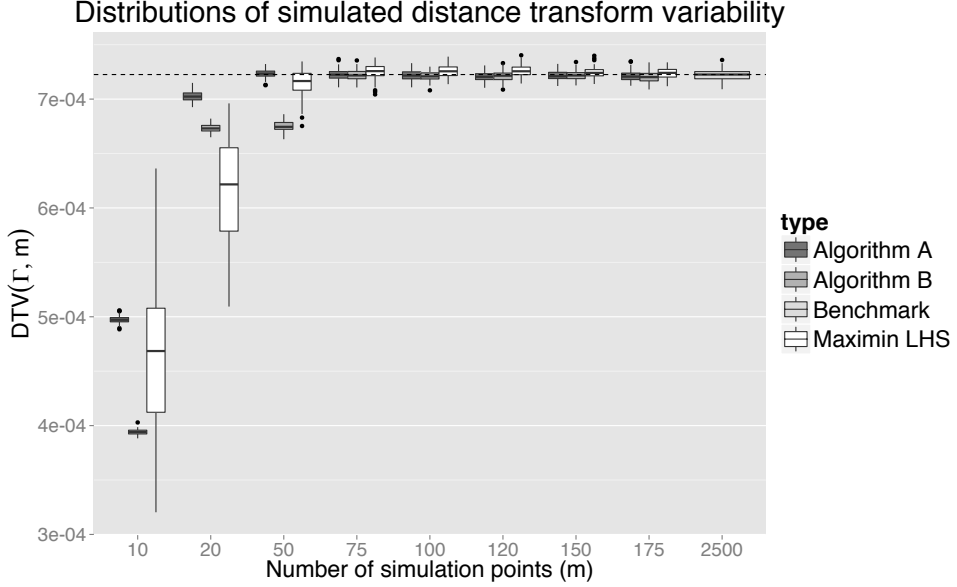
Figure 3.5: Comparison of the distributions of the simulated $\mathrm{DTV}(\Gamma; r)$ for different methods (from left to right, Algorithm A, Algorithm B and maximin LHS); the dashed horizontal line marks the median value of the benchmark ($m = 2500$) distribution.

over the grid of few simulations.

The distributions of $\mathrm{DTV}(\Gamma; r)$ obtained from quasi-realizations all approximate well the benchmark distribution with as little as 100 simulation points, independently of the way simulation points are selected. This effect might be enhanced by the low dimension of the example; nonetheless it suggests substantial savings in simulation costs.

The optimized designs (Algorithms A and B) achieve better approximations with fewer 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 B shows inferior precision compared with Algorithm A for very small values of m. This behaviour could be influenced by the dependency of the first simulation point on the starting point of the optimization procedure. In general, the choice between Algorithm A and Algorithm B is a trade-off between computational speed and precision. For low dimensional problems, or more in general, if only a small number of simulation points is needed, then
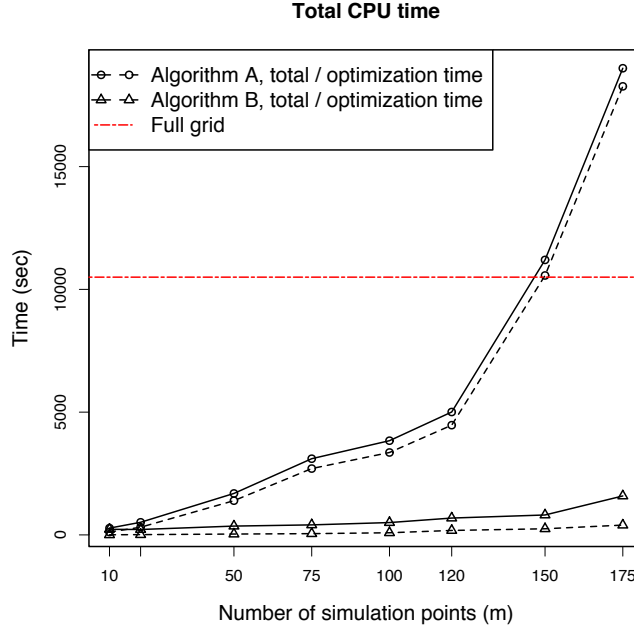
Figure 3.6: Total CPU time to obtain all realizations of Γ. Full grid simulations include only simulation time (dot-dashed horizontal line), while both algorithms include simulation point optimization (dashed lines) and simulation and interpolation times (solid lines).

Algorithm A could be employed at acceptable computational costs. However as the dimensionality increases more points are needed to approximate correctly full designs simulations, then Algorithm B obtains similar results to Algorithm A at a much lower computational cost. Both algorithms 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 at by the result on distance in measure shown in the previous section. In most practical situations Algorithm B yields the better trade-off between computational speed and precision, provided that enough simulation points are chosen.

Figure 3.6 shows the total CPU time for all the simulations in the experiment for Algorithm A, Algorithm B and the full grid simulations, computed on the cluster of the University of Bern with Intel Xeon E5649 2.53-GHz CPUs with 4 GB RAM. The CPU times for Algorithms 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.

3.6   Test case: Estimating length of critical level set in nuclear safety application

In this section we focus on a nuclear safety test case and we show that our method to generate quasi-realizations can be used to obtain estimates of the level set on high-resolution grids.

The problem at hand is a nuclear criticality safety assessment. In a system involving nuclear material it is important to control the chain reaction that may be produced by neutrons, which are both the initiators and the product of the reaction. An overproduction of neutrons in the radioactive material is not safe for storage or transportation. Thus, the criticality safety of a system is often evaluated with the neutron multiplication factor (k-effective, or k-eff) which returns the number of neutrons produced by a collision with one neutron. This number is usually estimated using a costly simulator. If k-eff > 1 the chain reaction is unstable; otherwise it is safe. In our case we consider a storage facility of plutonium powder, whose k-eff is modelled by two parameters: the mass of plutonium (MassPu) and the logarithm of the concentration of plutonium (logConcPu). The excursion set of interest is the set of safe input parameters $\Gamma^* = \{(\text{MassPu}, \text{logConcPu}) : \text{k-eff}(\text{MassPu}, \text{logConcPu}) \leq t\}$, where t is safety threshold, fixed here at $t = 0.95$. This test case was also presented in Chevalier et al. (2014a) to illustrate batch-sequential SUR strategies. The parameter space here is transformed into the unit square $[0, 1] \times [0, 1]$.

The set of interest is

$$\partial\Gamma^* = \{(\text{MassPu}, \text{logConcPu}) : \text{k-eff}(\text{MassPu}, \text{logConcPu}) = t\},$$

with $t = 0.95$, the level set of k-eff. We are interested in estimating this one-dimensional curve in $\mathbb{R}^2$. Since we have only a few evaluations of the function at points $\mathbf{X}_8 = \{\mathbf{x}_1, \ldots, \mathbf{x}_8\}$, shown in Figure 3.7, a direct estimate of $\partial\Gamma^*$ is not accurate. We rely instead on a random field model Z with prior distribution Gaussian, constant mean and a tensor product Matérn ($\nu = 3/2$) covariance kernel. The parameters of the covariance kernel are estimated by maximum likelihood with the package `DiceKriging`, Roustant et al. (2012). From the posterior distribution of Z, conditioned on evaluations of k-eff
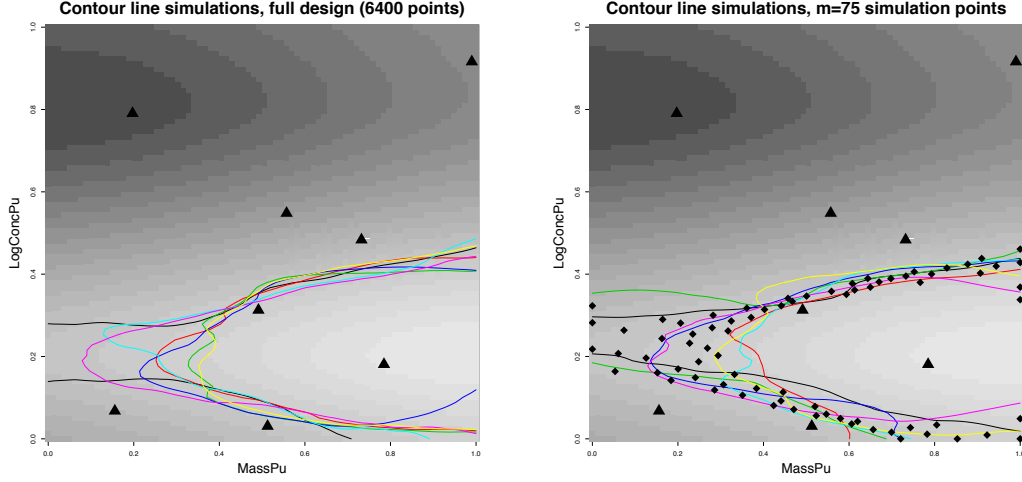
(a) Contour lines realizations obtained with full design simulations. The function k-eff was evaluated at the points denoted by triangles.

(b) Contour lines realizations obtained with simulations at $m = 75$ simulation points (diamonds) and predicted at the full design grid.

Figure 3.7: Realizations $\partial\Gamma$: full design simulations (a) and re-interpolated simulations at 75 locations (b).

at $\mathbf{X}_8$, it is possible to estimate $\partial\Gamma^*$. A plug-in estimate of $\partial\Gamma^*$ could be generated with the posterior mean $\mathfrak{m}_n$; however, this procedure alone does not provide a quantification of the uncertainties. Instead, from the posterior field we generated several realizations of $\partial\Gamma = \{(\mathrm{MassPu}, \mathrm{logConcPu}) : Z(\mathrm{MassPu}, \mathrm{logConcPu}) \mid (Z_{\mathbf{X}_8} = \mathrm{k\text{-}eff}(\mathbf{X}_8)) = 0.95\}$. This procedure requires simulations of the posterior field at high-quality grids; however, even in a two-dimensional parameter space, the procedure is computationally burdensome. In fact, while a discretization on a grid 50×50 delivers a low-quality approximation, simulations of the field at such grids are already expensive to compute. For this reason we choose to simulate the field at m appropriate simulation points and to predict the full simulations with the linear interpolator $\widetilde{Z}$ introduced in Equation (3.1).

Figure 3.7 shows few realizations of $\partial\Gamma$ discretized on an 80×80 grid, obtained with simulations of the field at all points of the design (Figure 3.7a) and with simulations at 75 simulation points, chosen with Algorithm B (Figure 3.7b). The two sets of curves seem to share similar properties. The expected distance in measure between $\partial\Gamma$ and $\partial\widetilde{\Gamma}$, as introduced in Definition 15, could be used here to quantify this similarity however, here we propose to use the arc length of each curve, defined as follows, as it is easier
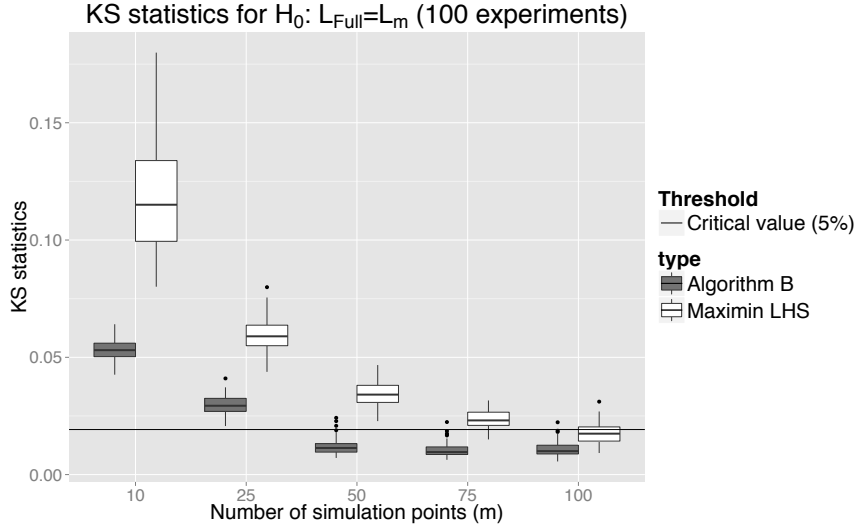
Figure 3.8: Distributions of KS statistic computed over 100 experiments for Kolmogorov-Smirnov test with null hypothesis $H_0 : L_{\text{full}} = L_m$. Simulation points chosen with Algorithm B or with maximin LHS design.

to interpret in our application.

Consider a regular grid $G = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$. For each realization, we select the points $G_{\partial\Gamma} = \{\mathbf{u} \in G : Z_{\mathbf{u}} \mid (Z_{\mathbf{X}_8} = \text{k-eff}(\mathbf{X}_8)) \in [0.95 - \varepsilon, 0.95 + \varepsilon]\}$, where ε is small. $G_{\partial\Gamma}$ contains all the points of the discrete design that have response ε−close to the target. We order the points in $G_{\partial\Gamma}$ in such a way that $\{\mathbf{u}_{i_1}, \ldots, \mathbf{u}_{i_J}\}$ are vertices of a piecewise linear curve approximating $\partial\Gamma$. We approximate the arc length of the curve with the sum of the segments' lengths: $\tilde{l}(\partial\Gamma) = \sum_{G_{\partial\Gamma}} \|\mathbf{u}_{i_{j-1}} - \mathbf{u}_{i_j}\|$. By computing the length for each realization we obtain a Monte Carlo estimate of the distribution of the arc length. We can now compare the distributions of the arc length obtained from reconstructed realizations simulated at few locations with the distribution obtained from simulations at the full grid in order to select the number of simulation points that leads to quasi-realizations for $\partial\Gamma$ whose length is indistinguishable from the full grid realizations' length.

Let us define the random variables $L_{\text{full}} = \tilde{l}(\partial\Gamma_{6400})$ and $L_m = \tilde{l}(\partial\Gamma_m)$, the arc lengths of the random set generated with full design simulations $(80 \times 80$ grid$)$ and the length of the random set generated with simulations at m points respectively. We compare the distributions of L_{full} and L_m with Kolmogorov-Smirnov tests for several values of m. The null hypothesis is $H_0 : L_{\text{full}} = L_m$. The distributions are approximated with $10,000$ simulations, either at the full grid design or at the selected m points. For each m, 100

repetitions of the experiment were computed, thus obtaining a distribution for the Kolmogorov-Smirnov (KS) statistic. Figure 3.8 shows the value of the KS statistic for each m, where the simulation points are obtained either with Algorithm B or with a maximin LHS design. For $m \geq 50$ optimized points, the KS statistic is below the critical value for at least 97% of the experiments, and thus it is not possible to distinguish the two length distributions with a significance level of 5%. If the simulation points are chosen with a maximin LHS design instead, the KS statistic is below the critical value for at least 67% of the experiments with $m = 100$ simulation points, as also shown in Figure 3.8. This result shows again the importance of choosing optimized simulation points. The approximation of L_{full} with L_m leads to substantial computational time savings. The computational time for $10,000$ simulations of the field at the full grid design ($6,400$ points) is 466 seconds, while the total time for finding 75 appropriate simulation points (with Algorithm B), simulate the field at these locations and re-interpolate the field at the full design is 48.7 seconds (average over 100 experiments).

The expected distance in measure introduced in Section 3.2.1 could also be used here to quantify how far the quasi-realizations are from the full grid realizations.

3.7 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. Often, however, the random field needs to be simulated at a fine design to obtain meaningful realizations 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 quasi-realizations of a conditional GRF that mitigates this problem. While the approach of predicting realizations of the field from simulations at few locations has already been introduced in the literature, this is the first attempt to define optimal simulation points based on a specific distance between random closed sets: the expected distance in measure. We showed in several examples that the quasi-realizations method reduces the computational cost due to conditional simulations of the field, however it does so relying on an approximation. In particular the random set quasi-realizations optimality with respect to the expected distance in measure does not necessarily guarantee that other properties of the set are correctly reproduced.

The quasi-realizations approach allowed us to study an uncertainty mea-

sure that, to the best of our knowledge, was not previously used in practice: the distance average variability. The estimation of the distance average variability is appealing when realizations of the excursion set on fine grids are computationally cheap. We showed on a two-dimensional test function that it is possible to reduce computational costs by at least one order of magnitude, thus making this technique practical. In general the quasi-realizations approach could improve the speed of distance average based methods as, for example, Jankowski and Stanberry (2010) and Jankowski and Stanberry (2012).

We presented a test case in safety engineering where we estimated the arc length's distribution of a level set in a two-dimensional parameter space. The level set was approximated by piecewise linear curve, the resolution of which depends on the simulation design. A Monte Carlo technique based on realizations of the excursion set obtained with full design simulations is computationally too expensive at high resolutions. Reconstructed simulations from simulations of the field at few well-chosen points re-interpolated on a fine design made this application possible. In particular we showed that the distribution of the arc length obtained with a full design simulation at a rough design, an 80×80 grid, was not significantly different than the distribution obtained from reconstructed sets with simulations at $m = 50$ well-chosen points, thus opening the way for estimates on higher resolution grids.

Conditional realizations of the excursion set can also be used to estimate the volume of excursion; in the appendix we show how to handle this problem with Monte Carlo simulations at fine designs.

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.

3.8   Appendix A: sketch of proof for Proposition 3

Here we present a sketch of proof based on a contradiction argument. For a more detailed and constructive proof see Azzimonti et al. (2016a), Appendix A.

Let us first assume that $s_{n,m}^2 \to 0$ μ-almost everywhere. The expected distance in measure can be rewritten, according to Equation (3.2), 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 $\rho_{n,m}(\mathbf{x}) \le 1$, it is sufficient by the dominated convergence theorem to prove that $\rho_{n,m} \to 0$ μ-almost everywhere.

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

$$\rho_{n,m}(\mathbf{x}) \le P_n\big(|Z(\mathbf{x}) - t| \le w\big) + P_n\big(|\widetilde{Z}(\mathbf{x}) - Z(\mathbf{x})| \ge w\big)$$
$$\le \frac{2w}{\sqrt{2\pi s_n^2(\mathbf{x})}} + \frac{s_{n,m}^2(\mathbf{x})}{w^2}.$$

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

$$\rho_{n,m}(\mathbf{x}) \le \frac{2\sqrt{s_{n,m}(\mathbf{x})}}{\sqrt{2\pi s_n^2(\mathbf{x})}} + s_{n,m}(\mathbf{x}) \to 0. \tag{3.7}$$

Since $s_{n,m}^2 \to 0$ μ-almost everywhere and $\rho_{n,m}(\mathbf{x}) = 0$ wherever $s_n^2(\mathbf{x}) = 0$, Equation (3.7) proves the sufficiency part of Proposition 3.

Conversely, let us assume by contradiction that $s_{n,m}(\mathbf{x}) \to c(\mathbf{x}) > 0$ for all $\mathbf{x}$ in a set $C \subset D$ with $\mu(C) > 0$. Moreover, there exists a set $B \subset C$, with $\mu(B) > 0$ such that $c(\mathbf{x}) > c_0$, for some $c_0 > 0$. Without loss of generality in what follows we restrict ourselves to B. Consider an arbitrary $\mathbf{x}_0 \in B$ then

$$\rho_{n,m}(\mathbf{x}_0) = P_n\left(Z(\mathbf{x}_0) \ge t, \widetilde{Z}(\mathbf{x}_0) < t\right) + P_n\left(Z(\mathbf{x}_0) < t, \widetilde{Z}(\mathbf{x}_0) \ge t\right)$$
$$\ge P_n\left(Z(\mathbf{x}_0) \ge t, \widetilde{Z}(\mathbf{x}_0) < t\right) = P_n\left(Z(\mathbf{x}_0) \ge t, Z(\mathbf{x}_0) < t + \epsilon_{n,m}(\mathbf{x}_0)\right)$$

$$= P_n\left(t \le Z(\mathbf{x}_0) < t + \epsilon_{n,m}(\mathbf{x}_0)\right),$$

where $\epsilon_{n,m}(\mathbf{x}_0) = Z(\mathbf{x}_0) - \widetilde{Z}(\mathbf{x}_0)$. Moreover we have

$$\rho_{n,m}(\mathbf{x}_0) \ge P_n\big(t \le Z(\mathbf{x}_0) < t + \epsilon_{n,m}(\mathbf{x}_0)\big)$$
$$= P_n\left(t \le Z(\mathbf{x}_0) < t + \epsilon_{n,m}(\mathbf{x}_0) \mid \epsilon_{n,m}(\mathbf{x}_0) \ge 0\right) P_n\left(\epsilon_{n,m}(\mathbf{x}_0) \ge 0\right)$$
$$+ P_n\left(t \le Z(\mathbf{x}_0) < t + \epsilon_{n,m}(\mathbf{x}_0) \mid \epsilon_{n,m}(\mathbf{x}_0) < 0\right) P_n\left(\epsilon_{n,m}(\mathbf{x}_0) < 0\right)$$
$$\ge P_n\left(\epsilon_{n,m}(\mathbf{x}_0) \ge 0\right) \int_0^\infty P_n\big(t \le Z(\mathbf{x}_0) < t + s_{n,m}(\mathbf{x}_0)u\big) \varphi(u) du$$
$$\ge \frac{1}{2} \int_0^\infty P_n\big(t \le Z(\mathbf{x}_0) < t + c(\mathbf{x}_0)u\big) \varphi(u) du = \kappa(\mathbf{x}_0) > 0.$$

The function $\kappa(\mathbf{x}_0)$ is strictly positive for $\mathbf{x}_0 \in B$, moreover notice that for any $u \geq 0$

$$P_n\big(t \leq Z(\mathbf{x}_0) < t+c(\mathbf{x}_0)u\big) \geq P_n\left(\frac{t - \mathfrak{m}_n(\mathbf{x}_0)}{s_n(\mathbf{x}_0)} \leq N < \frac{t - \mathfrak{m}_n(\mathbf{x}_0)}{s_n(\mathbf{x}_0)} + \frac{c_0 u}{s_n(\mathbf{x}_0)}\right),$$

where N is a standard Normal random variable. Since $\mathfrak{m}_n$ and s_n are continuous, we can lower bound this probability with a positive value independent of $\mathbf{x}_0$, by selecting the smallest interval farther away from zero. In particular, consider $\zeta_1 = \inf_D \frac{t-\mathfrak{m}_n}{s_n}$ and $\zeta_2 = \sup_D \frac{t-\mathfrak{m}_n}{s_n}$. Fix ζ equal to the value ζ_i with the largest absolute value. We have then

$$\kappa(\mathbf{x}_0) \geq \frac{1}{2}\int_0^\infty P_n\left(\frac{t - \mathfrak{m}_n(\mathbf{x}_0)}{s_n(\mathbf{x}_0)} \leq N < \frac{t - \mathfrak{m}_n(\mathbf{x}_0)}{s_n(\mathbf{x}_0)} + \frac{c_0 u}{s_n(\mathbf{x}_0)}\right)\varphi(u)du$$

$$> \frac{1}{2}\int_0^\infty P_n\left(\zeta \leq N < \zeta + \frac{c_0 u}{\inf_D s_n}\right)\varphi(u)du = \kappa_0 > 0$$

With this κ_0 we have then,

$$\mu\left(A_{n,m}^{\kappa_0}\right) \nrightarrow 0, \text{ for } m \to +\infty \qquad \text{where } A_{n,m}^{\kappa_0} = \{\mathbf{x} \in D : \rho_{n,m}(\mathbf{x}) \geq \kappa_0\},$$

which contradicts the sufficient condition.

3.9   Appendix B: estimating the distribution of a volume of excursion in six dimensions

In this section we show how it is possible to estimate the conditional distribution of the 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 3.2, 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 here we use Monte Carlo simulations to work out the posterior distribution of this volume (see Vazquez and Martinez, 2006; Adler, 2000). 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 mitigates this problem on a six-dimensional example.

We consider the analytical test function $h(\mathbf{x}) = -\log(-\operatorname{Hartman}_6(\mathbf{x}))$, where $\operatorname{Hartman}_6$ is the six-dimensional Hartman function (see Jones et al., 1998) 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 region 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 quasi-realizations of the excursion set. 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 (Scheuerer, 2009). This introduces a systematic bias on the volume of excursion for each realization because subsets of the excursion sets induced by small rougher variations of the true Gaussian field may not be intercepted by $\widetilde{Z}$. The effect changes the mean of the distribution, but it does not seem to influence the variance of the distribution. In the present setting 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 discretization design G, of size r, the distribution of the volume of excursion is obtained with the following steps: first the mean volume of excursion is estimated by integrating the coverage function of Γ over G; second the distribution is obtained by computing the volume of excursion for each quasi-realization of the excursion set; finally the distribution is centered in the mean value obtained with the first step. Figure 3.9a shows the absolute error on the mean between the full design simulation and the approximate simulations with and without bias correction. The optimal simulation points are computed with Algorithm B because for a large number of points it achieves results very similar to Algorithm A but at the same time the optimized points are much cheaper to compute, as shown 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 $V_m =$
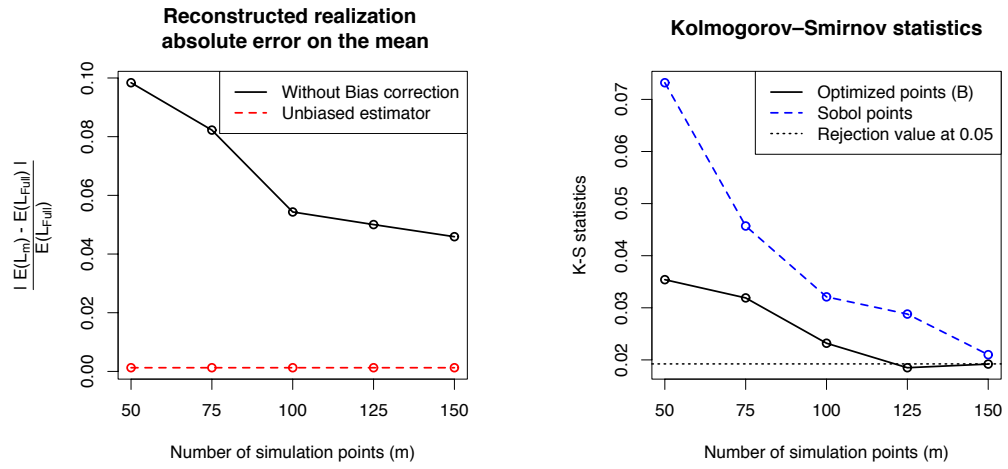
(a) Absolute error on the mean estimate between V_m and V_{full}. The red line is the error on the unbiased estimator, the black line represents the error on the estimator without bias correction.

(b) Kolmogorov-Smirnov statistic for testing the hypothesis $H_0 : V_m = V_{full}$. Simulations at points obtained with Algorithm B (black full line) or with space filling points (blue dashed line). The dotted horizontal line is the rejection value at level 0.05.

Figure 3.9: Analysis of volume distributions.

$\mu(\Gamma(\mathbf{E}_m))$ 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 two-sample Kolmogorov-Smirnov tests. Figure 3.9b shows the values of the Kolmogorov-Smirnov statistic for testing the null hypothesis $H_0 : V_m = V_{full}$, 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. 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 quasi-realizations from simulations at few optimal locations. In fact, the computational costs are significantly reduced by interpolating simulations: the CPU time needed to simulate on the full $10,000$ points design is $60,293$ 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 at the University of Bern with Intel Xeon E5649 2.53-GHz CPUs with 4 GB RAM.

Chapter 4

Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation

4.1 Introduction

Conservative estimates are a technique to estimate excursion sets of expensive to evaluate functions that aims at controlling the error of the estimate with a probabilistic condition. In chapter 2 we revisited this concept in a general framework for random closed sets, see, e.g. Definition 14, section 2.3. In Section 2.4.3 we specialized this concept to excursion set estimation with Gaussian random field (GRF) priors. In this chapter we focus on the computational challenges brought forward by this estimate and we introduce an algorithm to efficiently compute the orthant probability central to the estimation procedure. Let us briefly recall the framework. We are interested in estimating the set

$$\Gamma^\star = \{x \in D : f(x) \leq t\},$$

where $t \in \mathbb{R}$, $D \subset \mathbb{R}^d$ and $f : D \to \mathbb{R}$ is a continuous function which is expensive-to-evaluate. We consider the situation where only $n \geq 1$ evaluations of the function f are given at points $\mathbf{X}_n = \{x_1, \ldots, x_n\}$ and we denote

them with $\mathbf{f}_n = (f(x_1), \dots, f(x_n)) \in \mathbb{R}^n$.

In a Bayesian framework we consider f as one realization of a GRF $(Z_x)_{x \in D}$ with prior mean function $\mathfrak{m}$ and covariance kernel $\mathfrak{K}$. As detailed in Section 2.4, a prior distribution of the excursion set is obtained by thresholding Z, obtaining

$$\Gamma = \{x \in D : Z_x \leq t\}.$$

Letting $Z_{\mathbf{X}_n}$ denote the random vector $(Z_{x_1}, \dots, Z_{x_n})$, we can then condition Z on the observations $\mathbf{f}_n$ and obtain a posterior distribution for the field $Z_x \mid Z_{\mathbf{X}_n} = \mathbf{f}_n$. This gives rise to a posterior distribution for Γ. As in Section 2.4 we define as conservative estimate for $\Gamma^\star$, a solution of the optimization problem

$$\mathrm{CE}_\alpha(\Gamma) \in \arg\max_{K \in \mathfrak{C}_\rho} \{\mu(K) : P_n(K \subset \Gamma) \geq \alpha\}, \qquad (4.1)$$

where μ is a finite measure on D, $P_n(\cdot) = P(\cdot \mid Z_{\mathbf{X}_n} = \mathbf{f}_n)$ and $\mathfrak{C}_\rho$ is a parametric family of sets where the optimization procedure is conducted. Conservative estimates were introduced in Bolin and Lindgren (2015) for Gaussian Markov random fields however for general Gaussian process models they lead to two major computational issues.

First of all we need to select a family of sets to use for the optimization procedure in Equation (4.1). Here we follow Bolin and Lindgren (2015) and we select the Vorob'ev quantiles $Q_\rho, \rho \in [0, 1]$ as family of sets. See also Section 2.4 and Section 2.3.2 for details on this parametric family and Chapter 5 for a more detailed study on this choice. Among the advantages of the Vorob'ev family the fact that it is parametrized by one real number ρ makes the optimization straightforward. The conservative estimate is then the result of the optimization procedure in Equation (4.1). In order to solve this maximization we propose in Section 4.4 a dichotomy algorithm.

The second computational issue is central to this chapter. In fact, for each candidate Q we need to evaluate $P_{\mathrm{next}} = P_n(Q \subset \{Z_x \leq t\})$, the probability that Q is inside the excursion. This quantity can be approximated with a high dimensional Gaussian orthant probability. If the set Q is discretized over the points $c_1, \dots, c_r$, then

$$P_n(Q \subset \{Z_x \leq t\}) = P_n(Z_{c_1} \leq t, \dots, Z_{c_r} \leq t). \qquad (4.2)$$

In this chapter we introduce an algorithm to estimate the probability in Equation (4.2) efficiently, especially when r is high, e.g. $r \gg 500$.

The problem of estimating Gaussian orthant probabilities is more general than the setup just presented. In the remainder of the chapter we consider the random vector $\mathbf{X} = (X_1, \dots, X_d)$ and we assume that its distribution is

Gaussian, $\mathbf{X} \sim N_d(\mathbf{m}, \Sigma)$, $d > 1$. We are interested in estimating, for any fixed $t \in \mathbb{R}$, the following probability

$$\pi(t) = P(\mathbf{X} \le (t, \ldots, t)). \qquad (4.3)$$

The general problem of evaluating $\pi(t)$, which, for a full rank matrix Σ, is the integral of the multivariate normal density $\varphi(\cdot; \mathbf{m}, \Sigma)$ over the one-sided d-dimensional rectangle $(-\infty, t]^d$, has been extensively studied in moderate dimensions with many different methods. In low dimensions tables are available (see, e.g., Owen (1956) for $d = 2$). Furthermore, when the dimension is smaller than 20, there exist methods (see, e.g., Abrahamson (1964), Moran (1984) and Miwa et al. (2003)) exploiting the specific orthant structure of the probability in Equation (4.3). Currently, however, most of the literature uses numerical integration techniques to approximate the quantity. In moderate dimensions fast reliable methods are established to approximate $\pi(t)$ (see, e.g. Cox and Wermuth (1991)) and more recently the methods introduced in Schervish (1984); Genz (1992) and Hajivassiliou et al. (1996) (see also Genz and Bretz (2002), Ridgway (2016), Botev (2017) and the book Genz and Bretz (2009) for a broader overview) provide state-of-the-art algorithms when $d < 100$. Those techniques rely on fast quasi Monte Carlo (qMC) methods and are very accurate for moderate dimensions. Recently, the introduction of component-by-component randomly shifted lattice rules decreased the rate of convergence qMC based methods, see, e.g., Dick et al. (2013); Kuo and Sloan (2005); Nichols and Kuo (2014) and references therein. However, when d is larger than 1000 the current implementations of such methods are not computationally efficient or become intractable. Commonly used alternative methods are standard Monte Carlo (MC) techniques (see Tong (2012), Chapter 8 for an extensive review), for which getting accurate estimates can be computationally prohibitive.

We propose here a two step method that exploits the power of qMC quadratures and the flexibility of stochastic simulation. We rely on the following equivalent formulation.

$$\pi(t) = 1 - P(\max \mathbf{X} > t),$$

where $\max \mathbf{X}$ denotes $\max_{i=1,\ldots,d} X_i$. In the following we fix t, we denote $p = P(\max \mathbf{X} > t)$ and $E = \{1, \ldots, d\}$.

The central idea here is using a moderate dimensional subvector of $\mathbf{X}$ to approximate p and then correcting bias by Monte Carlo integration. Let us fix $q \ll d$ and define the active dimensions as $E_q = \{i_1, \ldots, i_q\} \subset E$. Let us further denote with $\mathbf{X}^q$ the q dimensional vector $\mathbf{X}^q = (X_{i_1}, \ldots, X_{i_q})$ and

with $\mathbf{X}^{-q}$ the $(d-q)$ dimensional vector $\mathbf{X}^{-q} = (X_j)_{j \in E \setminus E_q}$. Then,

$$p = P(\max \mathbf{X} > t) = p_q + (1 - p_q)R_q, \tag{4.4}$$
$$p_q = P(\max \mathbf{X}^q > t),$$
$$R_q = P(\max \mathbf{X}^{-q} > t \mid \max \mathbf{X}^q \le t).$$

The quantity p_q is always smaller or equal than p as $E_q \subset \{1, \ldots, d\}$. Selecting a non-degenerate vector $\mathbf{X}^q$, we propose to estimate p_q with the QRSVN algorithm (Genz et al., 2012) which is efficient as we choose a number of active dimensions q much smaller than d.

In Chevalier (2013), Chapter 6, the similar problem of approximating the non-exceedance probability of the maximum of a GRF Z based on a few well-selected points is presented. In that setting each component of $\mathbf{X}$ stands for the value of Z at one point of a discretization of the index set. Active dimensions (i.e. the well-selected points) were chosen by numerically maximizing p_q, and the remainder was not accounted for. Here a full optimization of the active dimensions is not required as we, instead, exploit the decomposition in Equation (4.4) to correct the bias introduced by p_q. Nonetheless the number of active dimensions and the dimensions themselves play an important role in the algorithm as explained in Section 4.2.1. In order to correct the bias of p_q we propose to estimate the reminder R_q with a standard MC technique or with a novel asymmetric nested Monte Carlo (anMC) algorithm. The anMC technique draws samples by taking into account the computational cost, resulting in a more efficient estimator.

In the remainder of the chapter, we propose an unbiased estimator for p and we compute its variance in Section 4.2. In Section 4.3 we introduce the anMC algorithm in the more general setting of estimating expectations depending on two vectors with different simulation costs. It is then explicitly applied to efficiently estimate R_q. Finally, in Section 4.4, we show an implementation of this method to compute conservative estimates of excursion sets for expensive to evaluate functions under non-necessarily Markovian GRF priors. All proofs are in Appendix 4.6.

4.2   An unbiased estimator for p

Equation (4.4) gives us a decomposition that can be exploited to obtain an unbiased estimator for p. In the following proposition we define the estimator and we compute its variance.

Proposition 4. *Consider $\widehat{p_q}$ and $\widehat{R_q}$, independent unbiased estimators of p_q and R_q respectively, then $\widehat{p} = \widehat{p_q} + (1 - \widehat{p_q})\widehat{R_q}$ is an unbiased estimator for p.*

Moreover its variance is

$$\mathrm{Var}(\widehat{p}) = (1 - R_q)^2 \, \mathrm{Var}(\widehat{p_q}) + (1 - p_q)^2 \, \mathrm{Var}(\widehat{R_q}) + \mathrm{Var}(\widehat{p_q}) \, \mathrm{Var}(\widehat{R_q}). \quad (4.5)$$

In what follows we present options for $\widehat{p_q}$ and $\widehat{R_q}$ that form an efficient computational strategy.

4.2.1 Quasi Monte Carlo estimator for p_q

The quantity p_q can also be computed as

$$p_q = 1 - P\left(\mathbf{X}^q \le t_q\right),$$

where t_q denotes the q dimensional vector $(t, \dots, t)$. The approximation of p_q thus requires only an evaluation of the c.d.f. of $\mathbf{X}^q$. Since we assume that $q \ll d$, then the dimension is low and we propose to estimate p_q with the estimator $\widehat{p_q}$ that uses the method QRSVN introduced in Genz (1992), Hajivassiliou et al. (1996) and refined in Genz and Bretz (2009). This method computes a randomized quasi Monte Carlo integration of the normal density. In particular we consider here the implementation of QRSVN with the variable reordering described in Genz and Bretz (2009, Section 4.1.3). The estimate's error is approximated with the variance of the randomized integration. The quantity $\widehat{p_q}^{\mathrm{G}}$ obtained with this procedure is an unbiased estimator of p_q, see Genz and Bretz (2009).

In order to compute $\widehat{p_q}^{\mathrm{G}}$ we need to choose q, the number of active dimensions, and then select the dimensions themselves. The decomposition of p in Equation (4.4) does not require optimal choices for those parameters as the remainder $\widehat{R_q}$ is estimated afterwards. Here we propose two heuristic techniques to make these choices in practice. Let us recall that the decomposition in Equation (4.4) leads to computational savings if we can approximate most of p with p_q for a small q. On the other hand a large number of active dimensions allows to intercept most of the probability mass in p. Here we adopt a heuristic approach and we select q by sequentially increasing the number of active dimensions until the relative change of $\widehat{p_q}^{\mathrm{G}}$ is less than the estimate's error. We detail the procedure in Algorithm 1.

The choice of q also influences the behaviour of the estimator for R_q. This aspect is discussed in more details in the next section. Let us now consider q fixed and let us study how to select the active dimensions in Algorithm 1.

The choice of active dimensions E_q plays an important role in the approximation of p because it determines the error $\widehat{p_q} - p$. Since this error

---

Algorithm 1: Select q, active dimensions and compute $\widehat{p}_q^{\,\mathrm{G}}$.

Input : q_0, small initial q, e.g. $q_0 = 10$, and q_{step} the increment of q

Output: q, $\widehat{p}_q^{\,\mathrm{G}}$

Compute $\widehat{p}_{q_0}^{\,\mathrm{G}}$ and save $\mathrm{err}(\widehat{p}_{q_0}^{\,\mathrm{G}}) = 3\sqrt{\mathrm{Var}(\widehat{p}_{q_0}^{\,\mathrm{G}})}$;

initialize $k = 0$;

repeat

    increment $k = k + 1$;

    $q_k := q_0 + k q_{\mathrm{step}}$;

    choose q_k active dimensions, compute $\widehat{p}_{q_k}^{\,G}$ and $\mathrm{err}(\widehat{p}_{q_k}^{\,G})$;

    compute $\Delta(\widehat{p}_{q_k}^{\,\mathrm{G}}) = \dfrac{\left| \widehat{p}_{q_k}^{\,\mathrm{G}} - \widehat{p}_{q_{k-1}}^{\,\mathrm{G}} \right|}{1 + \widehat{p}_{q_k}^{\,\mathrm{G}}}$;

until $\Delta(\widehat{p}_{q_k}^{\,G}) < \mathrm{err}(\widehat{p}_{q_k}^{\,G})$;

$q = q_k$ and $\widehat{p}_q^{\,\mathrm{G}} = \widehat{p}_{q_k}^{\,\mathrm{G}}$;

---

is always negative, we implement procedures to select E_q that exploit this property. Selecting E_q such that $P(\max \mathbf{X}^q > t)$ is numerically maximized, as in Chevalier (2013), optimally reduces the bias of $\widehat{p}_q$ as an estimator for p. Here we are not interested in a fully fledged optimization of this quantity as the residual bias is removed with the subsequent estimation of R_q, therefore, we exploit fast heuristics methods.

The basic tool used here is the excursion probability function:

$$p_t(i) = P(X_i > t) = \Phi\left(\frac{\mathbf{m}_i - t}{\sqrt{\Sigma_{i,i}}} \right),$$

where Φ is the standard normal c.d.f. The function p_t is widely used in spatial statistics (see, e.g. Bolin and Lindgren, 2015) and Bayesian optimization (see, e.g. Kushner, 1964; Bect et al., 2012). In our setting it can be used to identify the dimensions where we have a high probability of exceeding the threshold. The indices that realize a high value for p_t enable the identification of dimensions that actively contribute to the maximum. We propose the following methods.

Method A: sample q indices with probability given by p_t.

Method B: sample q indices with probability given by $p_t(1 - p_t)$.

These methods require only one evaluation of the normal c.d.f. at each element of the vector $\left(\frac{\mathbf{m}_i - t}{\sqrt{\Sigma_{i,i}}} \right)_{i=1,\dots,d}$, and are thus very fast. Both methods were already introduced for sequential evaluations of expensive to evaluate functions, see, e.g., Chevalier et al. (2014c).
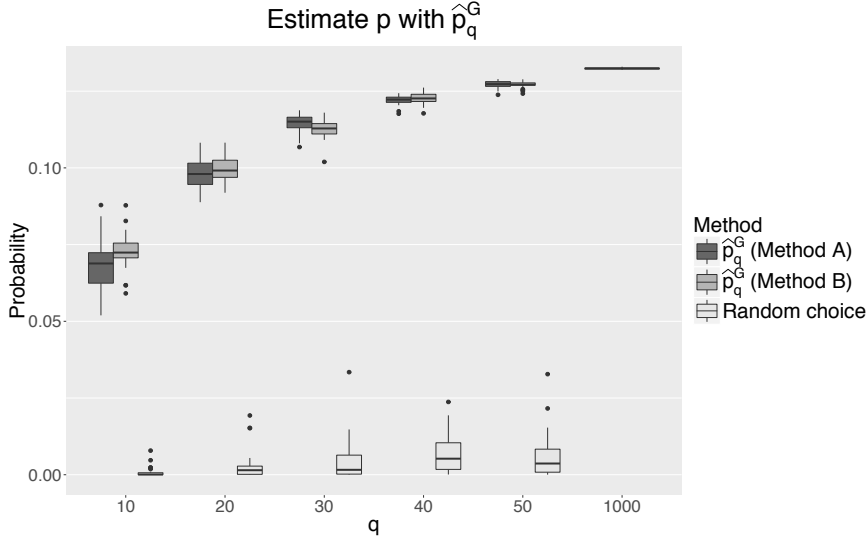
Figure 4.1: Distribution of $\widehat{p}_q{}^{\mathrm{G}}$ estimates obtained with different choices of active dimensions.

Figure 4.1 shows a comparison of the estimates p_q obtained with different methods to select E_q. We consider 30 replications of an experiment where $\widehat{p}_q{}^{\mathrm{G}}$ is used to approximate p. The dimension of the vector $\mathbf{X}$ is $d = 1000$, the threshold is fixed at $t = 11$. The vector $\mathbf{X}$ is obtained from a discretization of a six dimensional GRF on the first 1000 points of the Sobol' sequence (Bratley and Fox, 1988). The GRF has a tensor product Matérn ($\nu = 5/2$) covariance kernel and a non constant mean function $\mathfrak{m}$. The covariance kernel's hyperparameters are fixed as $\theta = [0.5, 0.5, 1, 1, 0.5, 0.5]^T$ and $\sigma^2 = 8$, see Rasmussen and Williams (2006), Chapter 4, for details on the parametrization. The two methods clearly outperform a random choice of active dimensions.

Methods A and B work well for selecting active dimensions when the mean vector $\mathbf{m}$ and the diagonal of the covariance matrix are anisotropic. In this situation both methods selects dimensions that are a good trade-off between a high variance and mean close to t. These dimensions usually contribute strongly to the total probability.

If the random vector $\mathbf{X}$ is a GRF discretized over a set of points $E_{\mathrm{spat}} = \{e_1, \ldots, e_d\} \subset \mathbb{R}^l$, then we can exploit this information when choosing E_q. Let us consider the sequence of vectors $(\boldsymbol{\delta}_j)_{j=1,\ldots,q}$, defined for each j as

$$\boldsymbol{\delta}_j = \prod_{k=1}^{j} \mathrm{dist}(e_{i_k}, E_{\mathrm{spat}}) \quad (j = 1, \ldots, q)$$

where $\mathrm{dist}(e_{i_k}, E_{\mathrm{spat}})$ denotes the d-dimensional vector of Euclidean distances

between e_{i_k} and each point in E_{spat} and $\{e_{i_1}, \ldots, e_{i_q}\}$ are the points corresponding to the selected active dimensions E_q. Methods A and B can then be adjusted by sampling the jth active dimension with probabilities given by $p_t \frac{\boldsymbol{\delta}_j}{\|\boldsymbol{\delta}_j\|}$ and $p_t(1-p_t)\frac{\boldsymbol{\delta}_j}{\|\boldsymbol{\delta}_j\|}$ respectively, where the product is intended component-wise.

In the example presented, the estimator $\widehat{p}_q^{\,\mathrm{G}}$ gives an inexpensive approximation of most of the probability mass with as few as 40 active dimensions, however it is intrinsically biased as an estimator of p. The remainder R_q enables to correct the bias of this first step.

4.2.2   Monte Carlo estimator for R_q

Debiasing $\widehat{p}_q^{\,\mathrm{G}}$ as an estimator of p can be done at the price of estimating

$$R_q = P\left(\max \mathbf{X}^{-q} > t \mid \max \mathbf{X}^q \leq t\right).$$

There is no closed formula for R_q, so it is approximated here via MC. Since $\mathbf{X}$ is Gaussian then so are $\mathbf{X}^q$, $\mathbf{X}^{-q}$ and $\mathbf{X}^{-q} \mid \mathbf{X}^q = x^q$, for any deterministic vector $x^q \in \mathbb{R}^q$.

In order to estimate $R_q = P\left(\max \mathbf{X}^{-q} > t \mid X_{i_1} \leq t, \ldots, X_{i_q} \leq t\right)$, we first generate n realizations $\mathbf{x}_1^q, \ldots, \mathbf{x}_n^q$ of $\mathbf{X}^q$ such that $\mathbf{X}^q \leq t_q$. Second, we compute the mean and covariance matrix of $\mathbf{X}^{-q}$ conditional on each realization $\mathbf{x}_l^q$, $l = 1, \ldots, n$ with the following formulas

$$\mathbf{m}^{-q|\mathbf{x}_l^q} = \mathbf{m}^{-q} + \Sigma^{-q,q}(\Sigma^q)^{-1}(\mathbf{x}_l^q - \mathbf{m}^q), \qquad \Sigma^{-q|q} = \Sigma^{-q} - \Sigma^{-q,q}(\Sigma^q)^{-1}\Sigma^{q,-q}, \tag{4.6}$$

where $\mathbf{m}^q, \Sigma^q$ and $\mathbf{m}^{-q}, \Sigma^{-q}$ are the mean vector and covariance matrix of $\mathbf{X}^q$ and $\mathbf{X}^{-q}$ respectively, $\Sigma^{-q,q}$ is the cross-covariance between the dimensions $E \setminus E_q$ and E_q, $\Sigma^{q,-q}$ is the transpose of $\Sigma^{-q,q}$. Given the mean and covariance matrix conditional on each sample $\mathbf{x}_l^q$, we can easily draw a realization $\mathbf{y}_l^{-q|q}$ from $\mathbf{X}^{-q} \mid \mathbf{X}^q = \mathbf{x}_l^q$. Once n couples $(\mathbf{x}_l^q, \mathbf{y}_l^{-q|q})$, $l = 1, \ldots, n$ are drawn from the respective distributions, an estimator for R_q is finally obtained as follows

$$\widehat{R}_q^{\,\mathrm{MC}} = \frac{1}{n}\sum_{l=1}^{n}\mathbf{1}_{\max \mathbf{y}_l^{-q|q} > t}.$$

The realizations of $\mathbf{X}^q$ truncated below t_q are obtained with a crude rejection sampling algorithm. The cost of this step is driven by the number of samples of $\mathbf{X}^q$ rejected and it can thus be very high for small acceptance probabilities. The accepted samples satisfy the condition $\mathbf{X}^q \leq t_q$ thus we have that the acceptance probability is $P(\mathbf{X}^q \leq t_q) = 1 - p_q$. This shows that

Estimate p with $\widehat{p}^{\,\text{GMC}}$
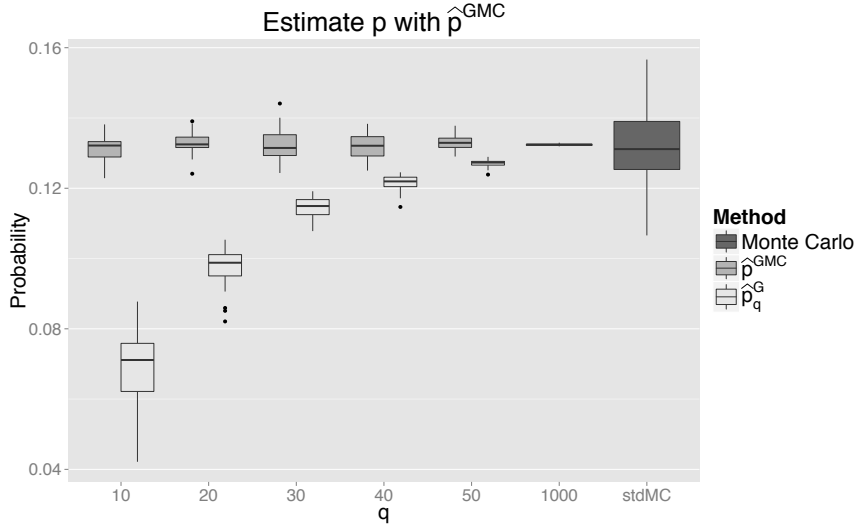
Figure 4.2: Estimate of p with $\widehat{p}^{\,\text{GMC}}$ for different values of q. A full MC estimation of the same quantity is shown for comparison

the choice of q and of the active dimensions plays an important role. If p_q is much smaller than p, then the rejection sampler will have a high acceptance probability, however the overall method will be less efficient as most of the probability is estimated in the remainder. On the other hand a choice of q and of active dimensions that leads to a value p_q very close to p will also lead to a slower rejection sampler as the acceptance probability would be small.

The second part of the procedure for $\widehat{R_q}^{\text{MC}}$, drawing samples from the distribution of $\mathbf{X}^{-q} \mid \mathbf{X}^q = \mathbf{x}_l^q$, is instead less dependent on q and generally less expensive than the fist step. The mean vector and covariance matrix computations requires only linear algebra operations as described in Equation (4.6) and realizations of $\mathbf{X}^{-q} \mid \mathbf{X}^q = \mathbf{x}_l^q$ can be generated by sampling from a multivariate normal distribution.

The difference in computational cost between the first step and the second step of the MC procedure can be exploited to obtain more efficient estimators. In Section 4.3 we present a new MC procedure that at a fixed computational cost reduces the variance of the estimate.

We conclude the section with a small numerical comparison. Let us denote with $\widehat{p}^{\,\text{GMC}}$ the unbiased estimator of p defined as

$$\widehat{p}^{\,\text{GMC}} = \widehat{p_q}^{\,\text{G}} + (1 - \widehat{p_q}^{\,\text{G}})\widehat{R_q}^{\text{MC}}.$$

Figure 4.2 shows the box plots of 30 replications of an experiment where p is approximated with $\widehat{p}^{\,\text{GMC}}$ for different choices of q. The set-up is the

same as in Figure 4.1. The core of the probability is approximated with $\widehat{p_q}^{\mathrm{G}}$ and the active dimensions are chosen with Method 1. The residual R_q is estimated with $\widehat{R_q}^{\mathrm{MC}}$. The remainder allows to correct the bias of $\widehat{p_q}^{\mathrm{G}}$ even with a small number of active dimensions. As comparison the results of the same experiment with a full MC estimator for p are also shown. For all experiments and for each method the number of samples was chosen in order to have approximately the same computational cost. The estimator $\widehat{p}^{\mathrm{GMC}}$ exploits an almost exact method to estimate the largest part of the probability p, therefore the MC estimator $\widehat{R_q}^{\mathrm{MC}}$ has less variance than a full MC procedure for a fixed computational cost.

4.3   Estimation of the residual with asymmetric nested Monte Carlo

In section 4.2, R_q was estimated by $\widehat{R_q}^{\mathrm{MC}}$. There exists many methods to reduce the variance of such estimators, including antithetic variables (Hammersley and Morton, 1956), importance sampling (Kahn, 1950; Kahn and Marshall, 1953) or conditional Monte Carlo (Hammersley, 1956) among many others, see, Lemieux (2009), Chapter 4, and Robert and Casella (2013), Chapter 4, for a broader overview. Here we propose a so-called asymmetric nested Monte Carlo (anMC) estimator for R_q that reduces the variance by a parsimonious multiple use of conditioning data.

The idea is to use an asymmetric sampling scheme that assigns the available computational resources by taking into account also the actual cost of simulating each component. This type of asymmetric sampling scheme was introduced in the particular case of comparing the performance of stopping times for a real-valued stochastic process in discrete times in Dickmann and Schweizer (2016). Here we introduce this procedure in a general fashion and then we detail how to use it as variance reduction for $\widehat{R_q}^{\mathrm{MC}}$. Consider two random elements $W \in \mathcal{W}$ and $Y \in \mathcal{Y}$, defined on the same probability space and not independent. We are interested in estimating

$$G = \mathbb{E}\left[g(W, Y)\right], \tag{4.7}$$

where $g : \mathcal{W} \times \mathcal{Y} \to \mathbb{R}$ is a measurable function, assumed integrable with respect to (W, Y)'s probability measure. Let us also assume that it is possible to draw realizations from the marginal distribution of W, Y and from the conditional distribution of $Y \mid W = w_i$, for each w_i sample of W. We can then obtain realizations (w_i, y_i), $i = 1, \ldots, n$ of (W, Y) by simulating w_i from the

distribution of W and then y_i from the conditional distribution $Y \mid W = w_i$, leading to:

$$\widehat{G} = \frac{1}{n} \sum_{i=1}^{n} g(w_i, y_i). \tag{4.8}$$

This MC estimator can actually be seen as the result of a two step nested MC procedure where, for each realization w_i, one inner sample y_i is drawn from $Y \mid W = w_i$. Note that the estimator $\widehat{R_q}^{\mathrm{MC}}$ used in Section 4.2 is a particular case of Equation (4.8) with $W = \mathbf{X}^q \mid \mathbf{X}^q \leq t_q$, $Y = \mathbf{X}^{-q}$ and $g(x, y) = \mathbf{1}_{\max y > t}$. As noted in Section 4.2, drawing realizations of $\mathbf{X}^q \mid \mathbf{X}^q \leq t_q$ has a higher computational cost than simulating $\mathbf{X}^{-q}$ because rejection sampling is required. More generally, if we denote with $C_W(n)$ the cost of n realizations of W and with $C_{Y \mid W}(m; w_i)$ the cost of drawing m conditional simulations from $Y \mid W = w_i$, then sampling several conditional realizations for a given w_i might bring savings if $C_W(1)$ is much higher than $C_{Y \mid W}(1; w_i)$.

In the proposed asymmetric sampling scheme for each realization w_i we sample m realizations $y_{i,1}, \ldots, y_{i,m}$ from $Y \mid W = w_i$. Assume that we sample with this scheme the couples $(w_i, y_{i,j})$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, then we can write the following estimator for G

$$\widetilde{G} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} g(w_i, y_{i,j}).$$

For a fixed number of samples, the estimator $\widetilde{G}$ may have a higher variance than $\widehat{G}$ due to the dependency between pairs sharing the same replicate of W. However, in many cases, it may be more relevant to focus on obtaining good estimates within a fixed time. If we set the computational budget instead of the number of samples and if $C_{Y \mid W}$ is smaller than C_W, then anMC may lead to an overall variance reduction thanks to an increased number of simulated pairs. We show in the remainder of this section that, in the case of an affine cost function $C_{Y \mid W}$, there exists an optimal number of inner simulations m diminishing the variance of $\widetilde{G}$ below that of $\widehat{G}$. Assume

$$C_W(n) = c_0 + cn \text{ and, for each sample } w_i$$
$$C_{Y \mid W}(m; w_i) = C_{Y \mid W}(m) = \alpha + \beta m,$$

with $c_0, c, \alpha, \beta \in \mathbb{R}^+$ dependent on the simulators chosen for W and $Y \mid W$. The second equation entails that the cost of conditional simulations does not depend on the conditioning value.

If $W = \mathbf{X}^q \mid \mathbf{X}^q \leq t_q$, $Y = \mathbf{X}^{-q}$ as in Section 4.2, then $Y \mid W$ is Gaussian with mean and covariance matrix described in Equation (4.6). In this case

the cost for sampling $Y \mid W$ is affine, with α describing pre-calculation times and β random number generation and algebraic operations.

Denote with $W_1, \ldots, W_n$ replications of W. For each W_i we consider the conditional distribution $Y \mid W_i$ and $Y_{1,i}, \ldots, Y_{m,i}$ replications from it. We study here the properties of $\widetilde{G}$ when the total simulation budget, denoted $C_{\text{tot}}(n, m)$ is fixed to $C_{\text{fix}} \in \mathbb{R}^+$. First observe that

$$C_{\text{tot}}(n, m) = c_0 + n(c + \alpha + \beta m).$$

Then we can derive the number of replications of W as a function of m:

$$N_{C_{\text{fix}}}(m) = \frac{C_{\text{fix}} - c_0}{c + \alpha + \beta m}.$$

The following proposition shows a decomposition of $\text{Var}(\widetilde{G})$ that is useful to find the optimal number of simulations m^* under a fixed simulation budget $C_{\text{tot}}(n, m) = C_{\text{fix}}$.

Proposition 5. *Consider n independent copies $W_1, \ldots, W_n$ of W and, for each W_i, m copies $Y_{i,j} = Y_j \mid W_i$ $j = 1, \ldots, m$, independent conditionally on W_i. Then,*

$$\text{Var}(\widetilde{G}) = \frac{1}{n} \text{Var}(g(W_1, Y_{1,1})) - \frac{m-1}{nm} \mathbb{E}\big[\text{Var}(g(W_1, Y_{1,1}) \mid W_1)\big]. \quad (4.9)$$

Corollary 2. *Under the same assumptions, $\widetilde{G}$ has minimal variance when*

$$m = \widetilde{m} = \sqrt{\frac{(\alpha + c)B}{\beta(A - B)}},$$

where $A = \text{Var}(g(W_1, Y_{1,1}))$ and $B = \mathbb{E}\big[\text{Var}(g(W_1, Y_{1,1}) \mid W_1)\big]$. Moreover denote with $\varepsilon = \widetilde{m} - \lfloor \widetilde{m} \rfloor$, then the optimal integer is $m^* = \lfloor \widetilde{m} \rfloor$ if

$$\varepsilon < \frac{(2\widetilde{m} + 1) - \sqrt{4(\widetilde{m})^2 + 1}}{2} \quad (4.10)$$

or $m^* = \lceil \widetilde{m} \rceil$ otherwise.

Proposition 6. *Under the same assumptions, if $m^* > \frac{2(\alpha+c)B}{(c+\alpha)B+\beta(A-B)}$ then $\text{Var}(\widetilde{G}) = \text{Var}(\widehat{G})\,[1 - \eta]$, where $\eta \in (0, 1)$.*

4.3.1 Algorithmic considerations

In order to compute m^*, we need to know $A = \mathrm{Var}(g(W_1, Y_{1,1}))$ and $B = \mathbb{E}\big[\mathrm{Var}(g(W_1, Y_{1,1}) \mid W_1)\big]$ and the constants c_0, c, α and β. A and B depend on the specific problem at hand and are usually not known in advance. Part of the total computational budget is then needed to estimate A and B. This preliminary phase is also used to estimate the system dependent constants c_0, c, α and β. Algorithm 2 reports the pseudo-code for anMC.

4.3.2 Estimate p with $\widehat{p}^{\mathbf{GanMC}}$

The anMC algorithm can be used to reduce the variance compared to the MC estimate proposed in Section 4.2.2. In fact, let us consider $W = \mathbf{X}^q \mid \mathbf{X}^q \le t_q$ and $Y = \mathbf{X}^{-q}$. We have that W is expensive to simulate as it requires rejection sampling while, for a given sample w_i, $Y \mid W = w_i$ is Gaussian with mean and covariance matrix described in Equation (4.6). In many situations, it is much cheaper to obtain samples from $Y \mid W = w_i$ than from W. Moreover, as noted earlier, R_q can be written in the form of Equation (4.7) with $g(x, y) = \mathbf{1}_{\max y > t}$. We can then use Algorithm 2 to calculate m^*, sample n^* realizations $w_1, \ldots, w_{n^*}$ of W and for each realization w_i obtain m^* samples $y_{i,1}, \ldots, y_{i,m^*}$ of $Y \mid W = w_i$. Then we can estimate R_q via

$$\widehat{R_q}^{\mathrm{anMC}} = \frac{1}{n^* m^*} \sum_{i=1}^{n^*} \sum_{j=1}^{m^*} \mathbf{1}_{\max y_{i,j} > t}.$$

Finally plugging in $\widehat{R_q}^{\mathrm{anMC}}$ and $\widehat{p}_q^{\mathrm{G}}$ in Equation (4.4), we obtain

$$\widehat{p}^{\mathrm{GanMC}} = \widehat{p}_q^{\mathrm{G}} + (1 - \widehat{p}_q^{\mathrm{G}})\widehat{R_q}^{\mathrm{anMC}}.$$

Figure 4.3 shows a comparison of results using 30 replications of the experiment presented in Section 4.2.2. Results obtained with a MC estimator are shown for comparison.

While the simulations of all experiments were obtained under the constraint of a fixed computational cost, the actual time to obtain the simulations was not exactly the same. In order to compare the methods in more general settings we further rely on the notion of efficiency. For an estimator $\widehat{p}$, we define the efficiency (Lemieux, 2009, Section 4.2) as

$$\mathrm{Eff}[\widehat{p}] = \frac{1}{\mathrm{Var}(\widehat{p})\,\mathrm{time}[\widehat{p}]},$$

where $\mathrm{time}[\widehat{p}]$ denotes the computational time of the estimator $\widehat{p}$.

Algorithm 2: Asymmetric nested Monte Carlo.

Input   : $\mathbf{m}_W, \mathbf{m}_Y, \Sigma_W, \Sigma_Y, \Sigma_{WY}, g, C_{\text{tot}}$

Output: $\widetilde{G}$

Part 0:   estimate c_0, c, β, α ;

initialize   compute the conditional covariance $\Sigma_{Y|W}$ and initialize n_0, m_0;

Part 1:   **for** $i \leftarrow 1$ **to** n_0 **do**

estimate A, B | simulate w_i from the distribution of W;

| compute the conditional mean $\mathbf{m}_{Y|W=w_i}$;

| draw m_0 simulations $y_{i,1}, \ldots, y_{i,m_0}$ from the conditional
| distribution $Y \mid W = w_i$;

| estimate $\mathbb{E}\left[g(W,Y) \mid W = w_i\right]$ with $\tilde{E}_i = \frac{1}{m_0} \sum_{j=1}^{m_0} g(w_i, y_{i,j})$;

| estimate $\text{Var}\left(g(W,Y) \mid W = w_i\right)$ with
| $\tilde{V}_i = \frac{1}{m_0 - 1} \sum_{j=1}^{m_0} (g(w_i, y_{i,j}) - \tilde{E}_i)^2$;

end

compute $\widetilde{m} = \sqrt{\dfrac{(\alpha+c)\frac{1}{n_0}\sum_{i=1}^{n_0} \tilde{V}_i}{\beta \frac{1}{n_0-1}\sum_{i=1}^{n_0}(\tilde{E}_i - \frac{1}{n_0}\sum_{i=1}^{n_0} \tilde{E}_i)^2}}$, m^* as in Corollary 2 and

$n^* = N_{C_{\text{fix}}}(m^*)$;

Part 2:   **for** $i \leftarrow 1$ **to** n^* **do**

compute $\tilde{G}$ | **if** $i \leq n_0$ **then**

| | **for** $j \leftarrow 1$ **to** m^* **do**

| | | **if** $j \leq m_0$ **then**

| | | | use previously calculated $\widetilde{E}_i$ and $\widetilde{V}_i$;

| | | **else**

| | | | simulate $y_{i,j}$ from the distribution $Y \mid W = w_i$;

| | | | compute $\widetilde{E}_i = \frac{1}{m^*} \sum_{j=1}^{m^*} g(w_i, y_{i,j})$;

| | | **end**

| | **end**

| **else**

| | simulate w_i from the distribution of W ;

| | compute the conditional mean $\mathbf{m}_{Y|W=w_i}$;

| | **for** $j \leftarrow 1$ **to** m^* **do**

| | | simulate $y_{i,j}$ from the conditional distribution $Y \mid W = w_i$;

| | **end**

| | compute $\widetilde{E}_i = \frac{1}{m^*} \sum_{j=1}^{m^*} g(w_i, y_{i,j})$;

| **end**

end

estimate $\mathbb{E}\left[g(W,Y)\right]$ with $\widetilde{G} = \frac{1}{n^*} \sum_{i=1}^{n^*} \tilde{E}_i$;
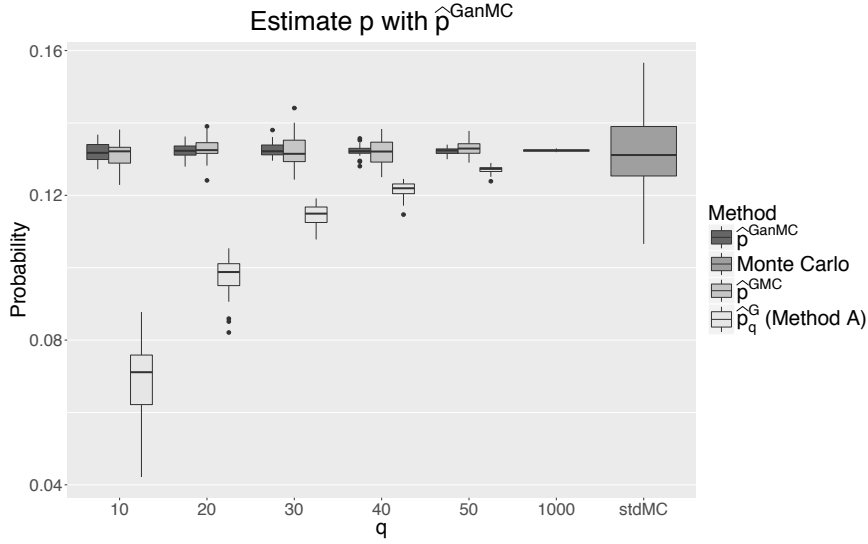
Figure 4.3: Comparison of results with $\widehat{p}_q^{\,G}$, $\widehat{p}^{\,GMC}$ and $\widehat{p}^{\,GanMC}$ on the example introduced in Figure 4.1.
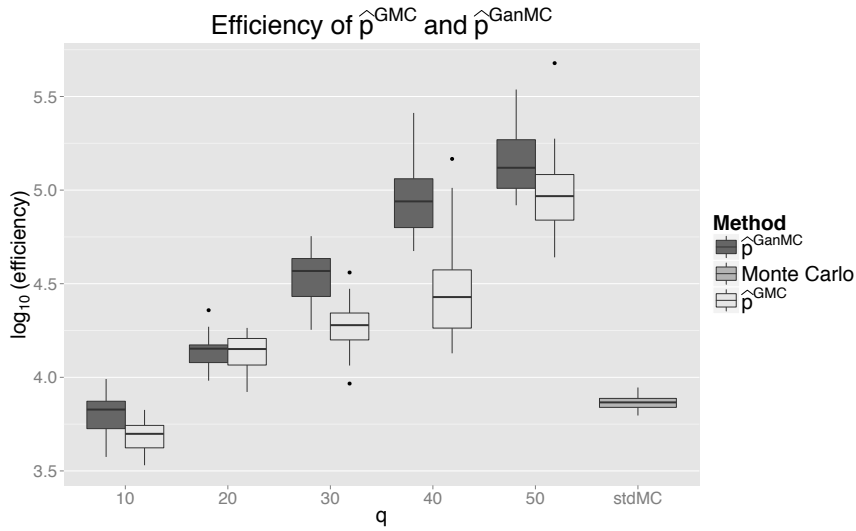
Figure 4.4: Efficiency of the estimators $\widehat{p}^{\,GMC}$ and $\widehat{p}^{\,GanMC}$ compared with the efficiency of a standard MC estimator on 30 replications of the experiment from Figure 4.3. Values in logarithmic scale.

Figure 4.4 shows a comparison of the efficiency of $\widehat{p}^{\mathrm{GMC}}$ and $\widehat{p}^{\mathrm{GanMC}}$ with a full Monte Carlo estimator. With as few as $q = 50$ active dimensions we obtain an increase in efficiency of around 10 times on average over the 30 replications of the experiment with the estimator $\widehat{p}^{\mathrm{GMC}}$. The estimator $\widehat{p}^{\mathrm{GanMC}}$ shows a higher median efficiency than the others for all $q \geq 20$.

4.3.3   Numerical studies

In this section we study the efficiency of the anMC method compared with a standard MC method for different choices of m. Here we do not select the optimal m^* defined in Corollary 2, but we study the efficiency as a function of m. In many practical situations, even if part 1 of Algorithm 2 does not render the optimal m^*, the anMC algorithm is still more efficient than a standard MC if the chosen m is close to m^*.

We consider a similar setup to the experiment presented in Section 4.2.1. Here we start from a GRF with tensor product Matérn ($\nu = 5/2$) and a non constant mean function $\mathfrak{m}$ different from the example in Section 4.2.1. The hyperparameters are fixed as $\theta = [0.5, 0.5, 1, 1, 0.5, 0.5]^T$ and $\sigma^2 = 8$. The GRF is then discretized over the first d points of the Sobol sequence to obtain the vector $\mathbf{X}$. We are interested in $1 - p = P(\mathbf{X} < t)$, with $t = 5$. We proceed by estimating p with $\widehat{p}^{\mathrm{GMC}}$ and $\widehat{p}^{\mathrm{GanMC}}$ for different choices of m. The initial part $\widehat{p}_q^{\mathrm{G}}$ is kept fixed with a chosen q and we compare the efficiency in the estimation of $\widehat{R_q}$. The number of outer simulations in the anMC algorithm is kept fixed to $n = 10,000$ and we only vary m.

Let us first look at this problem in the case $d = 1000$. The median estimated value for p is $\widehat{p} = 0.9644$. Most of the probability is estimated with p_q, in fact $\widehat{p}_q^{\mathrm{G}} = 0.9636$. Figure 4.5a shows $\mathrm{Eff}[\widehat{p}]$ computed with the overall variance of $\widehat{p}$. A choice of $m = 10$ leads to a median increase in efficiency of 73% compared to the MC case. In this example, the probability to be estimated is close to 1. The probability p_q is also close to 1 and thus the acceptance probability for $\widehat{R_q}$ is low. In this situation the anMC method is able to exploit the difference in computational costs to provide a more efficient estimator for $\widehat{R_q}$.

In order to study the effect of the acceptance probability on the method's efficiency we change the threshold in the previous example to $t = 7.5$ by keeping the remaining parameters fixed. The value of p is smaller, $\widehat{p} = 0.1178$. The number of active dimensions q, chosen with Algorithm 1, is smaller ($q = 90$) as the probability mass is smaller. The value of p_q ($\widehat{p}_q^{\mathrm{G}} = 0.1172$) is much smaller than in the previous case and this leads to a higher acceptance probability for $\widehat{R_q}$. Figure 4.5b shows efficiency of the method
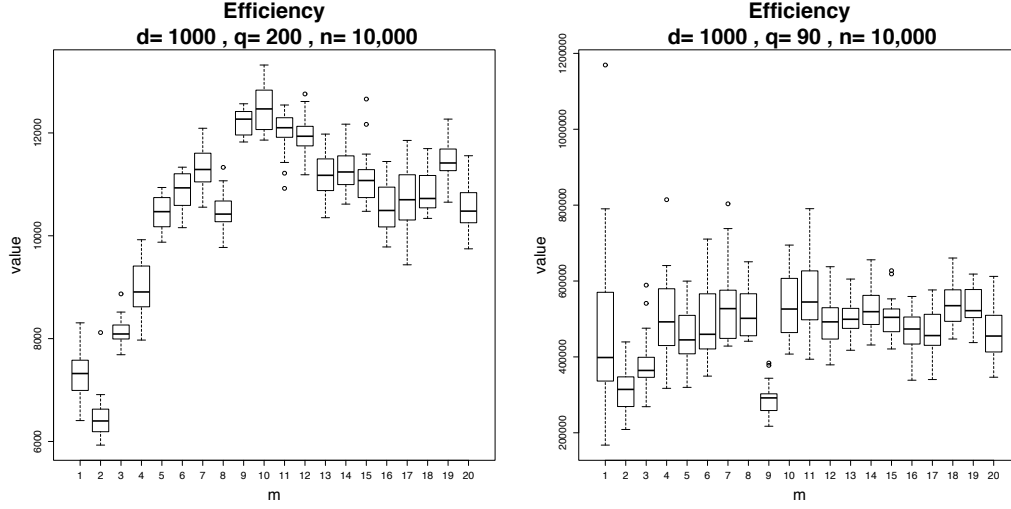
(a) High probability state, $t = 5$, $\widehat{p}^{\text{GanMC}} = 0.9644$.

(b) Low probability state, $t = 7.5$, $\widehat{p}^{\text{GanMC}} = 0.1178$.

Figure 4.5: Efficiency of $\widehat{p}^{\text{GanMC}}$ estimator versus the number of inner simulations m. For each m the experiment is reproduced 30 times.

as a function of m in this case. Here the anMC method does not bring any gain over the MC method as the the ratio between the cost of rejection sampling and the conditional simulations in $\widehat{R_q}$ is close to one. In this case the estimated m^* (1.91) is smaller than the minimum threshold of Proposition 6 that guarantees a more efficient anMC algorithm.

4.3.4   Comparison with state of the art

In this section we compare the GMC and GanMC methods introduced here with available state-of-the-art algorithms to estimate $\pi(t)$. The two methods introduced in the previous section are implemented in the **R** package `ConservativeEstimates` currently available on GitHub. We compare this implementation with:

QRSVN an implementation of Genz method (Genz and Bretz, 2009) in the **R** package `mvtnorm`, function `pmvnorm`;

GHK an implementation of GHK method (Geweke, 1991; Hajivassiliou and McFadden, 1998) in the **R** package `bayesm`, function `ghkvec`;

MET **R** implementation of the minimax-exponentially-tilted (MET) method (Botev, 2017) in the package `TruncatedNormal`, function `mvNcdf`;.
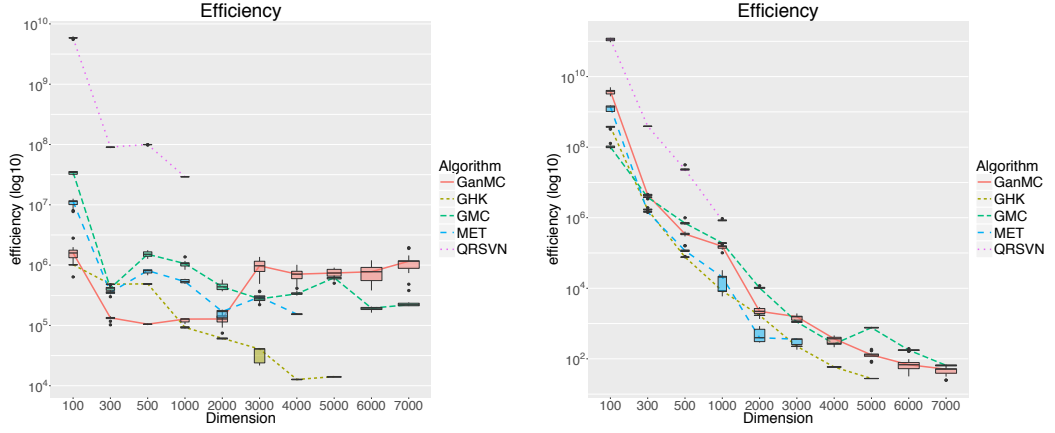
(a) Low $\pi(t)$ state, $t = 5$. The median estimated value for $p = 1 - \pi(t)$ ranges from 0.33845 to 0.99876.

(b) High $\pi(t)$ state, $t = 7.5$. The median estimated value for $p = 1 - \pi(t)$ ranges from 0.00315 to 0.32564.

Figure 4.6: Efficiency of the probability estimator versus the dimension d. For each d the experiment is reproduced 15 times.

We consider the example introduced in the previous section and we increase the dimension of the problem d by considering finer discretizations of the underlying GRF. For example the vector $\mathbf{X}$ of dimension $d = 100$ is obtained from the 6-dimensional GRF discretized on the first 100 points of the 6-dimensional Sobol' sequence. As the dimension d increases the probability $\pi(t)$ changes, thus giving us the possibility of exploring different setups.

Figure 4.6a presents a comparison of the estimator's efficiency for the problem of computing $\pi(t)$, with $t = 5$. This is a low probability setup, as the range of $\pi(t)$ varies between 0.66155 for $d = 100$ and 0.00124 for $d = 7000$. The most efficient algorithm is the QRSVN Genz method, however this implementation does not scale to dimensions higher than 1000. The GMC algorithm is the second most efficient in all dimensions except $d = 2000$ where it is the most efficient. The GanMC algorithm is instead the most efficient when d is greater than 2000. This effect is explained by the efficiency gains brought by $\widehat{R_q}^{\text{anMC}}$ when the rejection sampler is expensive. If $d > 2000$, the probability $P(\mathbf{X}^q \leq t_q)$ is always smaller than 0.01, thus the rejection sampler becomes much more expensive than the conditional sampler in the estimation of the remainder $\widehat{R_q}$. Algorithms GHK and MET allowed estimates until dimension $d = 5000$ and $d = 4000$ respectively before running in memory overflows. The GanMC algorithm is 45 times more efficient than the GHK algorithm for $d = 5000$ and 3.8 times more efficient than MET for $d = 4000$. It is also 5 times more efficient than GMC for $d = 7000$.

Figure 4.6b compares the estimators' efficiency for the computation of $\pi(t)$ with $t = 7.5$. As partially observed in the previous section this is a high probability setup as the median estimate of $\pi(t)$ ranges from 0.99685, for $d = 100$, to 0.67436, for $d = 7000$. Also in this case the QRSVN is the most efficient algorithm in low dimensions. The GMC and the GanMC algorithms however are the most efficient for all dimensions higher than 2000. The GanMC algorithm is 4 times more efficient than the MET for $d = 3000$ and 5 times more efficient than GHK for $d = 5000$. In this setup the computational cost of the rejection sampler in $\widehat{R_q}^{\text{anMC}}$ is not much higher than the conditional sampler. In fact, the acceptance probability of the rejection sampler is always higher than 0.6. This is the main reason why the GMC algorithm proves to be more efficient for most dimensions, in fact for $d = 5000$ it is 5.9 times more efficient than GanMC and for $d = 7000$ the ratio is 1.3. All computations were carried on the cluster of the University of Bern on machines with Intel Xeon CPU 2.40GHz and 16 GB RAM.

4.4 Application: efficient computation of conservative estimates

In this section we show that the GanMC method is a key step for the computation of conservative excursion set estimates relying on GRF models. We consider the conservative setup introduced in Section 4.1 and we remind here that a conservative estimate $\text{CE}_\alpha(\Gamma)$ is a solution of the optimization problem in Equation (4.1), recalled below for convenience.

$$\text{CE}_\alpha(\Gamma) \in \arg \max_{K \in \mathfrak{C}_\rho} \{\mu(K) : P_n(K \subset \Gamma) \geq \alpha\}.$$

As explained in Section 4.1 we select the family of Vorob'ev quantiles Q_ρ as $\mathfrak{C}_\rho$. Algorithm 3 details the steps used to obtain a conservative estimate in this framework.

At each iteration of the while loop in the algorithm we need to compute $P_{\text{next}} = P_n(Q \subset \{Z_x \leq t\})$, the probability that Q is inside the excursion. As shown in Equation (4.2), this quantity can be estimated as a r-dimensional orthant probability, where r is the size of the discretization. The use of $\widehat{p}^{\text{GanMC}}$ to approximate $1 - P_n(Q_{\rho'} \subset \{Z_x \leq t\})$ allows discretizations of the Vorob'ev quantiles at resolutions that seemed out of reach otherwise.

We apply Algorithm 3 to a two dimensional artificial test case. We consider as function f a realization of a Gaussian field $(Z_x)_{x \in D}$, where $D \subset \mathbb{R}^2$ is the unit square. We consider two parametrizations for the prior covariance kernel: a tensor product Matérn covariance kernel with $\nu = 5/2$, variance

---

Algorithm 3: Conservative estimates algorithm.

Input   :

- $\mathfrak{m}_k, \mathfrak{K}_k$, conditional mean and covariance of $Z \mid Z_{\chi_k} = f_k$;

- fine discretization design G;

Output: Conservative estimate for Γ^* at level α.

Part 0:   sort the points in G in decreasing order of $p_{\Gamma,k}$, with indices
$G_s = \{i_1, \ldots i_m\}$;

compute i_B, i_T   find the highest index i_T such that $\prod_{j=1}^{T} p_{\Gamma,k}(G_s)[i_j] \geq \alpha$;

find the highest index i_B such that $p_{\Gamma,k}(G_s)[i_B] \geq \alpha$;

evaluate mean and covariance matrix $\mathfrak{m}_k(i_B)$ and Σ_{i_B,i_B};

Part 1:   initialize $i_L = i_T$, $i_R = i_B$;

Initialize dichotomy   compute $P_L = P_k(Q_{\rho_{i_L}} \subset \{Z_x \leq t\})$, $P_R = P_k(Q_{\rho_{i_R}} \subset \{Z_x \leq t\})$;

Part 2:   **while** $P_R < \alpha$ *and* $(i_R - i_L) \geq 2$ **do**

optimization $\quad$ next evaluation $i_{\text{next}} = \frac{i_L + i_R}{2}$;

$\quad$ compute $P_{\text{next}} = P_k(Q_{\rho_{i_{\text{next}}}} \subset \{Z_x \leq t\})$;

$\quad$ **if** $P_{next} \geq \alpha$ **then**

$\quad\quad$ $i_L = i_{\text{next}}$, $i_R = i_R$;

$\quad$ **else**

$\quad\quad$ $i_L = i_L$, $i_R = i_{\text{next}}$;

$\quad$ **end**

end

$\sigma^2 = 0.5$ and range parameters $\theta = [0.4, 0.2]$ and a Gaussian covariance kernel with variance $\sigma^2 = 0.5$ and range parameters $\theta = [0.2, 0.4]$. In both cases we assume a prior constant mean function. We are interested in the set Γ^* with $t = 1$. For both cases we consider $n = 15$ evaluations of f at the same points chosen by Latin hypercube sampling. Figures 4.7a and 4.7b show the conservative estimate at level 95% compared with the true excursion, the Vorob'ev expectation and the 0.95-quantile for the Matérn and the Gaussian kernel. In both cases we see that the 0.95-quantile does not guarantee that the estimate is included in the true excursion with probability 0.95. The conservative estimates instead are guaranteed to be inside the true excursion with probability $\alpha = 0.95$. They correspond to Vorob'ev quantiles at levels 0.998 and 0.993 for Matérn and Gaussian respectively. The conservative estimates were obtained with a 100×100 discretization of the unit square. Such high resolution grids lead to very high dimensional probability calculations. In fact, the dichotomy algorithm required 11 computations of the probability $1 - P_k(Q_{\rho'} \subset \{Z_x \leq t\})$ for each case. The discretization's size for Q_ρ varied
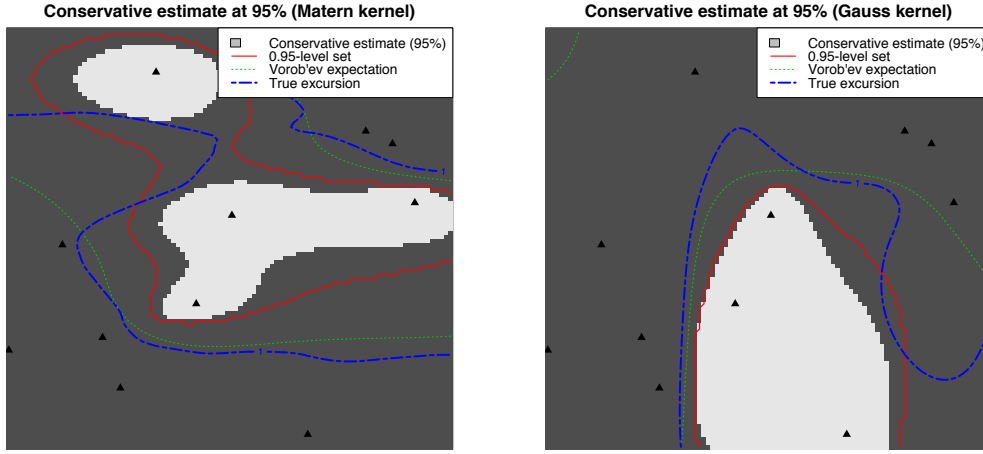
(a) Realization obtained with a Matérn kernel.

(b) Realization obtained with Gaussian kernel.

Figure 4.7: Conservative estimates at 95% (white region) for the excursion below $t = 1$. Both models are based on 15 evaluations of the function (black triangles). The true excursion level is plotted in blue, the Vorob'ev expectation in green and the 0.95-level set in red.

between 1213 and 3201 points in the Matérn kernel case and between 1692 and 2462 points in the Gaussian case. Such high dimensional probabilities cannot be computed with the current implementation of the algorithm by Genz; however, they could also be computed with a standard Monte Carlo at very high computational costs. Instead, with the proposed method, the total computational time on a laptop with Intel Core i7 1.7GHz CPU and 8GB of RAM was equal to 365 and 390 seconds respectively for Matérn and Gaussian kernel.

4.5 Discussion

In this chapter we introduced a new method to approximate high dimensional orthant Gaussian probabilities. The procedure resulted in estimators with greater efficiency than standard Monte Carlo, scalable on dimensions larger than 1000. We proposed two methods to estimate the remainder R_q in the decomposition of Equation (4.4): standard Monte Carlo and asymmetric nested Monte Carlo (anMC). Both methods showed higher efficiency than other state-of-the-art methods for dimensions higher than 1000. The proposed algorithms require a choice of the active dimensions in p_q, the low

dimensional part of the decomposition. In Section 4.2.1 we introduced heuristic algorithm to choose q and the active dimensions. This algorithm provides good results in case of dense covariance matrix with anisotropic diagonal and anisotropic mean vector. An alternative proposal is choosing the first q dimensions ordered according to the inverse Genz variable reordering proposed in Genz and Bretz (2009, Section 4.1.3). This ordering is similar to our proposed approach but it requires a full Cholesky decomposition of the covariance matrix, which could be prohibitive in very high dimensions.

The behaviour of the methods was analysed with numerical studies in Section 4.3. The anMC method proved to be very efficient if the orthant probability of interest has low to medium values. This method however relies on an initial step where several constants and probabilistic quantities are empirically estimated to choose the optimal m^*, the number of inner samples. In particular the cost parameters c, β, the slopes of the linear costs, might be hard to estimate if the constants c_0, α are very large. In this case Part 1 of Algorithm 2 might not choose the optimal m^*. In Section 4.3.3 we studied the behaviour of the anMC algorithm for different choices of m and we noticed that even if the chosen m is not optimal but it is close to optimal, the efficiency gain is very close to the optimal efficiency gain. Another aspect analysed in the numerical part is the behaviour of $\widehat{R_q}$ as a function of q and of the chosen active dimensions. In Section 4.3.3 we observed that the $\widehat{p}^{\text{GanMC}}$ estimator efficiency is mainly driven by the acceptance probability of the rejection sampler in $\widehat{R_q}^{\text{anMC}}$, which depends on $\widehat{p}_q$. This highlights the existence of a trade-off between $\widehat{p}_q^{\text{G}}$ and $\widehat{R_q}$. If the choice of q and active dimensions is not optimal, then the acceptance probability of the rejection sampler becomes larger, making the estimation of $\widehat{R_q}$ easier. By the same reasoning, if $\widehat{p}_q$ is closer to p the quantity $\widehat{R_q}$ becomes harder to estimate. In the second setting, however, the estimator $\widehat{R_q}^{\text{anMC}}$ becomes more efficient compared to $\widehat{R_q}^{\text{MC}}$ as the ratio between the computational costs becomes more favourable.

The efficiency of the overall method depends on the general structure of the Gaussian vector and there are situations where it brings only moderate improvements over a standard Monte Carlo approach. More in-depth studies of the relationships between the covariance structure and the efficiency of the method might be beneficial.

In the application section we showed that the estimator $\widehat{p}^{\text{GanMC}}$ made possible the computation of conservative estimates of excursion sets with general GRF priors. The **R** implementation of the algorithm is contained in the package `ConservativeEstimates` currently available on GitHub.

4.6 Appendix A: proofs

Proof of Proposition 4

Proof. We have that $\mathbb{E}[\widehat{p_q}] = p_q$ and $\mathbb{E}[\widehat{R_q}] = R_q$. Then we have

$$\operatorname{Var}(\widehat{p}) = \operatorname{Var}(\widehat{p_q}) + \underbrace{\operatorname{Var}((1 - \widehat{p_q})\widehat{R_q})}_{=\blacksquare} + 2\underbrace{\operatorname{Cov}(\widehat{p_q}, (1 - \widehat{p_q})\widehat{R_q})}_{=\blacktriangle}. \qquad (4.11)$$

We can write the variance $\blacksquare$ and the covariance $\blacktriangle$ as

$$\blacksquare = \operatorname{Var}((1 - \widehat{p_q})\widehat{R_q}) = (1 - p_q)^2 \operatorname{Var}(\widehat{R_q}) + R_q^2 \operatorname{Var}(\widehat{p_q}) + \operatorname{Var}(\widehat{p_q})\operatorname{Var}(\widehat{R_q}),$$

$$\blacktriangle = \operatorname{Cov}[\widehat{p_q}, (1 - \widehat{p_q})\widehat{R_q}] = -\operatorname{Var}(\widehat{p_q})R_q,$$

respectively, by exploiting the independence of $\widehat{p_q}$ and $\widehat{R_q}$. By plugging in those expressions in Equation (4.11) we obtain the result in Equation (4.5). $\square$

Proof of Proposition 5

Proof.

$$\operatorname{Var}(\widetilde{G}) = \frac{1}{n^2 m^2} \operatorname{Var}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} g(W_i, Y_{i,j})\right) = \frac{1}{nm^2} \operatorname{Var}\left(\sum_{j=1}^{m} g(W_1, Y_{1,j})\right)$$

$$= \frac{1}{nm^2} \sum_{j=1}^{m}\sum_{j'=1}^{m} \operatorname{Cov}\left(g(W_1, Y_{1,j}), g(W_1, Y_{1,j'})\right)$$

$$= \frac{1}{nm^2}\left[m\operatorname{Var}(g(W_1, Y_{1,1})) + m(m-1)\operatorname{Cov}(g(W_1, Y_{1,1}), g(W_1, Y_{1,2}))\right]$$

$$= \frac{1}{nm^2}\left[m\operatorname{Var}(g(W_1, Y_{1,1})) + m(m-1)\blacklozenge\right]. \qquad (4.12)$$

where the first equality is a consequence of the independence of $W_1, \ldots, W_n$ and the third equality is a consequence of the independence of $Y_{i,j}$ and $Y_{i,j'}$ conditionally on W_i. Moreover the covariance denoted by $\blacklozenge$ in Equation (4.12) can be written as follows.

$$\blacklozenge = \underbrace{\mathbb{E}\big[\operatorname{Cov}(g(W_1, Y_{1,1}), g(W_1, Y_{1,2}) \mid W_1)\big]}_{=0 \ Y_{1,1}, Y_{1,2} \text{ independent conditionally on } W_1} + \underbrace{\operatorname{Cov}\left(\mathbb{E}[g(W_1, Y_{1,1}) \mid W_1], \mathbb{E}[g(W_1, Y_{1,2}) \mid W_1]\right)}_{=\operatorname{Var}(\mathbb{E}[g(W_1, Y_{1,1})|W_1])}$$

$$= \operatorname{Var}\left(\mathbb{E}[g(W_1, Y_{1,1}) \mid W_1]\right) = \operatorname{Var}\left(g(W_1, Y_{1,1})\right) - \mathbb{E}\left[\operatorname{Var}\left(g(W_1, Y_{1,1}) \mid W_1\right)\right].$$
$$(4.13)$$

Equations (4.12) and (4.13) give the result in Equation (4.9). $\square$

Proof of Corollary 2

Proof. Denote with $e = \beta(A-B)$, $f = (\alpha+c)(A-B)+\beta B$, $g = (c+\alpha)B$, $h = C_{\text{tot}} - c_0$, then

$$\text{Var}(\widetilde{G})(m) = \frac{em^2 + fm + g}{hm}. \tag{4.14}$$

Observe that the first and second derivatives of $\text{Var}(\widetilde{G})$ with respect to m are respectively

$$\frac{\partial \text{Var}(\widetilde{G})}{\partial m} = \frac{1}{h}\left[e - \frac{g}{m^2}\right], \quad \frac{\partial^2 \text{Var}(\widetilde{G})}{\partial m^2} = \frac{2g}{hm^3}.$$

The second derivative is positive for all $m > 0$ then $\text{Var}(\widetilde{G})$ is a convex function for $m > 0$ and the point of minimum is equal to the zero of $\partial \text{Var}(\widetilde{G})/\partial m$, which is $m = \sqrt{g/e} = \widetilde{m}$.

Since $\text{Var}(\widetilde{G})$ is convex in m, the integer that realizes the minimal variance is either $\lfloor \widetilde{m} \rfloor$ or $\lceil \widetilde{m} \rceil$. By plugging in $m = \widetilde{m} - \varepsilon = \sqrt{g/e} - \varepsilon$ and $m = \widetilde{m} - \varepsilon + 1 = \sqrt{g/e} - \varepsilon + 1$ in Equation (4.14), we obtain the condition in Equation (4.10). $\qquad \square$

Proof of Proposition 6

Proof. First of all notice that the total cost of sampling $\widehat{G}$ is $C_{\text{tot}} = c_0 + n(c + C_{Y|W}) = c_0 + n(c + \alpha + \beta)$. By isolating n in the previous equation we obtain $n = \frac{\overline{C}_{\text{tot}}}{c+\alpha+\beta}$, where $\overline{C}_{\text{tot}} := C_{\text{tot}} - c_0$ for the sake of brevity, and, by computations similar to those in Proposition 5 we obtain

$$\text{Var}(\widehat{G}) = \frac{c + \alpha + \beta}{\overline{C}_{\text{tot}}} \text{Var}(g(W_1, Y_{1,1})) = \frac{c + \alpha + \beta}{\overline{C}_{\text{tot}}} A,$$

where $A = \text{Var}(g(W_1, Y_{1,1}))$. In the following we will also denote $B = \mathbb{E}\left[\text{Var}(g(W_1, Y_{1,1}) \mid W_1)\right]$ as in Corollary 2. Let us now substitute $N_{C_{\text{fix}}}(m^*)$ in Equation (4.9), thus obtaining

$$\begin{aligned}
\text{Var}(\widetilde{G}) &= \frac{(c + \alpha + \beta m^*)Am^* - (m^* - 1)(c + \alpha + \beta m^*)B}{\overline{C}_{\text{tot}}m^*} \\
&= \text{Var}(\widehat{G})\frac{(m^*)^2\beta(A - B) + m^*[(c + \alpha)(A - B) + \beta B] + (c + \alpha)B}{A(c + \alpha + \beta)m^*} \\
&= \text{Var}(\widehat{G})\frac{2(\alpha + c)B + m^*[(c + \alpha)(A - B) + \beta B]}{A(c + \alpha + \beta)m^*}, \tag{4.15}
\end{aligned}$$

where in Equation (4.15) we substituted $(m^*)^2$ from Corollary 2. By further rearranging the terms, we obtain

$$\text{Var}(\widetilde{G}) = \text{Var}(\widehat{G}) \left[1 - \frac{(m^* - 2)(c + \alpha)B + m^*\beta(B - A)}{A(c + \alpha + \beta)m^*} \right] = \text{Var}(\widehat{G}) \left[1 - \eta \right].$$

Since $A - B, B, c, \beta, \alpha$ are always positive, then $\eta < 1$ for all $m^* > 0$. Moreover $\eta > 0$ if

$$m^* > \frac{2(\alpha + c)B}{(\alpha + c)B + \beta(A - B)}.$$

$\square$

Chapter 5

Adaptive design of experiments for conservative excursion set estimation

This chapter is partly based on a working paper, following up a mini-symposium talk at SIAM UQ 2016, co-authored with David Ginsbourger, Clément Chevalier, Julien Bect and Yann Richet. A preprint is available at hal-01379642.

In this chapter we review stepwise uncertainty reduction (SUR) strategies for estimating excursion sets in the context of conservative estimates. See Chapter 2, Section 2.4.4 for a brief introduction to SUR strategies and Chapter 2, Section 2.4.3 for the basic definitions of conservative estimates. Let us consider a continuous function $f : \mathbb{X} \to \mathbb{Y}$, where $\mathbb{X}$ is a locally compact Hausdorff topological space and $\mathbb{Y} = \mathbb{R}^k$, $k \geq 1$. Our objective is to estimate the preimage of the closed subset $T \subset \mathbb{Y}$ under f, i.e. the set

$$\Gamma^\star = \{x \in \mathbb{X} : f(x) \in T\}.$$

In particular in this chapter we focus on the conservative estimate for $\Gamma^\star$ under the random field model introduced in Chapter 2, Section 2.3.3. Conservative estimates, introduced in Section 2.4.3, proved to be useful in reliability engineering (Azzimonti et al., 2015) and other fields (French and Sain, 2013; Bolin and Lindgren, 2015). The methods introduced in Chapter 4 can be used to compute the estimates given a static model where n evaluations of f are available, under a Gaussian random field (GRF) model. Here we focus on the sequential aspect and we study adaptive methods to reduce the uncertainties on these estimates. These methods are first introduced in the general framework where we assume that f is a realization of a GRF $(Z_x)_{x \in \mathbb{X}}$,

with finite second moments and with continuous paths. The formal definition of conservative estimates, their uncertainties and of stepwise uncertainty reduction strategies does not require stronger assumptions. However in Section 5.2.4 we specialize the case where Z is a real valued Gaussian random field defined on a subset of $\mathbb{R}^d, d \geq 1$. In this particular case it is possible to provide closed-form formulas for the sampling criteria.

5.1   Background on SUR strategies

The objective of sequential strategies for random field models is to select a sequence of points $X_1, X_2, \ldots, X_n$ that reduces the uncertainties on the posterior distribution of some quantity of interest. In our case the object of interest are conservative estimates for $\Gamma^\star$ defined as

$$\mathrm{CE}_{\alpha,n} \in \arg \max_{K \in \mathfrak{C}} \{\mu(K) : P_n(K \subset \Gamma) \geq \alpha\}, \tag{5.1}$$

where μ is a finite measure on $\mathbb{X}$ and P_n is the posterior probability given n evaluations. In this case the current state of the model is described by $\mathbf{X}_n \in \mathbb{X}^n$ and $Z_{\mathbf{X}_n} \in \mathbb{Y}^n$, with $\mathbf{X}_n$, the locations where the function f was evaluated and $Z_{\mathbf{X}_n}$ are the actual evaluations. Here we follow Ginsbourger and Le Riche (2010); Bect et al. (2012) and we define a strategy as a sequence of measurable functions $(s_n)_{n \in \mathbb{N}}$, where, for each $n \in \mathbb{N}$, the function $s_n : (\mathbb{X} \times \mathbb{Y})^n \to \mathbb{X}$ associates to the current state of the model $(\mathbf{X}_n, Z_{\mathbf{X}_n})$ a point $x_{n+1} \in \mathbb{X}$ where to evaluate the function f next.

In particular, we focus on Stepwise Uncertainty Reduction (SUR) strategies. A SUR strategy selects the next evaluation in order to reduce a particular uncertainty function. This notion was introduced by Fleuret and Geman (1999) in the field of image vision and it was later revised by Bect et al. (2012) in the framework of Gaussian Process modelling with the objective of estimating the probability of excursion $\mu(\Gamma^\star)$. In Chevalier et al. (2014a) the criteria proposed in Bect et al. (2012) were computed in the batch-sequential case and applied to the problem of identifying the set $\Gamma^\star$. In the remainder of this section we introduce the concepts of uncertainty function and SUR criterion more formally.

Definition 17. *Consider a model where n evaluations of f are given, we call uncertainty of an estimate the map*

$$H_n : (\mathbb{X} \times \mathbb{Y})^n \to \mathbb{R},$$

that associates to each vector of couples $(x_i, Z_{x_i})_{i=1,\ldots,n}$ a real value representing the uncertainty associated with the selected estimate.

If the locations and the evaluations are known then the uncertainty is a value associated with the current state of the model. However assume that $n > 1$ evaluations are known and we are interested in the uncertainty at $n + q$ evaluations. In this case both the locations $X_{n+1}, \ldots, X_{n+q}$ and the values $Z_{X_{n+1}}, \ldots, Z_{X_{n+q}}$ are random. For example, the location X_{n+2} might depend on the value $Z_{X_{n+1}}$. Then the quantity $H_{n+q}(\mathbf{X}_{n+q}, Z_{\mathbf{X}_{n+q}})$ is a random variable. More specifically, denote with $\mathcal{A}_n$ the σ-algebra generated by the couples $X_1, Z_{X_1}, \ldots, X_n, Z_{X_n}$. We denote with $\mathcal{H}_{n+q}$ the $\mathcal{A}_{n+q}$ measurable random variable that returns the uncertainty associated with $\mathcal{A}_{n+q}$, the $\sigma-$algebra generated by $(X_i, Z_{X_i})_{i=1,\ldots,n+q}$. In what follows we denote with $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \mid Z_{\mathbf{X}_n} = \mathbf{f}_n]$ the expectation conditioned on the event $Z_{\mathbf{X}_n} = \mathbf{f}_n$, where $\mathbf{X}_n$ is a fixed design of experiment and $\mathbf{f}_n = f(\mathbf{X}_n) \in \mathbb{Y}^n$.

Example 13 (Uncertainty of the excursion set measure). *In Bect et al. (2012), the object of interest is the probability of failure $\mu(\Gamma)$ defined for the excursion set of a real-valued field $\Gamma = \{x \in \mathbb{X} : Z_x \geq t\}$, with $t \in \mathbb{R}$, and where μ is a finite measure on $\mathbb{X}$. An intuitive way to quantify the uncertainty on this quantity is the quadratic loss $\mathcal{H}_n = \mathbb{E}_n\left[\left(\widehat{\mu(\Gamma)_n} - \mu(\Gamma)\right)^2\right]$ where*

$\widehat{\mu(\Gamma)_n} = \mathbb{E}_n[\mu(\Gamma)]$. In this case, $\widehat{\mu(\Gamma)_n}$ is an unbiased estimator of $\mu(\Gamma)$ and $\mathcal{H}_n$ is equal to $\mathrm{Var}_n(\mu(\Gamma))$.

In general, a SUR strategy aims at selecting locations $X_{n+1}^*, \ldots, X_{n+q}^*$ that reduce the uncertainty on a given quantity by minimizing the expected future uncertainty function given the current evaluations. Note that there are many ways to proceed with this minimization. If we consider a total budget $N > 0$, then an optimal finite horizon strategy is a sequence of random variables $\mathbf{X}_N = (X_1, \ldots, X_N) \in \mathbb{X}^N$ such that the expectation at time zero of the total uncertainty, i.e. $\mathbb{E}[\mathcal{H}_N]$, is minimized. As noted in Mockus (1989) and later in Osborne et al. (2009); Ginsbourger and Le Riche (2010); Bect et al. (2012) this strategy can be formally obtained by backward induction, however it is not practical as for horizons larger than a few steps it suffers from the curse of dimensionality. Several attempts have been proposed recently to relieve this computational issue, see, e.g. Marchant Matus et al. (2014); González et al. (2016) and references therein. Here we focus on sub-optimal strategies, also called *one-step lookahead strategies*, that greedily select the next location as the minimizer of the uncertainty at the next step.

Definition 18 (one-step lookahead criterion). *We call one-step lookahead sampling criterion a function $J_n : \mathbb{X} \to \mathbb{R}$ that associates to each point $x_{n+1} \in \mathbb{X}$ the expected uncertainty at the next step assuming this point is*

added

$$J_n(x_{n+1}) = \mathbb{E}_n[\mathcal{H}_{n+1} \mid X_{n+1} = x_{n+1}],$$

where $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot \mid Z_{\mathbf{X}_n} = \mathbf{f}_n]$ is the posterior expectation of the field after n evaluations of the function.

A generalization of this definition is the batch sequential one-step lookahead criterion where at each iteration we select a batch of q points $\mathbf{x}^{(q)} := (x_{n+1}, \ldots, x_{n+q}) \in \mathbb{X}^q$ that minimizes the future uncertainty. The choice of batch sequential criteria is often justified in practice because it is possible to run the evaluations of the function in parallel thus saving wall clock time. The batch sequential one-step lookahead sampling criterion is the function $J_n : (\mathbb{X})^q \to \mathbb{R}$ that associates to each batch of q points $\mathbf{x}^{(q)} := (x_{n+1}, \ldots, x_{n+q}) \in \mathbb{X}^q$ the expected uncertainty

$$J_n(\mathbf{x}^{(q)}) = \mathbb{E}_n\left[\mathcal{H}_{n+q} \mid X_{n+1} = x_{n+1}, \ldots, X_{n+q} = x_{n+q}\right].$$

Example 14 (SUR criterion for the excursion set measure)**.** *In Bect et al. (2012) the following criterion is proposed to minimize the uncertainty on $\widehat{\mu(\Gamma)}$, the excursion set measure.*

$$J_n(x) = \mathbb{E}_n\left[\left(\widehat{\mu(\Gamma)} - \mu(\Gamma)\right)^2 \mid X_{n+1} = x\right] = \mathrm{Var}_n\left(\mu(\Gamma) \mid X_{n+1} = x\right)$$

In Chevalier et al. (2014a) this sampling criterion is implemented to the batch sequential case. In the same paper fast formulas for evaluating the criteria are proposed under Gaussian process assumptions.

In the next section we revisit these concepts for the problem of computing conservative estimates of an excursion set.

5.2   Contributions

In this section we introduce the proposed sequential strategies for conservative estimates. In the first subsection we review the conservative estimate in a more concrete setting by specifying a family of subsets $\mathfrak{C}$ where the optimization in Equation (5.1) becomes feasible. Moreover we introduce ways to quantify uncertainties on the conservative estimate for a fixed design. In the second subsection we focus on the sequential aspect and we introduce new criteria for uncertainty reduction.

5.2.1 Preliminary notions and results

The conservative estimate introduced in Definition 14, Chapter 2 and recalled in Equation (5.1) above requires the specification of a family $\mathfrak{C}$ where to search for the optimal set CE_α. In practice it is necessary to choose a parametric family indexed by a parameter $\theta \in \mathbb{R}^k$. Consider a nested family $\mathfrak{C}_\theta$ indexed by a real number $\theta \in [0, 1]$, i.e. for each $\theta_1 > \theta_2$

$$C_{\theta_1} \subset C_{\theta_2}, \tag{5.2}$$

for any $C_{\theta_1}, C_{\theta_2} \in \mathfrak{C}_\theta$. Let us define $C_0 = \mathbb{X}$ and assume that $\mu(\mathbb{X}) < \infty$. This is often the case in our framework as $\mathbb{X}$ is either chosen as a compact subset of $\mathbb{R}^d$ with μ the Lebesgue measure or μ is a probability measure.

For each θ, we can define the function $\varphi_\mu : [0, 1] \to [0, +\infty)$ that associates to $\theta \in [0, 1]$ the value $\varphi_\mu(\theta) := \mu(C_\theta)$, with $C_\theta \in \mathfrak{C}_\theta$. It is a non increasing function of θ because the sets C_θ are nested. We further define the function $\psi_\Gamma : [0, 1] \to [0, 1]$ that associates to each θ the probability $\psi_\Gamma(\theta) := P(C_\theta \subset \Gamma)$. The function ψ_Γ is non decreasing in θ due to the nested property in Equation (5.2). In this set-up the computation of CE_α is reduced to finding the smallest $\theta = \theta^\star$ such that $\psi_\Gamma(\theta^\star) \geq \alpha$. The measure $\mu(CE_\alpha)$ is then equal to $\varphi_\mu(\theta^\star)$.

The Vorob'ev quantiles introduced in Equation (2.14), Chapter 2 are a family of nested closed sets that satisfy the property in Equation (5.2). Moreover they have the important property of being minimizers of the expected distance in measure among sets with the same measure.

Proposition 7 below generalizes a well-known result (Molchanov, 2005, Chapter 2) for the Vorob'ev expectation to all quantiles Q_ρ. The Vorob'ev expectation result is also presented in Corollary 3.

Proposition 7. *Consider a measure μ such that $\mu(\mathbb{X}) < \infty$. The Vorob'ev quantile*

$$Q_\rho = \{x \in \mathbb{X} : p_\Gamma(x) \geq \rho\}$$

minimizes the expected distance in measure with Γ among measurable sets M such that $\mu(M) = \mu(Q_\rho)$, i.e.

$$\mathbb{E}\left[\mu(Q_\rho \Delta \Gamma)\right] \leq \mathbb{E}\left[\mu(M \Delta \Gamma)\right],$$

for each measurable set M such that $\mu(M) = \mu(Q_\rho)$.

Proof. Let us consider a measurable set M such that $\mu(M) = \mu(Q_\rho)$. For

each $\omega \in \Omega$, we have

$$\mu(M\Delta\Gamma(\omega)) - \mu(Q_\rho\Delta\Gamma(\omega)) = 2\Big(\mu(\Gamma(\omega) \cap (Q_\rho \setminus M)) - \mu(\Gamma(\omega) \cap (M \setminus Q_\rho))\Big)$$
$$+ \mu(M) - \mu(Q_\rho).$$

By applying the expectation on both sides we obtain

$$\mathbb{E}\Big[\mu(M\Delta\Gamma) - \mu(Q_\rho\Delta\Gamma)\Big] =$$
$$= \mathbb{E}\Big[2\Big(\mu(\Gamma \cap (Q_\rho \setminus M)) - \mu(\Gamma \cap (M \setminus Q_\rho))\Big)\Big] + \mu(M) - \mu(Q_\rho)$$
$$= 2\int_{Q_\rho\setminus M} p_\Gamma(u)d\mu(u) - 2\int_{M\setminus Q_\rho} p_\Gamma(u)d\mu(u) + \mu(M) - \mu(Q_\rho),$$

where the second equality is obtained by interchanging the expectation with the integral by Tonelli's theorem. Moreover, since $p_\Gamma(x) \geq \rho$ for $x \in Q_\rho \setminus M$ and $p_\Gamma(x) \leq \rho$ for $x \in M \setminus Q_\rho$ we have

$$2\Big[\int_{Q_\rho\setminus M} p_\Gamma(u)d\mu(u) - \int_{M\setminus Q_\rho} p_\Gamma(u)d\mu(u)\Big] + \mu(M) - \mu(Q_\rho)$$
$$\geq 2\rho[\mu(Q_\rho \setminus M) - \mu(M \setminus Q_\rho)] + \mu(M) - \mu(Q_\rho)$$
$$= (2\rho - 1)[\mu(Q_\rho) - \mu(M)] = 0, \tag{5.3}$$

where in the last equality we exploited $\mu(Q_\rho) = \mu(M)$. $\qquad\square$

The Vorob'ev expectation, as defined in Definition 7, Chapter 2, is one particular quantile and it enjoys the same property.

Corollary 3. *The Vorob'ev expectation Q_{ρ_V} minimizes the expected distance in measure with Γ among all measurable (deterministic) sets M such that $\mu(M) = \mu(Q_{\rho_V})$. Moreover if $\rho_V \geq \frac{1}{2}$, then the Vorob'ev expectation also minimizes the expected distance in measure with Γ among all measurable sets M that satisfy $\mu(M) = \mathbb{E}[\mu(\Gamma)]$.*

Proof. The first statement is a direct consequence of Proposition 7 for $\rho = \rho_V$.

For the second statement, by definition we have that either $\mu(Q_{\rho_V}) = \mathbb{E}[\mu(\Gamma)]$ or $\mu(Q_\rho) < \mathbb{E}[\mu(\Gamma)] \leq \mu(Q_{\rho_V})$ for each $\rho > \rho_V$. In the first case we can directly apply Proposition 7. In the second case we can apply the same reasoning as in the proof of Proposition 7 however in Equation (5.3) we need to impose $\rho_V \geq \frac{1}{2}$ for obtaining the result.

$\qquad\square$

Example 15 (Counterexample to Corollary 3). *If $\mathbb{E}[\mu(\Gamma)] < \mu(Q_{\rho_V})$ and $\rho_V < \frac{1}{2}$ then Q_{ρ_V} is no longer the minimizer of the expected distance in measure with Γ among all measurable deterministic sets M such that $\mu(M) = \mathbb{E}[\mu(\Gamma)]$.*

Let Γ take value F with probability α or the empty set otherwise. Then $p_\Gamma(u) = \alpha\mathbf{1}_F(u)$ and $Q_{\rho_V} = F$ with $\rho_V = \alpha$, while $\mathbb{E}[\mu(\Gamma)] = \alpha\mu(F)$. Furthermore, $\mathbb{E}[\mu(\Gamma\Delta F)] = (1 - \alpha)\mu(F)$, while

$$\mathbb{E}[\mu(\Gamma\Delta M)] = \alpha\mu(F \setminus M) + (1 - \alpha)\mu(M)$$
$$= 2\alpha(1 - \alpha)\mu(F)$$

for $M \subset F$. Thus, the inequality does not hold for $\alpha < \frac{1}{2}$.

The Vorob'ev quantiles are thus an optimal family for conservative estimates with respect to the expected distance in measure. In general, however, the Vorob'ev quantile chosen for CE_α with this procedure is not the set S with the largest measure μ that has the property $P(S \subset \Gamma) \geq \alpha$ as shown in the counterexample below.

Example 16. *Consider a discrete set $D = \{x_1, x_2, x_3, x_4\}$, a random field $(Z_x)_{x\in D}$ and assume that $\Gamma = \{x \in D : Z_x \geq 0\}$. In this framework we show the existence of a conservative set at level $\alpha = 0.5$ larger than the largest Vorob'ev quantile with the same conservative property.*

Assume that for some $\rho_1 \in [0, 1]$

$$P(Q_{\rho_1} \subset \Gamma) = P(Z_{x_1} \geq 0, Z_{x_2} \geq 0) = 1/2,$$

where $Q_{\rho_1} = \{x_1, x_2\}$ is a Vorob'ev quantile at level ρ_1, that is $P(Z_{x_1} \geq 0), P(Z_{x_2} \geq 0) \geq \rho_1$.

Note that in the case where $Z_{x_1} \perp\!\!\!\perp Z_{x_2}$, the Vorob'ev level is automatically determined as $\rho_1 = \sqrt{2}/2$ and if $Z_{x_1} = Z_{x_2}$ a.s., then $\rho_1 = 1/2$. Let us assume here that $Z_{x_1} \neq Z_{x_2}$, which implies $\rho_1 \in (1/2, \sqrt{2}/2)$. Let us further denote with Ω_1 the subset of Ω such that for all $\omega \in \Omega_1$ $\min(Z_{x_1}(\omega), Z_{x_2}(\omega)) \geq 0$ and define $\Omega_2 = \Omega \setminus \Omega_1$.

We further fix the random variable Z_{x_3} as

$$Z_{x_3}(\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega_1 \\ -1 & \text{if } \omega \in \Omega_2. \end{cases}$$

Then $P(Z_{x_3} \geq 0) = P(\min(Z_{x_1}, Z_{x_2}) \geq 0) = P(Z_{x_1} \geq 0, Z_{x_2} \geq 0) = 1/2$. Moreover $P(\min(Z_{x_1}, Z_{x_2}, Z_{x_3}) \geq 0) = 1/2$, i.e. $\{x_1, x_2, x_3\}$ has the conservative property at level $\alpha = 0.5$.

Consider $\Omega_3 \subset \Omega_1$ with $P(\Omega_3) = 1/3$ and $\Omega_4 \subset \Omega_2$ with $P(\Omega_4) = 1/3$. Define

$$Z_{x_4}(\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega_3 \cup \Omega_4 \\ -1 & \text{otherwise.} \end{cases}$$

We now have that $P(\min(Z_{x_1}, Z_{x_2}, Z_{x_4}) \geq 0) = P(\min(Z_{x_1}, Z_{x_2}, Z_{x_3}, Z_{x_4}) \geq 0) = P(\Omega_3) = 1/3 < 1/2$ and $P(Z_{x_4} \geq 0) = 1/3 + 1/3 > 1/2$. Under this construction, assuming w.l.o.g. that $\rho_1 > 2/3$, the Vorob'ev quantiles are $Q_{\rho_1} = \{x_1, x_2\}, Q_{\rho_2} = \{x_1, x_2, x_4\}$, with $\rho_2 = 2/3$, and $Q_{0.5} = D$. The set $\{x_1, x_2, x_3\}$ is the largest subset of D with the conservative property therefore it is the unique conservative set at level $\alpha = 0.5$. However $\{x_1, x_2, x_3\}$ is not a Vorob'ev quantile.

In the remainder of this chapter we always consider $\mathfrak{C}$ as the family of Vorob'ev quantiles. Given an initial design $\mathbf{X}_n$ we can proceed as in Section 2.3.2 and compute the Vorob'ev quantiles $Q_{n,\rho}$. By exploiting the previously described optimization procedure we can obtain a conservative estimate at level α for $\Gamma^\star$ denoted here with $\mathrm{CE}_{\alpha,n}$. In the next subsection we introduce different ways to quantify the uncertainties on this estimate, while in Section 5.2.3 we introduce sequential strategies to reduce this uncertainty by adding new evaluations to the model.

5.2.2   Uncertainty quantification on $\mathrm{CE}_{\alpha,n}$

In this section we introduce several uncertainty functions for conservative estimates. Here we always consider a static scenario where n evaluations of f are available.

Our object of interest is the set of excursion, therefore we require uncertainty functions that take into account the whole set. In Chevalier et al. (2013); Chevalier (2013) the uncertainty on the Vorob'ev expectation is evaluated with the expected distance in measure between the current estimate Q_{n,ρ_n} and the set Γ. The following definition reviews this concept.

Definition 19 (Vorob'ev uncertainty). *Consider a Vorob'ev quantile Q_{ρ_n} at level ρ_n, computed after n evaluations of the objective function f. We call Vorob'ev uncertainty of the quantile Q_{ρ_n} the quantity*

$$\mathcal{H}_n(\rho_n) = \mathbb{E}_n[\mu(\Gamma \Delta Q_{n,\rho_n})]. \tag{5.4}$$

In what follows this uncertainty quantification function is applied to the Vorob'ev expectation, $\rho_n = \rho_{V,n}$, to the Vorob'ev median, $\rho_n = 0.5$, and to the conservative estimate at level α, $\rho_n = \rho_n^\alpha$.

Consider now the case of conservative estimates and fix the level α. Note that, by definition, the symmetric difference can be written as

$$\mathbb{E}_n[\mu(\Gamma \Delta Q_{n,\rho_n^\alpha})] = \mathbb{E}_n[\mu(Q_{n,\rho_n^\alpha} \setminus \Gamma)] + \mathbb{E}_n[\mu(\Gamma \setminus Q_{n,\rho_n^\alpha})]. \qquad (5.5)$$

Let us denote with $G_n^{(1)} = \mu(Q_{n,\rho_n^\alpha} \setminus \Gamma)$ the random variable associated with the measure of the first set difference and with $G_n^{(2)} = \mu(\Gamma \setminus Q_{n,\rho_n^\alpha})$ the random variable associated with the second one. Notice that for all $\omega \in \Omega$ such that $Q_{n,\rho_n^\alpha} \subset \Gamma(\omega)$, we have $G_n^{(1)}(\omega) = 0$.

Remark 3. *Consider the conservative estimate Q_{n,ρ_n^α}, then the ratio between the error $\mathbb{E}_n[G_n^{(1)}]$ and the measure $\mu(Q_{n,\rho_n^\alpha})$ is bounded by $1 - \alpha$, the chosen level for the conservative estimates.*

Proof. By applying the law of total expectation we obtain the following.

$$\begin{aligned}
\mathbb{E}_n[G_n^{(1)}] &= \mathbb{E}_n[G_n^{(1)} \mid Q_{n,\rho_n^\alpha} \subset \Gamma]P(Q_{n,\rho_n^\alpha} \subset \Gamma) \\
&+ \mathbb{E}_n[G_n^{(1)} \mid Q_{n,\rho_n^\alpha} \setminus \Gamma \neq \emptyset](1 - P(Q_{n,\rho_n^\alpha} \subset \Gamma)) \\
&\leq 0 + \mathbb{E}_n[G_n^{(1)} \mid Q_{n,\rho_n^\alpha} \setminus \Gamma \neq \emptyset](1 - \alpha) \leq \mu(Q_{n,\rho_n^\alpha})(1 - \alpha).
\end{aligned}$$

$\square$

In the conservative estimate case the two parts in Equation 5.5 do not contribute equally, therefore we can define a weighted Vorob'ev uncertainty as follows.

Definition 20 (Weighted Vorob'ev uncertainty)**.** *Consider the Vorob'ev quantile Q_{n,ρ_n^α} corresponding to the conservative estimate at level α for Γ and fix $\beta \geq 0$. The* weighted Vorb'ev uncertainty *is the function $\mathcal{H}_n^{\mathrm{W}}$ defined as*

$$\begin{aligned}
\mathcal{H}_n^{\mathrm{W}}(\rho_n^\alpha; \beta) &:= \mathbb{E}_n[\beta G_n^{(1)} + G_n^{(2)}] \\
&= \mathbb{E}_n[\beta \mu(Q_{n,\rho_n^\alpha} \setminus \Gamma) + \mu(\Gamma \setminus Q_{n,\rho_n^\alpha})] \qquad (5.6)
\end{aligned}$$

A conservative estimate is built to control the error $\mathbb{E}_n[G_n^{(1)}]$. By making a broad use of the hypothesis testing lexicon, we denote with *Type I* error at state n the quantity $\mathbb{E}_n[G_n^{(1)}]$ and with *Type II* error at state n the quantity $\mathbb{E}_n[G_n^{(2)}]$. This error defines the following uncertainty function for $\mathrm{CE}_{\alpha,n}$.

Definition 21 (Type II uncertainty). *Consider the Vorob'ev quantile Q_{n,ρ_n^α} corresponding to the conservative estimate at level α for Γ. The* Type II uncertainty *is the uncertainty function $\mathcal{H}_n^{\mathrm{T2}}$ defined as*

$$\mathcal{H}_n^{\mathrm{T2}}(\rho_n^\alpha) := \mathbb{E}_n[G_n^{(2)}] = \mathbb{E}_n[\mu(\Gamma \setminus Q_{n,\rho_n^\alpha})] \tag{5.7}$$

Conservative estimates at high levels α tend to select regions inside Γ, by definition. In particular if the number of function evaluations is high enough to have a good approximation of the function f, the conservative estimates with high α tend to be inside the true excursion set $\Gamma^\star$. In these situations the expected type I error is bounded by α, as shown in Remark 3, and thus usually very small, while type II error could be rather large. Type II uncertainty is thus a relevant quantity when evaluating conservative estimates. In the test case studies we also compute the expected type I error to check that it is consistently small.

The last uncertainty function introduced in this section is the measure of Q_{n,ρ_n^α} multiplied by -1. In fact by definition the conservative estimate has probability at least α of being inside the excursion. A large measure for Q_{n,ρ_n^α} is then a good indicator that most of the target set is covered by the conservative estimate.

Definition 22 (Uncertainty MEAS). *We denote the uncertainty function related to the measure μ with $\mathcal{H}_n^{\mathrm{MEAS}}$, defined as*

$$\mathcal{H}_n^{\mathrm{MEAS}}(\rho_n^\alpha) := -\mathbb{E}_n[\mu(Q_{n,\rho_n^\alpha})] \tag{5.8}$$

If $\mu(\mathbb{X}) < \infty$, then $\mathcal{H}_n^{\mathrm{MEAS}}(\rho_n^\alpha)$ is bounded from below by $-\mu(\mathbb{X})$. While the uncertainty functions are usually bounded from below by zero, this property is not required to develop SUR strategies, see, e.g. Bect et al. (2016).

5.2.3   Sequential strategies for conservative estimates

In this section we consider a current design of experiments $\mathbf{X}_n$, for some $n > 0$. We introduce one-step lookahead SUR strategies that select the next batch of $q > 0$ locations $X_{n+1}, \ldots, X_{n+q} \in \mathbb{X}$ in order to reduce the future uncertainty $\mathcal{H}_{n+q}$. In particular, we introduce one-step lookahead strategies associated with the uncertainties defined in Equation (5.4), (5.6), (5.7) and (5.8).

The first is an adaptation of the Vorob'ev criterion introduced in Chevalier (2013) and based on the Vorob'ev deviation (Vorob'ev, 1984; Molchanov, 2005; Chevalier et al., 2013). We define it as follows.

$$J_n(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha) = \mathbb{E}_n \left[\mathcal{H}_{n+q}(\rho_{n+q}^\alpha) \mid X_{n+1} = x_{n+1}, \dots, X_{n+q} = x_{n+q} \right]$$
$$= \mathbb{E}_n \left[\mathbb{E}_{n+q} \left[\mu(\Gamma \Delta Q_{n+q, \rho_{n+q}^\alpha}) \mid X_{n+1} = x_{n+1}, \dots, X_{n+q} = x_{n+q} \right] \right],$$

for $\mathbf{x}^{(q)} = (x_{n+1}, \dots, x_{n+q}) \in \mathbb{X}^q$, where $Q_{n+q, \rho_{n+q}^\alpha}$ is the Vorob'ev quantile obtained with $n + q$ evaluations of the function at level ρ_{n+q}^α, the conservative level obtained with $n + q$ evaluations. In what follows for the sake of brevity we denote with $\mathcal{A}_{n+q}(\mathbf{x}^{(q)})$ the σ-algebra generated by the couples $(x_{n+1}, Z_{x_{n+1}}), \dots, (x_{n+q}, Z_{x_{n+q}})$, where $(x_{n+1}, \dots, x_{n+q}) = \mathbf{x}^{(q)}$.

An approach leading to different criteria involves splitting the measure of the symmetric difference into two parts as in Equation (5.5). In the case of conservative estimates with high level α, each term of Equation (5.5) does not contribute equally to the expected distance in measure, as observed in Remark 3. It is thus reasonable to consider the following asymmetric sampling criterion

$$J_n^{\mathrm{W}}(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha, \beta) = \mathbb{E}_n \left[\mathcal{H}_{n+q}^W(\rho_{n+q}^\alpha) \mid X_{n+1} = x_{n+1}, \dots, X_{n+q} = x_{n+q} \right]$$
$$= \mathbb{E}_n \left[\mathbb{E} \left[\beta G_n^{(1)}(Q_{n+q, \rho_{n+q}^\alpha}) + G_n^{(2)}(Q_{n+q, \rho_{n+q}^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right],$$

for $\mathbf{x}^{(q)} \in \mathbb{X}^q$, where $\beta \geq 0$ is a penalization constant. Of particular interest is the criterion $J_n^{\mathrm{T2}}(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha) := J_n(\mathbf{x}^{(q)}; \rho_{n+q}^\alpha, \beta = 0)$ where the criterion selects the next evaluation at a location that minimizes the expected type II error. In Section 5.2.4 this criterion is derived for a generic level $\rho_n \in [0, 1]$, under more restrictive assumptions on the field and on $\Gamma^\star$.

The last criterion studied here for conservative estimates relies on the uncertainty function $\mathcal{H}_n^{\mathrm{MEAS}}$. We can define the measure based criterion as

$$J_n^{\mathrm{MEAS}}(\mathbf{x}^{(q)}; \rho_n^\alpha) = \mathbb{E}_n \left[-\mathbb{E} \left[\mu(Q_{n+q, \rho_{n+q}^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right].$$

Since we are interested in minimizing this criterion we consider the equivalent function to maximize

$$\widetilde{J_n}^{\mathrm{MEAS}}(\mathbf{x}^{(q)}; \rho_n^\alpha) = \mathbb{E}_n \left[\mathbb{E} \left[\mu(Q_{n+q, \rho_{n+q}^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right].$$

Note that this criterion selects points that are meant to increase the measure of the estimate and it is only reasonable for conservative estimates where the conservative condition on $Q_{\rho_n^\alpha}$ leads to sets with finite measure in expectation.

In practice, there is no closed form formula for the criteria presented above with ρ_{n+q}^α. In the implementation part we replace this level with ρ_n^α, the last level computed.

5.2.4 Implementation

In this section we detail the algorithmic aspects of the criteria.

The notions of an estimate uncertainty and of sequential criteria can be defined in the generic setting introduced in Section 5.1. However, in order to provide formulae to implement the criteria we need to restrict ourselves to a more specific framework. In this section we fix $\mathbb{X} \subset \mathbb{R}^d$, a compact subset of $\mathbb{R}^d$, and $\mathbb{Y} = \mathbb{R}$. These choices are common especially in engineering and other scientific applications. We assume that Z is a Gaussian field with constant prior mean $\mathfrak{m}$ and covariance kernel $\mathfrak{K}$. While this assumption might seem restrictive at first, the choice of an appropriate covariance kernel $\mathfrak{K}$ allows us to define prior distributions over a very large space of functions. Moreover this choice gives us closed form formulae for the posterior mean and covariance of the field thus making the implementation of the criteria less computationally demanding. Note that such formulae could be extended to the universal kriging (UK) case with a Gaussian improper prior on the trend parameters. Here we restrict ourselves to the constant mean setting for brevity. Finally we derive the previously introduced sampling criteria in closed-form for the set $\Gamma^\star = \{x \in \mathbb{X} : f(x) \in T\}$ with $T = [t, +\infty)$, where $t \in \mathbb{R}$ is a fixed threshold. It is straightforward to compute the criteria for $T = (-\infty, t]$ and also to extend them for intersections of such intervals.

The formulas for the criteria introduced here all rely on the posterior coverage probability function $p_{n,\Gamma}$ which, in what follows, will be denoted as p_n if there is no ambiguity on the set Γ. Recall that, under the previous assumptions, for each $n \geq 0$, $p_n(x) = \Phi\left((\mathfrak{m}_n(x) - t)/s_n(x)\right)$, with $x \in \mathbb{X}$, where Φ is the c.d.f. of the univariate standard Normal distribution, $\mathfrak{m}_n$ is the posterior mean of the field and $s_n(x) = \sqrt{\mathfrak{K}_n(x,x)}$ for all $x \in \mathbb{X}$, is the posterior standard deviation. Let us now fix the vector of new evaluations $\mathbf{x}^{(q)} \in \mathbb{X}^q$. Assuming that $s_{n+q}(x) > 0$ for each $x \in \mathbb{X}$, the coverage function $p_{n+q}(\mathbf{x}^{(q)})$ can be written as

$$p_{n+q}(\mathbf{x}^{(q)})(x) = \Phi\left(a_{n+q}(x) + \mathbf{b}_{n+q}^T Y_q\right), \tag{5.9}$$

where

$$a_{n+q}(x) = \frac{\mathfrak{m}_n(x) - t}{s_{n+q}(x)},$$

$$\mathbf{b}_{n+q}(x) = \frac{\mathfrak{K}_n(x, \mathbf{x}^{(q)})K_q^{-1}}{s_{n+q}(x)},$$

$$\mathfrak{K}_n(x, \mathbf{x}^{(q)}) = (\mathfrak{K}(x, x_{n+1}), \ldots, \mathfrak{K}(x, x_{n+q})),$$

K_q is covariance matrix with elements $[\mathfrak{K}_n(x_{n+i}, x_{n+j})]_{i,j=1,\ldots,q}$ and $Y_q \sim N_q(0, K_q)$ is a q dimensional normal random vector that represents the un-

known centred response. In practice here we exploit the kriging update formulas (Chevalier et al., 2014b; Emery, 2009) for faster updates of the posterior mean and covariance when new evaluations are added.

The first criterion introduced in Section 5.2.3 is based on the symmetric difference between the set Γ and the Vorob'ev quantile $Q_{n,\rho}$. In Chevalier (2013), Chapter 4.2, the formula for this criterion in this framework was derived for the Vorob'ev expectation, i.e. the quantile at level $\rho = \rho_V$. In the following remark we first extend this result to any quantile ρ.

Remark 4 (Criterion J_n). *Under the previously introduced assumptions the criterion J_n can be expanded in closed form as*

$$J_n(\mathbf{x}^{(q)}; \rho_n) = \mathbb{E}_n \left[\mathbb{E} \left[\mu \left(\Gamma \Delta Q_{n_q, \rho_n} \right) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right]$$

$$= \int_{\mathbb{X}} \left(2\Phi_2 \left(\begin{pmatrix} a_{n+q}(u) \\ \Phi^{-1}(\rho_n) - a_{n+q}(u) \end{pmatrix}; \begin{pmatrix} 1 + \gamma_{n+q}(u) & -\gamma_{n+q}(u) \\ -\gamma_{n+q}(u) & \gamma_{n+q}(u) \end{pmatrix} \right) \right.$$

$$\left. - p_n(u) + \Phi \left(\frac{a_{n+q}(u) - \Phi^{-1}(\rho_n)}{\gamma_{n+q}(u)} \right) \right) d\mu(u), \tag{5.10}$$

where $\Phi_2(\cdot; \Sigma)$ is the bivariate centred Normal distribution with covariance matrix Σ and $\Phi^{-1}(u)$ denotes the quantile at level u of the standard Normal distribution and

$$\gamma_{n+q}(u) = \sqrt{\mathbf{b}_{n+q}^T(u) K_q \mathbf{b}_{n+q}(u)}.$$

The proof of the previous remark is a simple adaptation of the proof in Chevalier (2013), Chapter 4.2.

Proposition 8. *The criterion $J_n^{\mathrm{W}}(\cdot; \rho_n^\alpha, \beta)$ can be expanded in closed form as*

$$J_n^{\mathrm{W}}(\mathbf{x}^{(q)}; \rho_n^\alpha, \beta) = \mathbb{E}_n \left[\mathbb{E} \left[\beta G_n^{(1)}(Q_{n+q, \rho_n^\alpha}) + G_n^{(2)}(Q_{n+q, \rho_n^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right]$$

$$= \int_{\mathbb{X}} \left((\beta + 1) \Phi_2 \left(\begin{pmatrix} a_{n+q}(u) \\ \Phi^{-1}(\rho_n^\alpha) - a_{n+q}(u) \end{pmatrix}; \begin{pmatrix} 1 + \gamma_{n+q}(u) & -\gamma_{n+q}(u) \\ -\gamma_{n+q}(u) & \gamma_{n+q}(u) \end{pmatrix} \right) \right.$$

$$\left. - \beta p_n(u) + \beta \Phi \left(\frac{a_{n+q}(u) - \Phi^{-1}(\rho_n^\alpha)}{\gamma_{n+q}(u)} \right) \right) d\mu(u). \tag{5.11}$$

Proof. The proof is a simple adaptation of the Remark 4. See, e.g., Chevalier (2013). $\square$

By taking $J_n^{\mathrm{W}}(\mathbf{x}^{(q)}; \rho_n^\alpha, \beta)$ with $\beta = 0$ we obtain a closed form for the Type II uncertainty criterion $J_n^{\mathrm{T2}}(\mathbf{x}^{(q)}; \rho_n^\alpha)$ as

$$J_n^{\mathrm{T2}}(\mathbf{x}^{(q)}; \rho_n^\alpha) = \mathbb{E}_n \left[\mathbb{E} \left[G_n^{(2)}(Q_{n+q, \rho_n^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right] \tag{5.12}$$

$$= \int_{\mathbb{X}} \Phi_2 \left(\begin{pmatrix} a_{n+q}(u) \\ \Phi^{-1}(\rho_n^\alpha) - a_{n+q}(u) \end{pmatrix}; \begin{pmatrix} 1 + \gamma_{n+q}(u) & -\gamma_{n+q}(u) \\ -\gamma_{n+q}(u) & \gamma_{n+q}(u) \end{pmatrix} \right) d\mu(u).$$

Proposition 9. *The criterion J_n^{MEAS} can be expanded in closed form as*

$$J_n^{\mathrm{MEAS}}(\mathbf{x}^{(q)}; \rho_n^\alpha) = \mathbb{E}_n \left[\mathbb{E} \left[\mu(Q_{n+q,\rho_n^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right]$$

$$= \int_{\mathbb{X}} \Phi \left(\frac{a_{n+q}(u) - \Phi^{-1}(\rho_n^\alpha)}{\gamma_{n+q}(u)} \right) d\mu(u). \qquad (5.13)$$

Proof. The indicator function of the set Q_{n+q,ρ_n^α} can be written as $\mathbb{1}_{p_{n+q}(x) \geq \rho_n^\alpha}$. By Tonelli's theorem we exchange the expectation with the integral over $\mathbb{X}$ and we obtain

$$\mathbb{E}_n \left[\mathbb{E} \left[\mu(Q_{n+q,\rho_n^\alpha}) \mid \mathcal{A}_{n+q}(\mathbf{x}^{(q)}) \right] \right] = \int_{\mathbb{X}} \mathbb{E}_n \left[\mathbb{1}_{p_{n+q}(u) \geq \rho_n^\alpha} \right] d\mu(u)$$

$$= \int_{\mathbb{X}} P_n \left(p_{n+q}(u) \geq \rho_n^\alpha \right) d\mu(u).$$

By substituting the expression in Equation (5.9) we obtain

$$\int_{\mathbb{X}} P_n \left(p_{n+q}(u) \geq \rho_n^\alpha \right) d\mu(u) = \int_{\mathbb{X}} P_n \left(a_{n+q}(u) + \mathbf{b}_{n+q}^T(u) Y_q \geq \Phi^{-1}(\rho_n^\alpha) \right) d\mu(u)$$

$$= \int_{\mathbb{X}} \Phi \left(\frac{a_{n+q}(u) - \Phi^{-1}(\rho_n^\alpha)}{\gamma_{n+q}(u)} \right) d\mu(u)$$

$$\square$$

The sampling criteria, implemented above in Equation (5.10), (5.11), (5.12) and (5.13), are used to select the next evaluations of the function f as a part of a larger algorithm that provides conservative estimates for $\Gamma^\star$. In practice the estimation procedure often starts with a small initial design $\mathbf{X}_n$, $n \geq 1$, where n is often chosen as a function of the input space dimension. In $\mathbb{X} = \mathbb{R}^d$, as a rule of thumbs, the initial number of evaluations is often $n = 10d$. In our framework, often, the initial design is chosen as space filling, such as a Latin hypercube sample (LHS) design or points from a low discrepancy sequence such as the Halton and the Sobol' sequence. In Algorithm 4 we summarize the main steps for computing conservative estimates and evaluating their uncertainties. The conservative level ρ_n^α is computed

with Algorithm 3 detailed in Chapter 4.

Algorithm 4: Sequential conservative excursion set estimation.

Input : N_{tot} maximum number of evaluations, n size of initial design, q size of batches, function f, threshold t, criterion J, uncertainty function H

Output: $Q_{N_{tot},\rho^\alpha_{N_{tot}}}$ for $\Gamma^\star$ and the uncertainty value H_N

initial DoE select initial DoE $\mathbf{X}_n$, e.g., with space filling design;

evaluate the function f at $\mathbf{X}_n$;

compute the posterior model $Z \mid \mathcal{A}_n$ and the estimate Q_{n,ρ^α_n} ;

$i = n$;

while *i less than N_{tot}* **do**

    $i = i + q$;

update the model   select $\mathbf{x}^{(q)}$ by minimizing $J_i(x)$;

    evaluate the function f at $\mathbf{x}^{(q)}$;

    update the posterior model $Z \mid \mathcal{A}_i$;

    compute the conservative estimate Q_{i,ρ^α_i} ;

post-processing   evaluate the uncertainty function H_i on Q_{i,ρ^α_i} ;

end

optional post-processing on $Q_{N_{tot},\rho^\alpha_{N_{tot}}}$;

A global approach to the optimization of the criteria in Equations (5.11), (5.10), (5.12) and (5.13) often leads to an expensive optimization problem. Instead here we rely on a greedy heuristic approach and optimize the criterion sequentially. First we optimize the criterion with $q = 1$, to find $x^{1,*}$, the first point of the batch. We then find the jth point of the batch by fixing $x^{1,*}, \ldots, x^{j-1,*}$ and by optimizing the criterion at the vector $\widetilde{\mathbf{x}(j)} = (x^{1,*}, \ldots, x^{(j-1),*}, x)$ as a function of the last component only. This heuristic method has already been used in Chevalier et al. (2014a) and in Chevalier et al. (2012) to optimized the batch-sequential criteria introduce there. Here we conducted the single optimizations with the function `genoud` (Mebane and Sekhon, 2011), a genetic algortithm with BFGS descents with numerical gradients.

5.3 Numerical experiments

In this section we apply the proposed sequential uncertainty reduction methods to different test cases. First we develop a benchmark study with Gaussian process realizations to study the different behaviour of the proposed strategies. Then, we apply the methods to two reliability engineering test cases. The first is a coastal flood test case where the set of interest repre-

Strategy number	criterion	parameters
Benchmark 1	IMSE	
Benchmark 2	tIMSE	target=t
A	$J_n(\cdot; \rho_n)$	$\rho_n = 0.5$
B	$J_n(\cdot; \rho_n)$	$\rho_n = \rho_n^\alpha$, $\alpha = 0.95$
C	$J_n^{\text{MEAS}}(\cdot; \rho_n)$	$\rho_n = \rho_n^\alpha$, $\alpha = 0.95$
D	$J_n^{\text{T2}}(\cdot; \rho_n)$	$\rho_n = 0.5$
E	$J_n^{\text{T2}}(\cdot; \rho_n^\alpha)$	$\alpha = 0.95$
F (hybrid strategy)	$J_n^{\text{T2}}(\cdot; \rho_n^\alpha) + \text{IMSE}$	2 iterations IMSE, 1 iteration with E

Table 5.1: Adaptive strategies implemented.

sents the offshore conditions that do not lead the water level at the coast to be larger than a critical threshold above which flood would occur. The data was provided by Jérémy Rohmer and Déborah Idier from the French geological survey (Bureau de Recherches Géologiques et Minières, BRGM). The second test case instead is a nuclear criticality safety problem where the set of interest is the set of unsafe parameters for a nuclear storage facility. The data was provided by Yann Richet from the French Institute of Nuclear Safety (Institut de Radioprotection et de Sûreté Nucléaire, IRSN).

In all test cases we test the strategies detailed in Table 5.1. The strategy $J_n^W(\cdot; \rho_n^\alpha, \beta)$ was also implemented and tested in those test cases for the values $\beta = 0.1$ and $\beta = 10$. The results of those tests are not reported here as they were either too similar to the result obtained with other strategies: J_n for $\beta = 10$ or J_n^{T2} for $\beta = 0.1$.

All computations are carried out in the **R** programming language (R Core Team, 2016), with the packages `DiceKriging` (Roustant et al., 2012) and `DiceDesign` (Franco et al., 2013) for Gaussian modelling, `KrigInv` (Chevalier et al., 2014c) for already existing sampling criterion and `ConservativeEstimates` (Azzimonti and Ginsbourger, 2016) to compute the conservative estimates.

5.3.1 Benchmark study: Gaussian processes

Let us start with a benchmark study for the different strategies introduced in Section 5.2 on Gaussian process realizations in two and five dimensions.

The following set-up is shared by both test cases. We consider the unit hyper-cube $\mathbb{X} = [0, 1]^d \subset \mathbb{R}^d$, $d = 2, 5$ and we defined a Gaussian random field $(Z_x)_{x \in \mathbb{X}}$ with prior constant mean $\mathfrak{m} = 0$ and a tensor product Matérn covariance kernel $\mathfrak{K}$ with $\nu = 3/2$. The parameters of the covariance kernel

Test case	d	covariance parameters	m_{doe}	n_{init}
GP	2	$\nu = 3/2,\ \theta = [0.2, 0.2]^T,\ \sigma^2 = 1$	10	3
GP	5	$\nu = 3/2,\ \theta = \sqrt{\frac{5}{2}}[0.2, 0.2, 0.2, 0.2, 0.2]^T,\ \sigma^2 = 1$	10	6
Costal	2	$\nu = 5/2,\ \text{MLE for } \theta, \sigma^2, \sigma^2_{\text{noise}}$	10	10
Nuclear	2	$\nu = 5/2,\ \text{MLE for } \theta, \sigma^2, \sigma^2_{\text{noise}}$	10	10

Table 5.2: Parameter choice for each test case.

are fixed as in Table 5.2, see Example 2, Chapter 2 for the kernel's explicit formula. The objective is to obtain a conservative estimate at level $\alpha = 0.95$ for $\Gamma = \{x \in \mathbb{X} : Z_x \geq 1\}$. We consider μ as the Lebesgue measure on $\mathbb{X}$.

We consider an initial design of experiments $\mathbf{X}_{n_{\text{init}}}$, obtained with the function `optimumLHS` from the package `lhs`. The field is simulated at $\mathbf{X}_{n_{\text{init}}}$. The size of the initial design of experiment is chosen small to highlight the differences between the sequential strategies. We select the next evaluation by minimizing each criteria detailed in Table 5.1. Each criterion is run for $n = 30$ iterations, updating the model with $q = 1$ new evaluations at each step. For a fixed DoE we consider 10 different realizations of $Z_{\mathbf{X}_{n_{\text{init}}}}$ and we reinitialize the value of the seed at the same value before selecting $x_{n_{\text{init}}+1}$. For a given realization of $Z_{\mathbf{X}_{n_{\text{init}}}}$, each strategy is then run on a different realization as each new evaluation is sampled from Z. The seed of the random number generator is re-initialized at the same value in iteration $n_{\text{init}} + 1$ for each strategy. We consider m_{doe} different initial designs $\mathbf{X}_{n_{\text{init}}}$ of equal size and we report the median value of the uncertainty functions over the repetitions for each strategy at the last iteration.

Dimension 2

We evaluate the strategies by looking at the expected type I and type II errors for $Q_{\rho_n^\alpha}$, defined in Section 5.2, and by computing the measure $\mu(Q_{\rho_n^\alpha})$. For each of these quantities we report the median result over the replications obtained after $n = 30$ evaluations for each initial design.

Expected type I error does not vary much among the different strategies as it is controlled by the probabilistic condition imposed on the estimate, as shown in Section 5.2.1. The distributions of expected type I error and the total time over the different initial DoEs are shown in Appendix 5.5.

Figure 5.1 shows the distribution of expected type II error and the volume $\mu(Q_{\rho_{33}^\alpha})$, i.e. after 30 new evaluations are added. The strategies A, B, C, E, F all provide better uncertainty reduction for conservative estimates than a standard IMSE strategy or than a tIMSE strategy. In particular strategy

(a) Median type II error for $Q_{\rho_n^\alpha}$.   (b) Measure $\mu(Q_{\rho_n^\alpha})$.
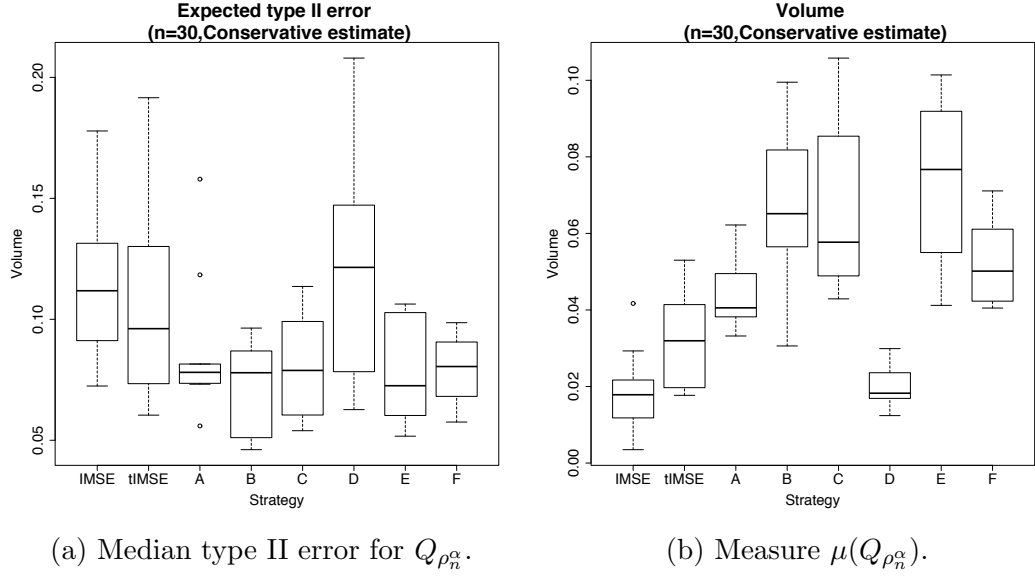
Figure 5.1: Median expected type II errors and measure of the estimate across the different designs of experiments after $n = 30$ iterations of each strategy. Test case in dimension 2.

E has the lowest expected type 2 error while at the same time providing an estimate with the largest measure, thus yielding a conservative set with large measure which is likely to be included in the actual excursion set. All estimates obtained in this study are very conservative: the median ratio between the expected type I error and the estimate's volume is 0.03%, thus much smaller than the upper bound $1 - \alpha = 5\%$ computed in Remark 3. On the other hand the expected type II error is in median 178% bigger than the estimate's volume.

Dimension 5

In Figure 5.2 we show the distribution of expected type II errors for $Q_{\rho_{33}^\alpha}$ and the estimate's measure $\mu(Q_{\rho_{33}^\alpha})$ obtained with different starting design of experiments, for each strategy. The distributions of expected type I error and the total time for those computations are shown in Appendix 5.5.

In this test case the differences between the strategies are less clear. The IMSE strategy provides conservative estimates with small measure and with slightly larger expected type II error. Strategies A, B, C, E provide a good trade off between small expected type II error and large measure of the estimate, however they are not clearly better than the other strategies in this case. Also in this case the estimates provided by the method are very

(a) Median type II error for $Q_{\rho_n^\alpha}$.

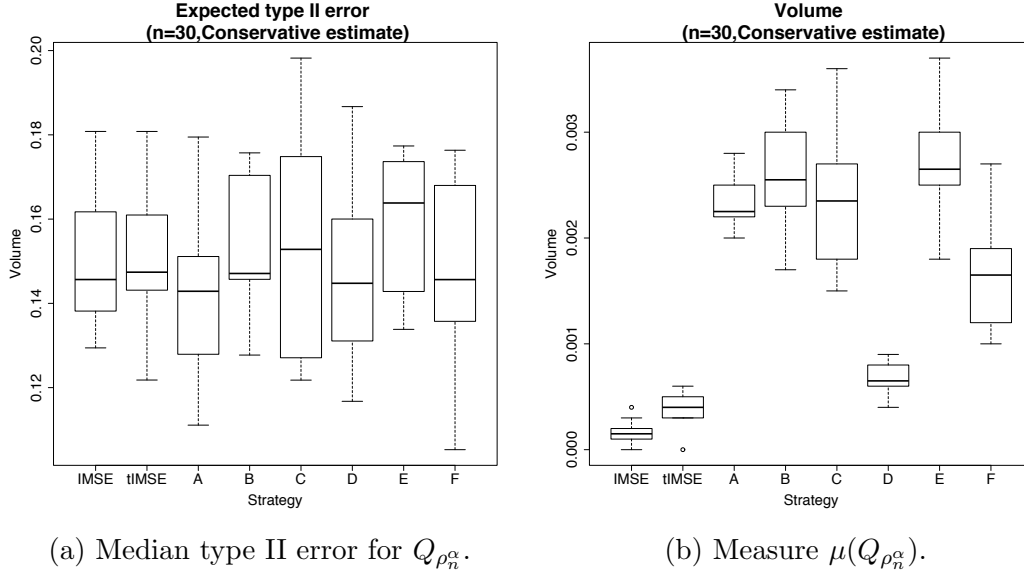(b) Measure $\mu(Q_{\rho_n^\alpha})$.

Figure 5.2: Median expected type II errors and measure of the estimate across different designs of experiments after $n = 30$ iterations of each strategy. Test case in dimension 5.

conservative. The median ratio over all DoEs and all replications between expected type I error and volume is 0.33%, which is smaller than the upper bound of 5% computed in Remark 3. The expected type II error is instead 3 orders of magnitude larger than the estimate's volume. This indicates that we have only recovered a small portion of the true set $\Gamma^\star$, however, under the model, the estimate is very conservative.

5.3.2 Coastal flood test case

In this section we test our strategies on a coastal flood test case introduced in Rohmer and Idier (2012).We focus here on the problem of studying the parameters that lead to floods on the coastlines. This study is often conducted with full grid simulations, see, e.g. Idier et al. (2013) and references therein. This type of simulations require many hours of computational time and render set estimation problems often infeasible. The use of meta-models, recently revised in this field (see, e.g. Rohmer and Idier, 2012, and references therein) eases this computational issue.

We consider a simplified coastal flood case as described in Rohmer and Idier (2012). The water level at the coast is modelled as a deterministic function $f : \mathbb{X} \subset \mathbb{R}^d \to \mathbb{R}$, assuming steady offshore conditions, without solving the flood itself inland. The input space $\mathbb{X} = [0.25, 1.50] \times [0.5, 7]$

represents the variables storm surge magnitude S and significant wave height H_s. We are interested in recovering the set

$$\Gamma^\star = \{x \in \mathbb{X} : f(x) \leq t\},$$

where $t = 2.15$ is a fixed threshold representing the critical height of water for inundations. The region $\Gamma^\star$ is the set of input parameters that leads to a safe value. In order to evaluate the quality of the meta-model and of our estimates we compare the results with a grid experiment of 30×30 runs carried out in Rohmer and Idier (2012).

Here we select a Gaussian process prior $(Z_x)_{x \in \mathbb{X}} \sim GP(\mathfrak{m}, \mathfrak{K})$, with parameters chosen as described in Table 5.2. We assume that the function evaluations are noisy with noise mean equal to zero and homogeneous variance $\sigma_0^2 \geq 0$, estimated from the data. We study the behaviour of sequential strategies starting from $m_{\mathrm{doe}} = 10$ initial DoEs, with equal size $n = 10$ but with different points. The initial DoEs are chosen with a maximin LHS design $\mathbf{X}_n = \{x_1, \ldots, x_n\} \subset \mathbb{X}, n = 10$ with the function `optimumLHS` from the package `lhs`. We evaluate the function f at those points obtaining $\mathbf{f}_{10} = (f(x_1), \ldots, f(x_{10})) \in \mathbb{R}^n$. Figure 5.3a shows the true function f we are aiming at reconstructing, the critical level $t = 2.15$ and one initial design of experiments. The parameters of the covariance kernel are estimated with maximum likelihood from the evaluations $\mathbf{f}_{10}$.

We are interested in conservative set estimates for $\Gamma^\star$ at level $\alpha = 0.95$, as defined in Section 5.2.1. The measure μ is the Lebesgue measure on $\mathbb{X}$.

We proceed to add 20 evaluations with the strategies detailed in Table 5.1. In most practical situations, the covariance hyper-parameters are not known a priori and thus need to be estimated from the data. An initial value for the hyper-parameters is usually obtained from the initial design $\mathbf{X}_n$ with maximum likelihood estimators. If these values are considered accurate enough, they can be fixed throughout the sequential procedure. Most often, however, they are re-estimated at each step, i.e. the function is evaluated at the selected points and the covariance hyper-parameters are estimated with maximum likelihood. See Appendix 5.5.2 for a brief study on the effect of hyper-parameter estimation on the strategies. Here we estimate the hyper-parameters with maximum likelihood for each initial design and we proceed to update those estimates at each step. Figure 5.3b shows an example of conservative estimate obtained from one of the initial DoEs after 30 function evaluations, where the locations are chosen with strategy E.

Figure 5.4 shows the computed volume of the estimate and the true type II error at the last iteration of each strategy, after 30 evaluations of the function. The true type II error is computed by comparing the conservative
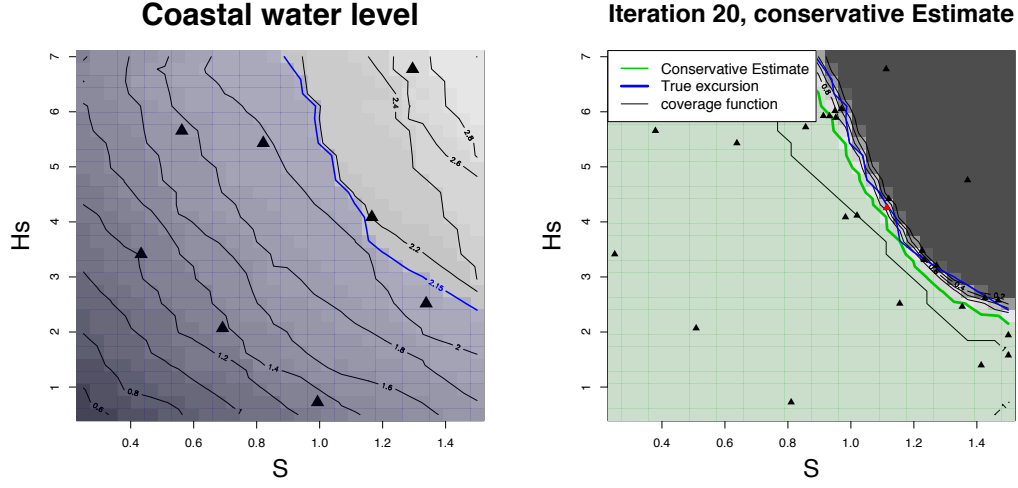
(a) Water level, set of interest (shaded blue, volume=77.56%) and one initial DoE (black triangles).

(b) Conservative estimate (shaded green, $\alpha = 0.95$) after 30 function evaluations (Strategy E).

Figure 5.3: Coastal flood test case. True function (left) and conservative estimate after 30 function evaluations (right).
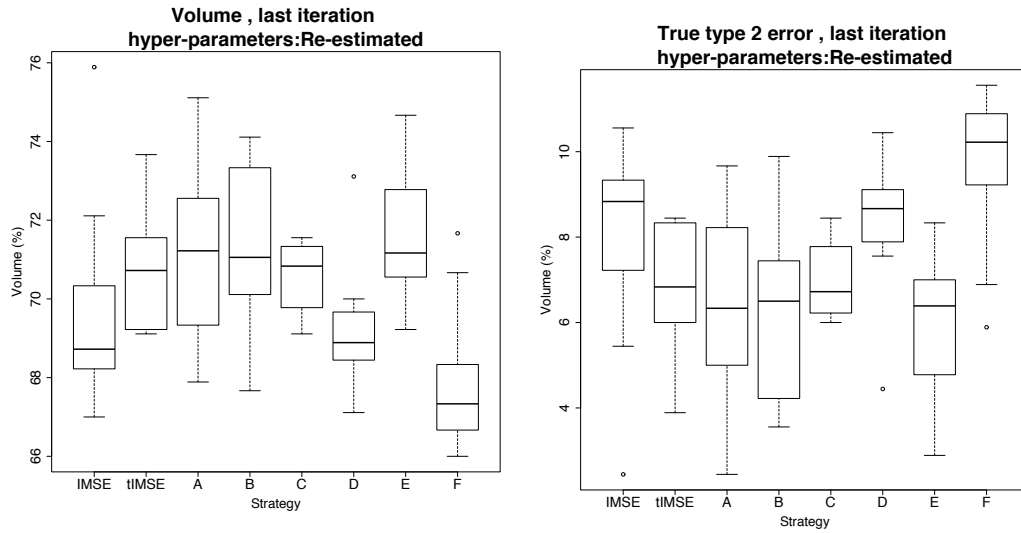
(a) Volume of conservative estimate, $\mu(Q_{\rho_n^\alpha})$.

(b) True type II error.

Figure 5.4: Coastal flood test case, randomized initial DoEs results. Values of the uncertainties for each strategy at the last iteration.
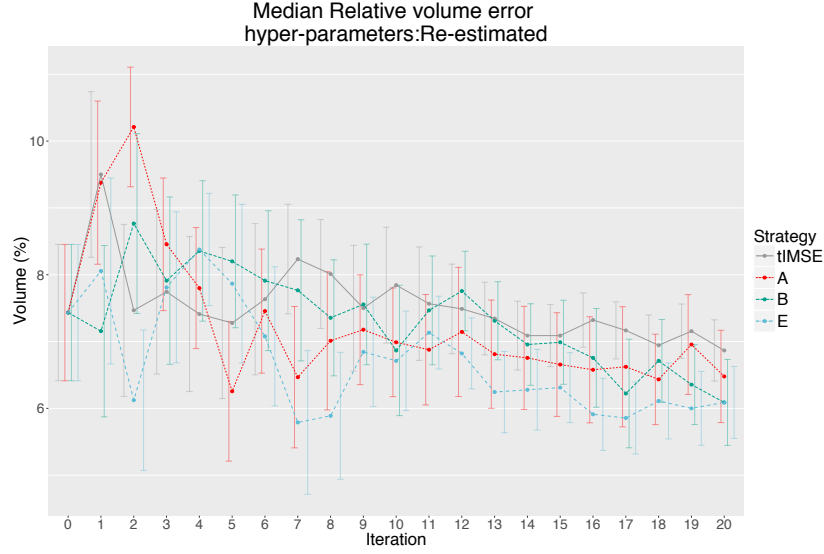
Figure 5.5: Coastal flood test case. Relative volume error as a function of iteration number. Results plotted for strategies tIMSE, A, B, E.

estimate with an estimate of $\Gamma^\star$ obtained from evaluations of the function at a 30×30 grid. Monte Carlo integration over this grid of evaluations leads to a volume of $\Gamma^\star$ equal to 77.56%.

Strategies A, B, E provide estimates with higher volume and lower type II error in median than IMSE and tIMSE. For all strategies the true type I error is zero for almost all initial DoEs, thus indicating that all strategies lead to conservative estimates.

Figure 5.5 shows the behaviour of relative volume error as a function of the iteration number for Strategies tIMSE, A, B, E. The hyper-parameter re-estimation causes the model to be overconfident at the initial iterations, thus increasing the relative volume error. As the number of evaluations increases the hyper-parameter estimates become more stable and each iteration the relative error decreases as conservative estimates are better included in the true set.

5.3.3 Nuclear criticality safety test case

In this section we apply the proposed strategies in a reliability engineering test case from the French Institute of nuclear safety (IRSN).

The problem at hand concerns a nuclear storage facility and we are interested in estimating the set of parameters that lead to a safe storage of the material. The safety of such system is closely linked to the production

of neutrons. In fact, since neutrons are both the product and the initiator of nuclear reactions, an overproduction could lead to a chain reaction. For this reason the safety of such systems is usually evaluated with the neutron multiplication factor, here called $k-$effective or k-eff $: \mathbb{X} \to [0, 1]$. The input space $\mathbb{X} = [0.2, 5.2] \times [0, 5]$ represents the fissile material density, PuO_2, and the water thickness, H_2O. A value k-eff$(PuO_2, H_2O) > 1$ denotes a configuration of parameters which is not safe. We are thus interested in recovering the set of safe configurations

$$\Gamma^{\star} = \{(PuO_2, H_2O) \in \mathbb{X} : \text{k-eff}(PuO_2, H_2O) \leq 0.92\},$$

where the threshold was chosen at 0.92 for safety reasons.

In general, the values of k-eff are computed with expensive computer experiments thus our objective is to provide an estimate for $\Gamma^{\star}$ from few evaluations of k-eff and to quantify its uncertainty. As in the previous sections here we study in particular adaptive methods to reduce the uncertainties on the estimate.

The values for the function k-eff were obtained with a MCMC simulator whose results have an heterogeneous noise variance. In Appendix 5.5.3 we consider the noise homogeneous and we estimate it from the data. Such procedure, however, does not take into account the heterogeneous noise thus leading to possibly wrong estimates. Moreover the criteria introduced are not implemented to take into account an heterogeneous noise variance in the choice of the optimal points. This could lead to the choice of sub-optimal locations in cases where the noise variance is strongly heterogeneous. In order to avoid such pitfalls we select a deterministic function by replacing the k-eff values with a smoothed approximation. The approximation is computed from a 50×50 grid of evaluations of k-eff in $\mathbb{X}$. We assume that k-eff is a Gaussian random field with mean zero, tensor product Matérn ($\nu = 5/2$) covariance kernel and heterogeneous noise variance equal to the MCMC variance. We then compute the kriging mean of this field given 2500 observations at the 50×50 grid and we use this function as the true function. In what follows we denote with k-eff the smoothed function.

We consider the function k-eff as a realization of a Gaussian random field and we fix a prior $(Z)_{x \in \mathbb{X}} \sim GP(\mathfrak{m}, \mathfrak{K})$ with constant mean function $\mathfrak{m}$ and tensor product Matérn covariance kernel with $\nu = 5/2$. We consider $m_{\text{doe}} = 10$ different initial DoEs of size $n_0 = 10$. In particular we consider LHS designs obtained with the function `optimumLHS` from the package `lhs` in **R**. We evaluate k-eff at each design obtaining the vector of evaluations k-eff$(\mathbf{X}_{10}) = (\text{k-eff}(x_1), \ldots, \text{k-eff}(x_{10}))$.

Figure 5.6a shows the true values of the function, the set $\Gamma^{\star}$ and one initial design of experiments. The initial evaluations are used to estimate
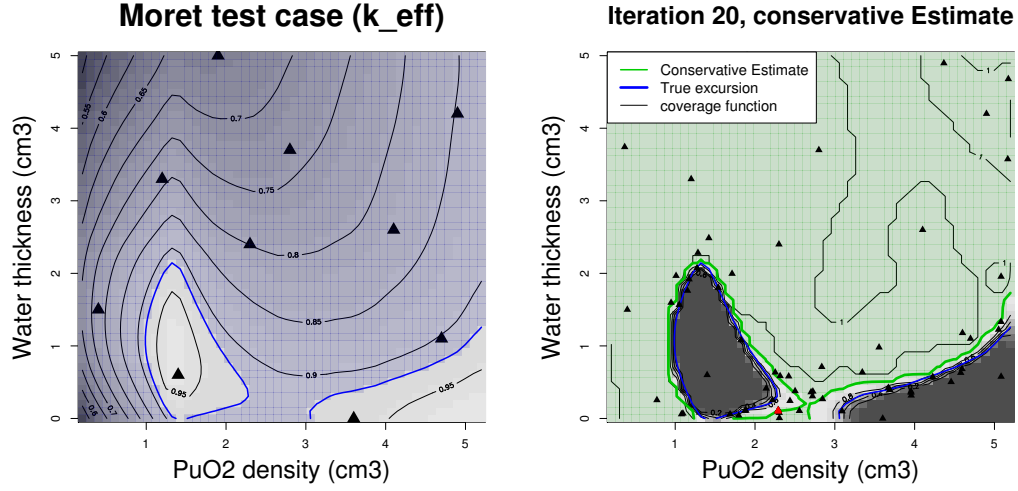
(a) One initial design of experiments (b) Conservative estimate (shaded green, (black triangles) for k-eff, set of interest $\alpha = 0.95$) after 70 function evaluations (shaded blue, volume=88.16%).   (Strategy E).
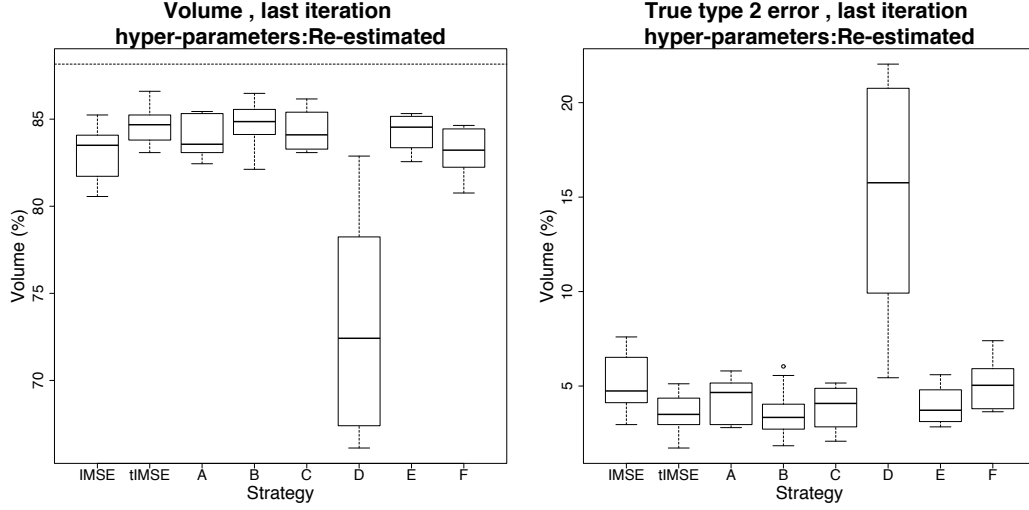
Figure 5.6: Nuclear criticality safety test case. Smoothed function (left) and conservative estimate (right) after 70 evaluations.

the covariance function's hyper-parameters with maximum likelihood. The starting model provides a conservative estimate for $\Gamma^\star$ at level $\alpha = 0.95$. Here we choose the Lebesgue measure μ on $\mathbb{X}$.

We now test how to adaptively reduce the uncertainty on the estimate with different strategies. We run $n = 20$ iteration of each strategy and at each step we select a batch of $q = 3$ new points where to evaluate k-eff. The covariance hyper-parameters are re-estimated at each iteration by adding the 3 new evaluations to the model. A brief study on the effect of hyper-parameter estimation on the strategies is reported in Appendix 5.5.4.

Figure 5.7 shows a comparison of the conservative estimate's volume and its type II error at the last iteration, i.e. after 70 evaluations of the function, for each initial DoE and each strategy. Also in this test case the strategies A, B, E perform well both in terms of final volume and true type II error. Strategy C compares better to the other strategy than in the previous test cases. In this example strategy C provides large volumes, however it also has a larger type II error.

Figure 5.8 shows the relative volume error as a function of the iteration number for strategies tIMSE, A, B, C, E. The relative volume error is computed by comparing the conservative estimate with an estimate of $\Gamma^\star$ obtained from evaluations of k-eff on a grid 50×50. The volume of $\Gamma^\star$ com-

(a) Final volume of conservative estimate, $\mu(Q_{\rho_n^\alpha})$, and true volume $\mu(\Gamma^\star)$ (dashed line, 88.16%).

(b) True type II error at the final iteration computed comparing the estimate with $\Gamma^\star$.

Figure 5.7: Nuclear criticality safety test case, randomized initial DoEs results. Values of the uncertainties for each strategy at the last iteration.
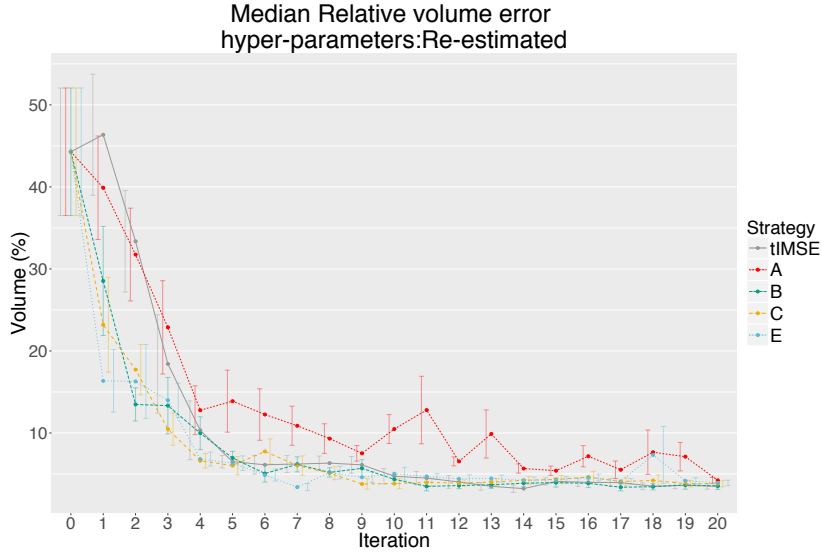

Figure 5.8: Nuclear criticality safety test case. Relative volume error as a function of the iteration number. Strategies tIMSE, A, B, C, E.

puted with Monte Carlo integration from this grid of evaluations is equal to
88.16%. All strategies presented show a strong decrease in relative volume
error in the first 6 iterations, i.e. until 28 evaluations of k-eff are added.
In particular strategies B, C, E show the strongest decline in error in the
first 3 iterations. Strategy A while also showing a strong decrease in relative
volume error, it returns estimates with higher errors at almost all iterations
except for the last one. Overall, as in the previous test cases, strategy E, the
minimization of the expected type II error, seem to provide the best uncer-
tainty reductions both in terms of relative volume error and in terms of type
II error.

5.4   Discussion

In this chapter we introduced sequential uncertainty reduction strategies for
conservative estimates. This type of set estimates proved to be useful in re-
liability engineering, however they could be of interest in all situations when
overestimating the set has more important consequences than underestimat-
ing it. The conservative aspect however is strongly relying on the goodness of
the underlying random field model. In fact, since the estimator CE is based
on a global quantity, a model that approximates badly the true function will
not lead to reliable estimates. For a fixed model, this aspect might be tam-
pered by increasing the level of confidence α. In the test cases presented here
we fixed $\alpha = 0.95$, however in real life situations, testing different levels of α,
such as, e.g. $\alpha = 0.99, 0.995$, and comparing the results is a good practice.

The sequential strategies proposed here provide a way to reduce the un-
certainties on the estimate by adding new function evaluations. In particular
strategy E, i.e. the criterion $J_n^{\mathrm{r}2}(\cdot; \rho_n^{\alpha})$, resulted among the best criteria in
terms of Type 2 uncertainty and relative volume error in all test cases. In
this work we mainly focused on showing the differences between the strategy
with a-posteriori measures of the uncertainties. Nonetheless the expected
type I and II errors could be used to provide stopping criteria for the sequen-
tial strategies. Further studies in this direction are needed to understand the
limit behaviour of these quantity as the number of evaluation increases.

The reliability engineering test cases gave a first glimpse on the practical
use for conservative SUR strategies. However these test cases also high-
lighted shortcomings of the current implementation and possible future im-
provements. In particular, the data provided for the nuclear criticality safety
test case was obtained with a noisy function. In this work we studied the
strategies on a smoothed approximation of this noisy function, thus avoiding
the heterogeneous case. In cases where the noise variance can be considered

homogeneous, it should not strongly affect the uncertainty reduction strategies in their current implementation. However, if the noise variance varies widely, then the current implementation of the criteria might select suboptimal locations. More studies are required to implement and test criteria that organically take into account heterogeneous noise variance.

The conservative SUR strategies proposed here focus on reducing the uncertainties by adding evaluations points close to the boundary of Γ. This choice however does not necessarily lead to a better global model than space filling strategies. Since conservative estimates are strongly dependent on the underlying model the benefits of a well identified boundary might be overshadowed by a bad global model. On the other end, a strategy that explores the space, such as IMSE, might lead to a good global model with less reliable conservative estimates because the boundary is not well identified. Finally in the test case section we only briefly touched the fundamental problem of hyper-parameter estimation. The conservative SUR strategies were not implemented to provide good hyper-parameter estimation. In both test cases we obtained rather stable parameter estimations which did not lead to odd behaviours of conservative estimates, however in general this might not be the case. The sequential behaviour of hyper-parameters maximum likelihood estimators under SUR strategies needs to be studied in more details. The experimental results highlighted the need for strategies that are able to select locations that provide a good trade off between exploration of the space, boundary identification and reliable hyper-parameter estimation.
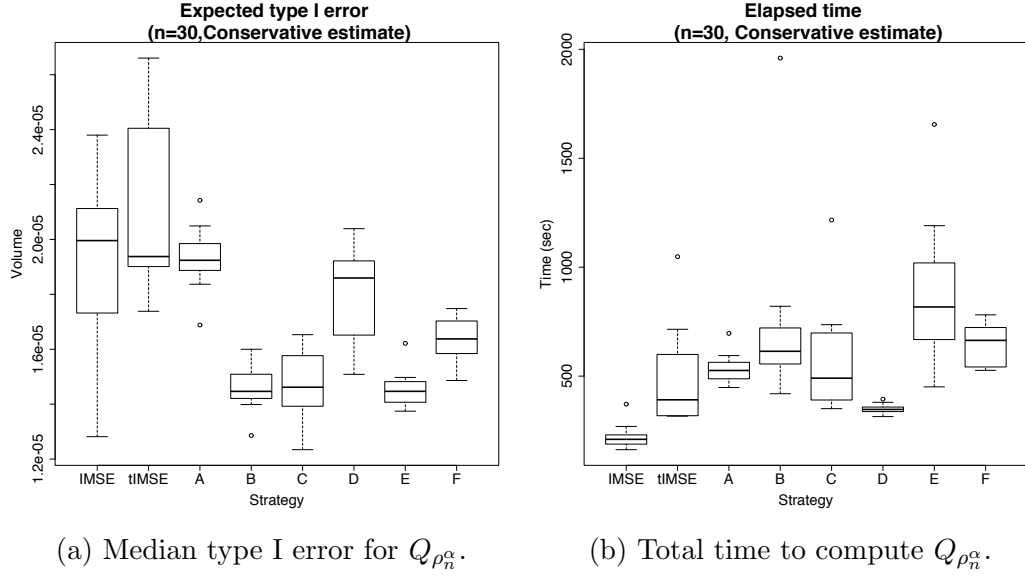
(a) Median type I error for $Q_{\rho_n^\alpha}$.   (b) Total time to compute $Q_{\rho_n^\alpha}$.

Figure 5.9: Two-dimensional Gaussian realizations test case. Type I error of conservative estimate and elapsed time in the 2-dimensional test case.

5.5   Appendix: additional numerical results

5.5.1   Benchmark study on Gaussian processes

In this section we present more details about the test cases on Gaussian process realizations described in Section 5.3.1.

Dimension 2

Figure 5.9a shows the type I error for each strategy in Table 5.1. Strategies B, C and E show a lower type I error with respect to the other strategies, however the all strategies present very low type I error compared to the total expected volume of the set.

Figure 5.9b shows the total time required to evaluate the criteria and to compute at each step the conservative estimate. The computational time is mainly driven by the size of the conservative estimate. In fact, for conservative estimates with large volumes, the othant probabilities involved in its computation are higher dimensional.

Dimension 5

Figure 5.10 shows the type I error and the total time required to evaluate the criteria and to compute at each step the conservative estimate. In this case

(a) Median type I error for $Q_{\rho_n^\alpha}$.

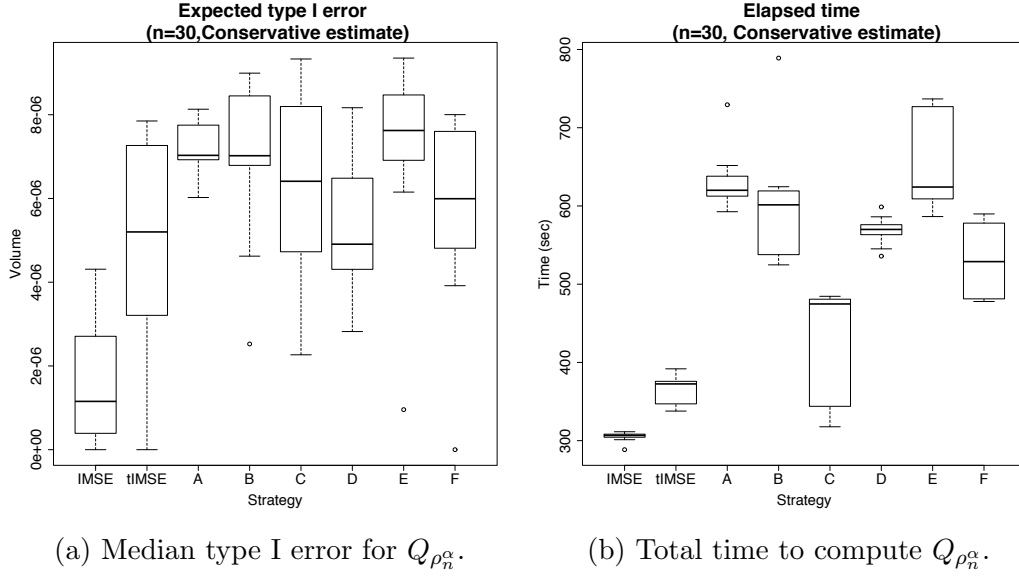(b) Total time to compute $Q_{\rho_n^\alpha}$.

Figure 5.10: Five-dimensional Gaussian realizations test case. Type I error for conservative estimate and elapsed time in the 5-dimensional test case.

strategies IMSE presents a lower type I error. This might be caused by the better global approximation obtained with a model based on a space filling design.

Figure 5.10b shows the total time required to evaluate the criteria and to compute at each step the conservative estimate.

5.5.2 Coastal flood test case: hyper-parameter estimation

In this section we explore the behaviour of the strategies under different scenarios for the covariance hyper-parameters:

1. *fixed initial hyper-parameters (FI)*: the covariance parameters are fixed throughout the sequential procedure as the maximum likelihood estimates obtained from the initial evaluations $\mathbf{f}_{10}$;

2. *re-estimated hyper-parameters (RE)*: at each step the hyper-parameter estimates are updated with the new evaluations of the function;

We consider the same experimental setup of Section 5.3.2, we fix only one initial DoE of $n = 10$ points chosen with the function `lhsDesign` from the package `DiceDesign` (Franco et al., 2013). We run $n = 20$ iterations of each strategy in Table 5.1 where at each iteration we select one evaluation of f.
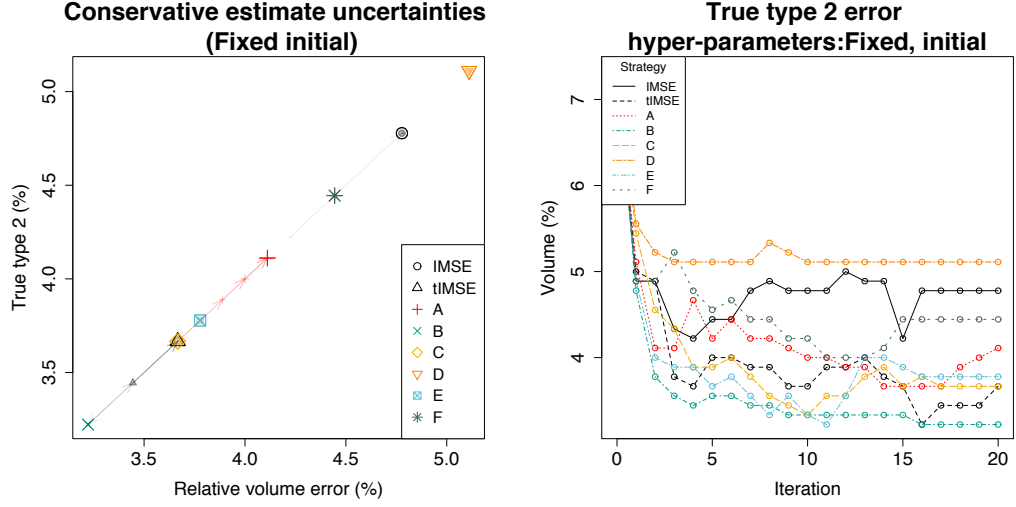
(a) Type II error versus relative volume error for $Q_{\rho_n^\alpha}$ for the last 5 iterations.

(b) Type II error of $Q_{\rho_n^\alpha}$ at each n computed with respect to the true set.

Figure 5.11: Coastal flood test case. True type II error and relative volume error for $Q_{\rho_n^\alpha}$, with no re-estimation of the covariance parameters.

Let us start with the case where we estimate the hyper-parameters only using the evaluations of f at $\mathbf{X}_{10}$. Figures 5.11 show the type II error and the relative total volume error. These errors are both computed comparing the conservative estimate with the true set obtained with evaluations of f on a fine grid. True type I error, not shown, is equal to zero for each strategy at each iteration. Type II error decreases as a function of the evaluations number for all strategies, in particular strategies B, tIMSE, C and E provide a good uncertainty reduction. In particular strategies B, C and E provide a faster uncertainty reduction in the first 10 iterations than the other strategies. Strategy D focuses on reducing the type II error for the Vorob'ev median and it is not well adapted to reduce uncertainties on conservative estimates.

In practice, the covariance parameters are often re-estimated after a new evaluation is added to the model. While this technique should improve the model it leads to better conservative estimates only if the hyper-parameter estimation is stable and reliable. Conservative estimates are in fact based on the coverage probability function and in particular on high quantiles of this function.

Figure 5.12 shows the Type II and relative volume error computed comparing the conservative estimate to the true set in the case where covariance parameters are re-estimated at each step. During the first 10 iterations, all strategies except IMSE show a small (less than 1%) type I error, not shown,
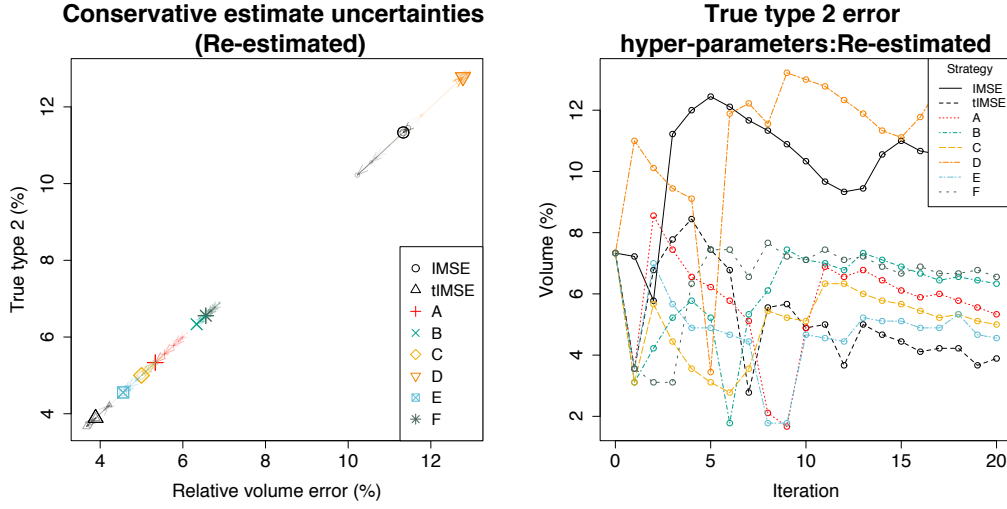
(a) Type II error versus relative volume error for $Q_{\rho_n^\alpha}$ at each n.

(b) Type II error for $Q_{\rho_n^\alpha}$ computed with respect to the true set at each n.

Figure 5.12: Coastal flood test case. True type II error and relative volume error for $Q_{\rho_n^\alpha}$ at each n, with re-estimation of the covariance parameters at each step.

which becomes equal to zero for all strategies after iteration 10.

The strategies IMSE and D still show the worst behaviour among the tested strategies both in terms of true type II error and of relative volume error, however the remaining strategies do not show big differences. The tIMSE show the best behaviour closely followed by Strategy E, C, A. The differences between the final estimated set obtained with these four strategies are small and they are mainly due to a difference in the hyper-parameter estimation. The tIMSE strategy produces more stable hyper-parameter estimators than Strategy E, where the range parameters decrease in the last steps. This change leads to smaller more conservative set estimates.

If the covariance hyper-parameters are kept fixed, Figure 5.11b shows that the true type II error tends to stabilize because the conservative estimate is the best according to the current model. On the other hand the re-estimation of the parameters leads to a more unstable type II error which also indicates that the underlying model is adapting to the new observations.

The re-estimation of the covariance parameters at each steps might lead to instabilities in the maximum likelihood estimators. In this test case the parameter estimation is very stable, however further studies are required to better understand the behaviour of maximum likelihood estimators under these circumstances.
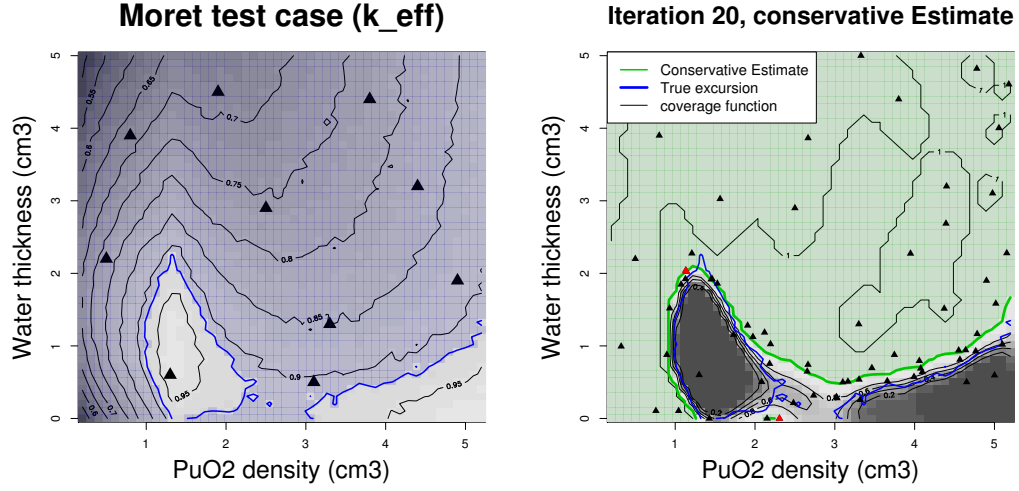
(a) One initial design of experiments (b) Conservative estimate (shaded green,
(black triangles) for k-eff, set of interest $\alpha = 0.95$) after 70 function evaluations
(shaded blue, volume=87.92%).   (Strategy E).

Figure 5.13: Nuclear criticality safety test case, homogeneous noise. True
function (left) and conservative estimate (right)

5.5.3   Nuclear criticality safety test case: homogeneous noise results

Here we report the results on the true function k-eff, under the assumption
of unknown homogeneous noise variance. The data provided by IRSN have
an heterogeneous noise variance, however the variability of such variance is
not very high, so in a first approximation it can be considered homogeneous.

As in Section 5.3.3, we assume that k-eff is realization of a Gaussian
random field and we fix a prior $(Z)_{x \in \mathbb{X}} \sim GP(\mathfrak{m}, \mathfrak{K})$ with constant mean
function $\mathfrak{m}$ and tensor product Matérn covariance kernel with $\nu = 5/2$. We
consider the function evaluations noisy with homogeneous noise mean equal
to zero and variance $\sigma_0^2 \geq 0$, estimated from the data. We consider $m_{\mathrm{doe}} = 10$
different initial DoEs of size $n_0 = 10$ chosen in the same manner.

Figure 5.13a shows the true values of the function, the set $\Gamma^\star$ and one
initial design of experiments. The initial evaluations are used to estimate
the covariance function's hyper-parameters with maximum likelihood. The
starting model provides a conservative estimate for $\Gamma^\star$ at level $\alpha = 0.95$. The
measure μ is the Lebesgue measure on $\mathbb{X}$.

We now test how to adaptively reduce the uncertainty on the estimate
with the strategies in Table 5.1. We run $n = 20$ iteration of each strategy
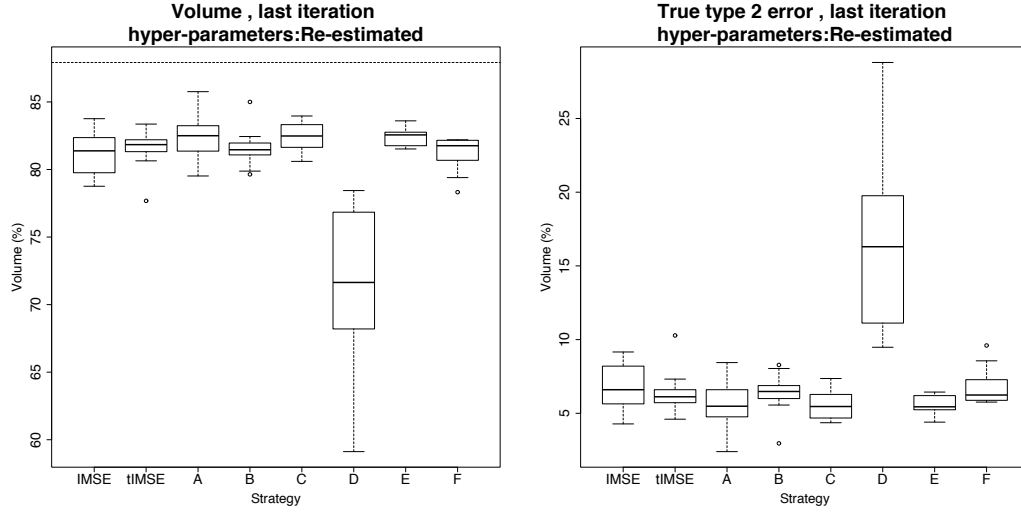
(a) Final volume of conservative esti-
mate, $\mu(Q_{\rho_n^\alpha})$, and true volume $\mu(\Gamma^\star)$
(dashed line, 87.92%).

(b) True type II error at the final iter-
ation computed comparing the estimate
with $\Gamma^\star$.

Figure 5.14: Nuclear criticality safety test case, randomized initial DoEs
results. Values of the uncertainties for each strategy at the last iteration.

and at each step we select a batch of $q = 3$ new points where to evaluate
k-eff. The covariance hyper-parameters are re-estimated at each iteration
by adding the 3 new evaluations to the model. Figure 5.13b shows the
conservative estimate for $\Gamma^\star$ computed after 70 function evaluations with
locations chosen with Strategy E.

Figure 5.14 shows a comparison of the conservative estimate's volume and
its type II error at the last iteration, i.e. after 70 evaluations of the function,
for each initial DoE and each strategy. Also in this test case the strategies
A, B, E perform well both in terms of final volume and true type II error.
Strategy C also performs well in this test case, as opposed to the previous
examples. Figure 5.15 shows the relative volume error as a function of the
iteration number for strategies tIMSE, A, B, C, E. The relative volume error
is computed by comparing the conservative estimate with an estimate of $\Gamma^\star$
obtained from MCMC mean value of k-eff on a grid 50×50. The volume of $\Gamma^\star$
computed with Monte Carlo integration from this grid of evaluations is equal
to 87.92%. All strategies presented show a strong decrease in relative volume
error in the first 10 iteration, i.e. until 40 evaluations of k-eff are added. In
particular strategies B, C, E show the strongest decline in error in the first 5
iterations. In this test case strategy A shows a stronger decrease in median
relative error compared to the smoother test case presented in Section 5.3.3.

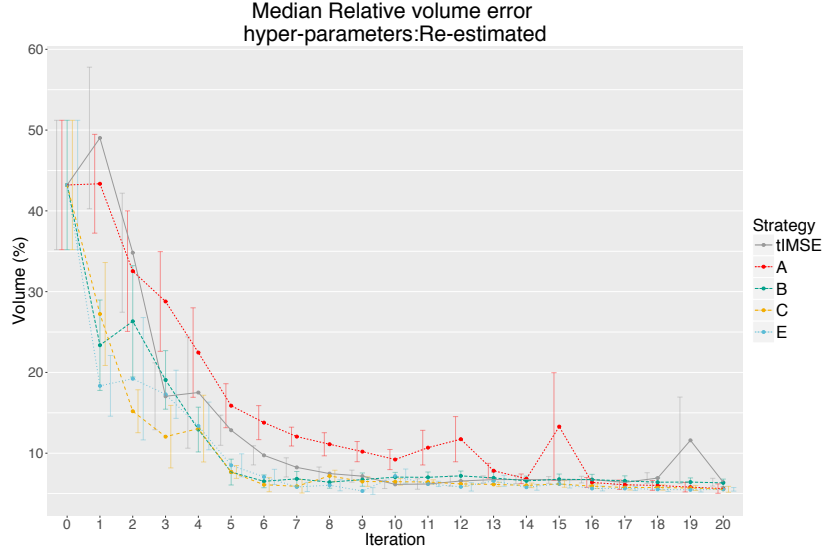Figure 5.15: Nuclear criticality safety test case, homogeneous noise. Relative volume error for iteration number. Strategies tIMSE, A, B, C, E.

Overall, as in the previous test cases, strategy E, the minimization of the expected type II error, seem to provide the best uncertainty reductions both in terms of relative volume error and in terms of type II error.

5.5.4 Nuclear criticality safety test case: hyper-parameter estimation

As in the previous test case the re-estimation of the covariance parameters could have an important effect on conservative estimates. In this section we explore the behaviour of the strategies in the IRSN test case under the scenarios for the hyper-parameters introduced in Section 5.5.2: *fixed initial hyper-parameters (FI)* and *re-estimated hyper-parameters (RE)*.

We consider the same experimental setup as in Section 5.5.3, however here we fix one initial design of experiment $\mathbf{X}_{15} = \{x_1, \ldots, x_{15}\} \subset \mathbb{X}$ chosen with the function `maximinESE_LHS` from the package `DiceDesign` (Franco et al., 2013) which implements an Enhanced Stochastic Evolutionary (ESE) algorithm for LHS design. We evaluate k-eff at that design obtaining the vector of evaluations k-eff$(\mathbf{X}_{15}) = ($k-eff$(x_1), \ldots, $k-eff$(x_{15}))$.

Figure 5.16 shows the Type I, II error and the relative volume error computed comparing the conservative estimate to the true set in the case where the covariance parameters are kept fixed. Figure 5.16a shows the values of type II versus relative volume errors for the last 5 iterations, i.e.
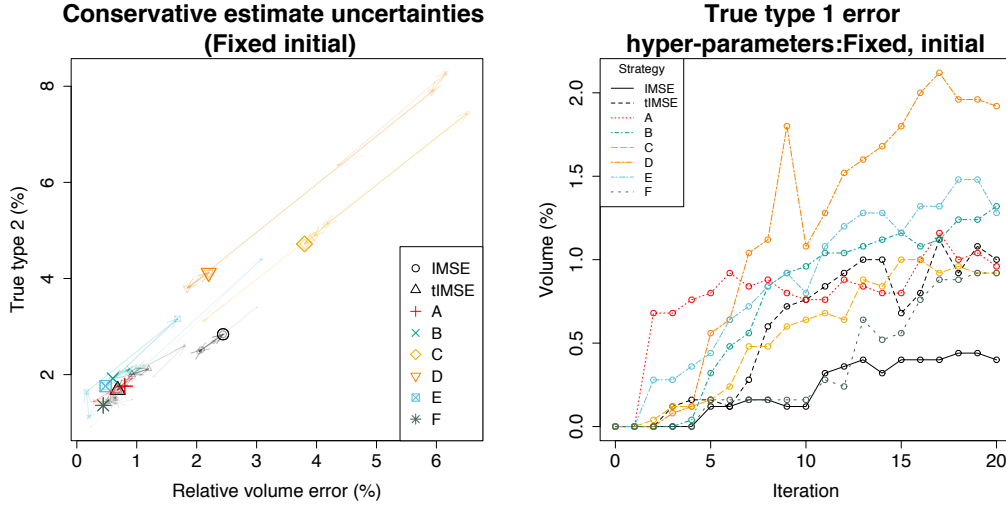
(a) Type II vs relative volume error of $Q_{\rho_n^\alpha}$ with respect to the true set for the last 5 iterations (15 evaluations).

(b) Type I error of $Q_{\rho_n^\alpha}$ computed with respect to the true set $\Gamma^\star$ at each iteration.

Figure 5.16: Nuclear criticality safety test case. True type II, relative volume and type I errors for $Q_{\rho_n^\alpha}$ with no re-estimation of the covariance parameters.

the last 15 evaluations. The points do not lie all on the diagonal which implies that type I errors are not zero. In fact, Figure 5.16b shows an increase in true type I error for all strategies as new evaluations are added. This effect, while limited, might be caused by the initial model inaccuracy. Strategies C, D and IMSE are the worst performers both in terms of true type II error and in terms of relative volume error. The other strategies show a very similar behaviour. The decrease of type II error is partially compensated by an increase in type I error, however the overall relative error decreases.

In the more realistic scenario where the covariance parameters are re-estimated at each step we obtain smaller, more conservative sets with each strategy. For example, the estimate's volume obtained at the final iteration, after 75 function evaluations, with Strategy F and re-estimation of the parameters is 4% smaller than the value obtained with fixed initial parameters.

Figure 5.17 shows the Type II versus relative volume errors and true type I error computed comparing the true set with the conservative estimate obtained with covariance parameters re-estimated at each step.

The re-estimation of the parameters leads to a more accurate global model. This accuracy is reflected in a decrease of all errors as evaluations are added. In this scenario the strategies E, F, A show a bigger uncertainty reduction both in terms of type II error and in terms of relative volume error.

(a) Type II versus relative volume errors (b) True type I error of $Q_{\rho_n^\alpha}$ with $\alpha = $ of $Q_{\rho_n^\alpha}$ with respect to the true set $\Gamma^\star$ for 0.95, computed with respect to $\Gamma^\star$ at each the last 5 iterations (15 k-eff evaluations). iteration.
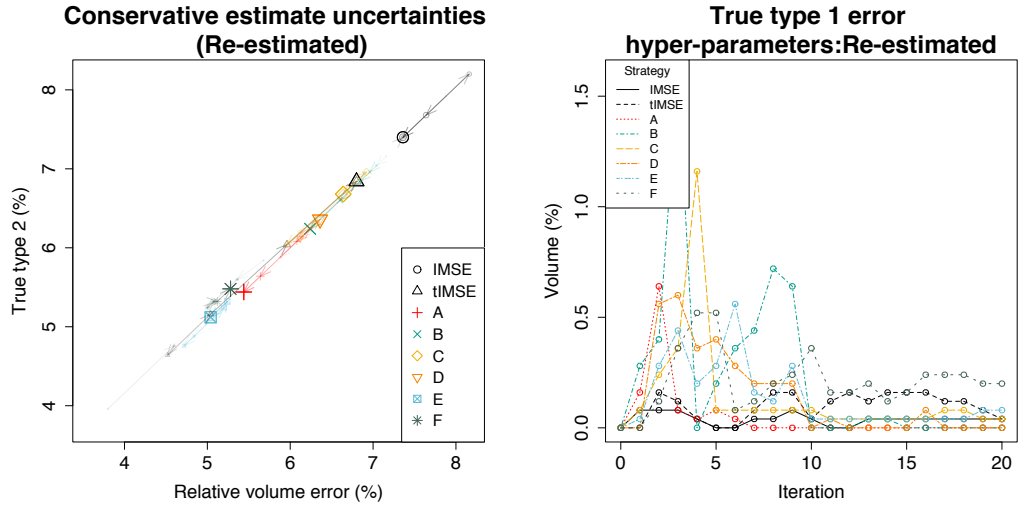
Figure 5.17: Nuclear criticality safety test case. True type II, relative volume and type I error for $Q_{\rho_n^\alpha}$ with re-estimation of the covariance parameters at each step.

Chapter 6

Conclusion and future works

In this manuscript we analysed the problem of estimating excursion sets of expensive to evaluate functions with Bayesian methods relying on Gaussian random field priors.

These methods are increasingly studied in the literature, both from theoretical and applied perspectives. In Chapter 2, we briefly revisited the state of the art in Bayesian set estimation with Gaussian process priors and we highlighted some of the links with the theory of random closed sets. One of those links is exposed in Chapter 3, where we revisited the concept of distance average expectation and its notion of variability in the framework of excursion sets estimation. Distance average variability brings a new way to quantify uncertainties on the excursion set. Its computation, however, requires conditional simulations of the posterior field discretized on high resolution designs. We provided a technique to compute this quantity from posterior field quasi-realizations. This method allows uncertainty quantification on excursion sets, with the distance average variability, for a fraction of the computational cost required to obtain the same assessment with full posterior field realizations. The quasi-realization technique proved to be reliable also for studying contour lines and excursion sets volume distributions. The quasi-realizations generated with this method, however, come from a reconstructed process which is smoother than the original one. In fact, this leads, in some cases, to excursion set estimates that are too smooth. In general however, the quasi-realization technique could make the computation of other empirical quantities related to excursion sets more affordable and it could open the way for different excursion set estimators.

Another technique, recently introduced (Bolin and Lindgren, 2015) in the literature of Bayesian set estimation, is conservative set estimation. Conservative estimates give a probabilistic statement on the accuracy of the estimate under the model. In particular, they are helpful in situations where the

consequences of underestimating an excursion set are much less important than overestimating it, such as in reliability engineering, climate science applications and geological studies. While the probabilistic statement given by these estimates can be very useful, it makes them computationally difficult to obtain as they require many successive evaluations of high dimensional Gaussian orthant probabilities. In Chapter 4, we introduced a method to efficiently compute Gaussian orthant probabilities and we implemented it in two algorithms: GMC and GanMC. Both algorithms were tested against state-of-the-art techniques and they showed substantial efficiency increases for high dimensional examples. Nevertheless, the behaviour of the method underlying both algorithms in general settings requires further studies. For example, a better understanding of the effect of different covariance structures could lead to better estimators. The main difference between the two algorithms is that, while GMC uses a standard Monte Carlo method, GanMC introduces a novel asymmetric nested Monte Carlo technique that exploits a model of computational costs to provide a more efficient estimator. This technique is used in our orthant probability algorithm to compute a specific conditional probability. We show on numerical test cases how, depending on the type of probability to be estimated, the efficiency gains brought by anMC can be more or less substantial. Asymmetric nested Monte Carlo, however, could bring more efficient MC estimators also in other applications. The anMC method is particularly well suited when the quantity to be sampled is a function of two correlated random variables that have very different sampling costs. If the cost functions satisfy the assumptions outlined in Chapter 4, an implementation of anMC could bring large computational savings. The practical feasibility of such extensions however depends on the availability of a good model for the cost functions.

Conservative estimates, as presented in the literature (Bolin and Lindgren, 2015) and in Chapter 4, are based on a fixed design of experiments. In many practical situations, however, it is important to adaptively reduce the uncertainty on these estimates by evaluating the expensive function at meaningful locations. In fact, the choice of adaptive sequential designs is a central topic in the field of computer experiments. Such adaptive methods have been successfully employed in Bayesian optimization where the object of interest is the minimum of the function. In Bayesian set estimation, similar methods have been proposed, see e.g. Chevalier (2013), where a sequential method relying on the Vorob'ev variability is developed. In Chapter 5, we presented new sequential strategies for conservative estimates. We introduced them in the framework of Stepwise Uncertainty Reduction (SUR) methods (Bect et al., 2012) and we presented three numerical studies to compare the introduced criteria with state-of-the-art criteria. In the test

cases presented, criteria adapted to conservative estimates provide better uncertainty reduction when compared to generic criteria, such as the IMSE criterion. In Chapter 5, the strategies were introduced in an empirical Bayes framework. This setup raises further questions on the behaviour of covariance hyper-parameters when re-estimated from new evaluations. In fact, while in the test cases presented the hyper-parameter estimation was very stable, conservative estimates strategies are not necessarily well adapted to hyper-parameter estimation. Further studies are needed to develop hybrid or global criteria that simultaneously reduce uncertainties and provide consistent hyper-parameter estimation. Moreover, the convergence of a class of SUR criteria was proven very recently (Bect et al., 2016). Further work might involve the extension of such result to the criteria introduced in Chapter 5. Finally, from an application point of view, conservative estimates allow to control the overestimation of the set under the model, therefore they are well suited for reliability engineering and safety applications. In such applications, however, it is of paramount importance to pair these strategies with diagnostic tools such as visualizations and appropriate stopping criteria.

The methods proposed in this manuscript require a Gaussian process (GP) model to surrogate the function. As such, their outcome strongly depends on the quality of this underlying model. However, they are flexible on the type of Gaussian process model. In fact, the introduced set estimation techniques could be applied almost directly with sparse underlying GP model (see, e.g., Snelson and Ghahramani, 2006; Hensman et al., 2013, and references therein) or low rank models (see, e.g. Banerjee et al., 2008; Katzfuss, 2016, and references therein). They could also be adapted to fully Bayesian GP models and to other types of GP models in relatively straightforward ways. Future work in this direction could lead to set estimation techniques that better account for uncertainty or that are better adapted to industrial scale problems.

Further theoretical studies of functionals related to random closed sets could lead to the development of alternative set estimates. For example, the study of differential properties of the coverage probability function of an excursion set might help to summarize its distribution in a more effective way. Finally, much work is still needed to bridge the gap between the probabilistic theory of random closed sets and the more applied Bayesian set estimation world. This manuscript made a first, partial attempt in this direction; however much work remains to be done.

The methods introduced in Chapter 4 of this thesis were developed in the **R** package `ConservativeEstimates` in order to spread their use. The package implements the conservative estimates algorithm along with GMC and GanMC algorithms for orthant probabilities and it is available on GitHub.

Bibliography

Abrahamson, I. (1964). Orthant probabilities for the quadrivariate normal distribution. *The Annals of Mathematical Statistics*, 35(4):1685–1703. (Cited on page 75.)

Adler, R. J. (2000). On excursion sets, tube formulas and maxima of random fields. *The Annals of Applied Probability*, 10(1):1–74. (Cited on page 68.)

Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*. Springer, Boston. (Cited on pages 11 and 12.)

Artstein, Z. and Vitale, R. A. (1975). A strong law of large numbers for random compact sets. *The Annals of Probability*, 3(5):879–882. (Cited on page 23.)

Aumann, R. J. (1965). Integrals of set-valued functions. *Journal of Mathematical Analysis and Applications*, 12(1):1–12. (Cited on page 23.)

Ayala, G., León, T., and Zapater, V. (2005). Different averages of a fuzzy set with an application to vessel segmentation. *IEEE Transactions on Fuzzy Systems*, 13(3):384–393. (Cited on page 39.)

Azzimonti, D., Bect, J., Chevalier, C., and Ginsbourger, D. (2016a). Quantifying uncertainties on excursion sets under a Gaussian random field prior. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):850–874. (Cited on pages 43 and 66.)

Azzimonti, D. and Ginsbourger, D. (2016). Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation. *Under revision. Preprint, hal-01289126*. (Cited on pages 73 and 114.)

Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J., and Richet, Y. (2016b). Adaptive Design of Experiments for Conservative Estimation of Excursion Sets. In *2016 SIAM Conference on Uncertainty Quantification*, Lausanne, Switzerland. (Cited on page 4.)

Azzimonti, D., Ginsbourger, D., Chevalier, C., and Richet, Y. (2015). Conservative estimates of excursion sets in reliability engineering. In *47èmes Journées de Statistique de la SFdS*. (Cited on pages 4 and 99.)

Baddeley, A. J. and Molchanov, I. S. (1998). Averaging of random sets based on their distance functions. *Journal of Mathematical Imaging and Vision*, 8(1):79–92. (Cited on pages 2, 3, 28, 29, 30, 45, 49, and 58.)

Baíllo, A., Cuevas, A., and Justel, A. (2000). Set estimation and nonparametric detection. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 28(4):765–782. (Cited on pages 7 and 35.)

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848. (Cited on page 137.)

Bayarri, M. J., Berger, J. O., Calder, E. S., Dalbey, K., Lunagomez, S., Patra, A. K., Pitman, E. B., Spiller, E. T., and Wolpert, R. L. (2009). Using statistical and computer models to quantify volcanic hazards. *Technometrics*, 51(4):402–413. (Cited on pages 1 and 43.)

Bect, J., Bachoc, F., and Ginsbourger, D. (2016). A supermartingale approach to Gaussian process based sequential design of experiments. Preprint at hal-01351088. (Cited on pages 108 and 137.)

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793. (Cited on pages 3, 37, 42, 43, 44, 56, 78, 100, 101, 102, and 136.)

Binois, M., Ginsbourger, D., and Roustant, O. (2015). Quantifying uncertainty on Pareto fronts with Gaussian process conditional simulations. *European Journal of Operational Research*, 243(2):386–394. (Cited on page 44.)

Bolin, D. and Lindgren, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):85–106. (Cited on pages 2, 31, 32, 39, 44, 74, 78, 99, 135, and 136.)

Botev, Z. I. (2017). The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:1–24. (Cited on pages 75 and 89.)

Bratley, P. and Fox, B. L. (1988). Algorithm 659: Implementing Sobol's quasirandom sequence generator. *ACM Transactions on Mathematical Software (TOMS)*, 14(1):88–100. (Cited on pages 41, 55, 57, and 79.)

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208. (Cited on page 56.)

Chen, V. C., Tsui, K.-L., Barton, R. R., and Meckesheimer, M. (2006). A review on design, modeling and applications of computer experiments. *IIE transactions*, 38(4):273–291. (Cited on page 41.)

Chevalier, C. (2013). *Fast uncertainty reduction strategies relying on Gaussian process models*. PhD thesis, University of Bern. (Cited on pages 37, 42, 76, 78, 106, 108, 111, and 136.)

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014a). Fast kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465. (Cited on pages 1, 43, 51, 55, 62, 100, 102, and 113.)

Chevalier, C. and Ginsbourger, D. (2014). Fast computation of the multi-points expected improvement with applications in batch selection. In *7th International Conference on Learning and Intelligent Optimization*, pages 59–69. Springer. (Cited on page 42.)

Chevalier, C., Ginsbourger, D., Bect, J., and Molchanov, I. (2013). Estimating and quantifying uncertainties on level sets using the Vorob'ev expectation and deviation with Gaussian process models. In Uciński, D., Atkinson, A., and Patan, C., editors, *mODa 10 Advances in Model-Oriented Design and Analysis*. Physica-Verlag HD. (Cited on pages 2, 36, 44, 106, and 108.)

Chevalier, C., Ginsbourger, D., and Emery, X. (2014b). Corrected kriging update formulae for batch-sequential data assimilation. In *Mathematics of Planet Earth*, Lecture Notes in Earth System Sciences, pages 119–122. Springer Berlin Heidelberg. (Cited on pages 55 and 111.)

Chevalier, C., Picheny, V., , and Ginsbourger, D. (2012). *KrigInv: Kriging-based Inversion for Deterministic and Noisy Computer Experiments*. R package version 1.3. (Cited on page 113.)

Chevalier, C., Picheny, V., and Ginsbourger, D. (2014c). The KrigInv package: An efficient and user-friendly R implementation of kriging-based inversion algorithms. *Computational Statistics & Data Analysis*, 71:1021–1034. (Cited on pages 38, 54, 78, and 114.)

Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty, Second Edition.* Wiley, New York. (Cited on pages 11, 12, 18, and 44.)

Cody, W. J. (1993). Algorithm 715: SPECFUN–a portable FORTRAN package of special function routines and test drivers. *ACM Transactions on Mathematical Software (TOMS)*, 19(1):22–30. (Cited on page 38.)

Cox, D. R. and Wermuth, N. (1991). A simple approximation for bivariate and trivariate normal integrals. *International Statistical Review*, 59(2):263–269. (Cited on page 75.)

Cressie, N. A. C. (1993). *Statistics for spatial data.* Wiley, New York. (Cited on page 50.)

Cuevas, A. and Fraiman, R. (2010). Set estimation. In Kendall, W. S. and Molchanov, I., editors, *New Perspectives in Stochastic Geometry*, pages 374–397. Oxford Univ. Press, Oxford. (Cited on page 44.)

Cuevas, A., González-Manteiga, W., and Rodríguez-Casal, A. (2006). Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19. (Cited on page 7.)

Devroye, L. and Wise, G. (1980). Detection of abnormal behaviour via non-parametric estimation of the support. *SIAM Journal on Applied Mathematics*, 38(3):480–488. (Cited on pages 7 and 35.)

Dick, J., Kuo, F. Y., and Sloan, I. H. (2013). High-dimensional integration: the quasi-Monte Carlo way. *Acta Numerica*, 22:133–288. (Cited on page 75.)

Dickmann, F. and Schweizer, N. (2016). Faster comparison of stopping times by nested conditional Monte Carlo. *Journal of Computational Finance*, pages 101–123. (Cited on page 82.)

Dubourg, V., Sudret, B., and Bourinet, J.-M. (2011). Reliability-based design optimization using kriging surrogates and subset simulation. *Structural and Multidisciplinary Optimization*, 44(5):673–690. (Cited on pages 37 and 43.)

Dudley, R. M. (2002). *Real analysis and probability.* Cambridge University Press. (Cited on page 11.)

Durrande, N. (2011). *Étude de classes de noyaux adaptées à la simplification et à l'interprétation des modèles d'approximation. Une approche fonctionnelle et probabiliste.* PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne. (Cited on page 14.)

Emery, X. (2009). The kriging update equations and their application to the selection of neighboring data. *Computational Geosciences*, 13(3):269–280. (Cited on page 111.)

Felzenszwalb, P. and Huttenlocher, D. (2004). Distance transforms of sampled functions. Technical report, Cornell University. (Cited on page 59.)

Fleuret, F. and Geman, D. (1999). Graded learning for object detection. In *Proceedings of the workshop on Statistical and Computational Theories of Vision of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR/SCTV).* (Cited on page 100.)

Franco, J. (2008). *Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes.* PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne. (Cited on pages 37 and 41.)

Franco, J., Dupuy, D., Roustant, O., Damblin, G., and Iooss, B. (2013). *DiceDesign: Designs of Computer Experiments.* R package version 1.3. (Cited on pages 114, 127, and 132.)

French, J. P. and Sain, S. R. (2013). Spatio-temporal exceedance locations and confidence regions. *The Annals of Applied Probability*, 7(3):1421–1449. (Cited on pages 1, 31, 32, 43, and 99.)

Gayraud, G. and Rousseau, J. (2007). Consistency results on nonparametric Bayesian estimation of level sets using spatial priors. *Test*, 16(1):90–108. (Cited on page 7.)

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1(2):141–149. (Cited on pages 55, 75, and 77.)

Genz, A. and Bretz, F. (2002). Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, 11(4):950–971. (Cited on page 75.)

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities.* Springer-Verlag. (Cited on pages 75, 77, 89, and 94.)

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Bornkamp, B., and Hothorn, T. (2012). *mvtnorm: Multivariate Normal and t Distributions.* R package version 0.9-9992. (Cited on page 76.)

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing science and statistics: Proceedings of the 23rd symposium on the interface*, pages 571–578. (Cited on page 89.)

Gibbs, M. N. (1998). *Bayesian Gaussian processes for regression and classification.* PhD thesis, University of Cambridge. (Cited on page 19.)

Ginsbourger, D., Bay, X., Roustant, O., and Carraro, L. (2012). Argumentwise invariant kernels for the approximation of invariant functions. In *Annales de la Faculté de Sciences de Toulouse*, volume 21, pages 501–527. (Cited on page 11.)

Ginsbourger, D. and Le Riche, R. (2010). Towards Gaussian process-based optimization with finite time horizon. In *mODa 9–Advances in Model-Oriented Design and Analysis*, volume 52, pages 89–96. Springer. (Cited on pages 100 and 101.)

Ginsbourger, D., Roustant, O., and Durrande, N. (2016). On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, 170:117–128. (Cited on page 11.)

González, J., Osborne, M., and Lawrence, N. D. (2016). GLASSES: Relieving the myopia of Bayesian optimisation. In *19th International Conference on Artificial Intelligence and Statistics*, pages 790–799. (Cited on page 101.)

Gramacy, R. and Polson, N. (2011). Particle learning of Gaussian process models for sequential design and optimization. *Journal of Computational and Graphical Statistics*, 20(1):102–118. (Cited on page 18.)

Gramacy, R. B. and Lee, H. K. H. (2009). Adaptive design and analysis of supercomputer experiments. *Technometrics*, 51(2):130–145. (Cited on pages 51 and 55.)

Hajivassiliou, V., McFadden, D., and Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of Econometrics*, 72(1):85–134. (Cited on pages 75 and 77.)

Hajivassiliou, V. A. and McFadden, D. L. (1998). The method of simulated scores for the estimation of LDV models. *Econometrica*, 66(4):863–896. (Cited on page 89.)

Hall, P. and Molchanov, I. (2003). Sequential methods for design-adaptive estimation of discontinuities in regression curves and surfaces. *Annals of Statistics*, 31(3):921–941. (Cited on page 44.)

Hamada, M., Martz, H. F., Reese, C. S., and Wilson, A. G. (2001). Finding near-optimal Bayesian experimental designs via genetic algorithms. *The American Statistician*, 55(3):175–181. (Cited on page 55.)

Hammersley, J. (1956). Conditional Monte Carlo. *Journal of the ACM (JACM)*, 3(2):73–76. (Cited on page 82.)

Hammersley, J. M. and Morton, K. W. (1956). A new Monte Carlo technique: antithetic variates. *Mathematical proceedings of the Cambridge philosophical society*, 52(03):449–475. (Cited on page 82.)

Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics*, 35(4):403–410. (Cited on pages 13 and 18.)

Hartigan, J. (1987). Estimation of a convex density contour in two dimensions. *Journal of the American Statistical Association*, 82(397):267–270. (Cited on page 7.)

Hensman, J., Fusi, N., and Lawrence, N. D. (2013). Gaussian processes for big data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*. (Cited on page 137.)

Idier, D., Rohmer, J., Bulteau, T., and Delvallée, E. (2013). Development of an inverse method for coastal risk management. *Natural Hazards and Earth System Sciences*, 13(4):999–1013. (Cited on page 117.)

Jankowski, H. and Stanberry, L. (2010). Expectations of random sets and their boundaries using oriented distance functions. *Journal of Mathematical Imaging and Vision*, 36(3):291–303. (Cited on pages 39 and 66.)

Jankowski, H. and Stanberry, L. (2012). Confidence regions for means of random sets using oriented distance functions. *Scandinavian Journal of Statistics*, 39(2):340–357. (Cited on pages 39 and 66.)

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2):131–148. (Cited on page 41.)

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492. (Cited on pages 42, 44, 50, 57, and 69.)

Joseph, V. R., Hung, Y., and Sudjianto, A. (2008). Blind kriging: A new method for developing metamodels. *Journal of mechanical design*, 130(3):031102. (Cited on page 18.)

Kahn, H. (1950). Random sampling (Monte Carlo) techniques in neutron attenuation problems–I. *Nucleonics*, 6(5):27–33. (Cited on page 82.)

Kahn, H. and Marshall, A. W. (1953). Methods of reducing sample size in Monte Carlo computations. *Journal of the Operations Research Society of America*, 1(5):263–278. (Cited on page 82.)

Katzfuss, M. (2016). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*, published on line. (Cited on page 137.)

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464. (Cited on page 19.)

Krige, D. G. (1951). A statistical method for mine variation problems in the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52:119–139. (Cited on page 17.)

Kudo, H. (1954). Dependent experiments and sufficient statistics. *Natural Science Report. Ochanomizu University*, 4:905–927. (Cited on page 23.)

Kuo, F. Y. and Sloan, I. H. (2005). Lifting the curse of dimensionality. *Notices of the American Mathematical Society*, 52(11):1320–1328. (Cited on page 75.)

Kushner, H. J. (1964). A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, 86(1):97–106. (Cited on page 78.)

Lantuéjoul, C. (2002). *Geostatistical simulation: models and algorithms*. Springer, Berlin. (Cited on page 44.)

Lemieux, C. (2009). *Monte Carlo and quasi-Monte Carlo sampling*. Springer. (Cited on pages 82 and 85.)

Lewis, T., Owens, R., and Baddeley, A. (1999). Averaging feature maps. *Pattern Recognition*, 32(9):1615–1630. (Cited on page 39.)

Marchant Matus, R., Ramos, F., and Sanner, S. (2014). Sequential Bayesian optimisation for spatial-temporal monitoring. In *30th Conference on Uncertainty in Artificial Intelligence (UAI 2014)*, Corvallis, Oregon, United States. AUAI Press. (Cited on page 101.)

Marmin, S., Chevalier, C., and Ginsbourger, D. (2016). Differentiating the Multipoint Expected Improvement for Optimal Batch Design. In *Machine Learning, Optimization, and Big Data*, volume 9432 of *Lecture Notes in Computer Science*, pages 37–48. (Cited on page 42.)

Matérn, B. (1960). *Spatial Variation*. PhD thesis, Stockholm University. (Cited on page 13.)

Matérn, B. (1986). *Spatial variation (second edition)*, volume 36 of *Lecture Notes in Statistics*. Springer Science & Business Media. (Cited on page 13.)

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58:1246–1266. (Cited on page 17.)

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245. (Cited on page 41.)

Mebane, W. R. J. and Sekhon, J. S. (2011). Genetic optimization using derivatives: The rgenoud package for R. *Journal of Statistical Software*, 42(11):1–26. (Cited on pages 55 and 113.)

Miwa, T., Hayter, A., and Kuriki, S. (2003). The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):223–234. (Cited on page 75.)

Mockus, J. (1989). *Bayesian Approach to Global Optimization. Theory and Applications*. Kluwer Academic Publisher, Dordrecht. (Cited on page 101.)

Mockus, J., Tiesis, V., and Zilinskas, A. (1978). The application of Bayesian methods for seeking the extremum. In Dixon, L. and Szego, E. G., editors, *Towards Global Optimization*, volume 2, pages 117–129. Elsevier. (Cited on page 42.)

Molchanov, I. (2005). *Theory of Random Sets*. Springer, London. (Cited on pages 2, 3, 21, 22, 23, 24, 25, 28, 36, 47, 48, 49, 103, and 108.)

Molchanov, I. S. (1998). A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics*, 25(1):235–242. (Cited on page 7.)

Moran, P. (1984). The Monte Carlo evaluation of orthant probabilities for multivariate normal distributions. *Australian Journal of Statistics*, 26(1):39–44. (Cited on page 75.)

Neal, R. M. (1998). Regression and classification using Gaussian process priors (with discussion). *Bayesian Statistics*, 6:475–501. (Cited on pages 18 and 19.)

Nichols, J. A. and Kuo, F. Y. (2014). Fast CBC construction of randomly shifted lattice rules achieving $O(n^{-1+\delta})$ convergence for unbounded integrands over R^s in weighted spaces with POD weights. *Journal of Complexity*, 30(4):444–468. (Cited on page 75.)

Oakley, J. (1999). *Bayesian Uncertainty Analysis for Complex Computer Codes*. PhD thesis, University of Sheffield. (Cited on page 50.)

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(1):1–42. (Cited on page 17.)

Osborne, M. A., Garnett, R., and Roberts, S. J. (2009). Gaussian processes for global optimization. In *3rd international conference on learning and intelligent optimization (LION3)*, pages 1–15. (Cited on page 101.)

Owen, D. B. (1956). Tables for computing bivariate normal probabilities. *The Annals of Mathematical Statistics*, 27(4):1075–1090. (Cited on page 75.)

Patel, J. K. and Read, C. B. (1996). *Handbook of the normal distribution*, volume 150. CRC Press. (Cited on page 9.)

Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., and Kim, N.-H. (2010). Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, 132(7):071008. (Cited on page 42.)

Polonik, W. (1995). Measuring mass concentration and estimating density contour clusters - an excess mass approach. *Annals of Statistics*, 23(3):855–881. (Cited on page 7.)

Pronzato, L. and Müller, W. G. (2012). Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3):681–701. (Cited on pages 37 and 41.)

Qian, P. Z. G. and Wu, C. F. J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experimentsierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2):192–204. (Cited on page 19.)

Qian, P. Z. G., Wu, H., and Wu, C. F. J. (2008). Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics*, 50(3):383–396. (Cited on page 18.)

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. (Cited on page 114.)

Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541. (Cited on pages 7, 42, 44, and 51.)

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. MIT Press. (Cited on pages 11, 13, 14, 44, and 79.)

Reitzner, M., Spodarev, E., Zaporozhets, D., et al. (2012). Set reconstruction by Voronoi cells. *Advances in Applied Probability*, 44(4):938–953. (Cited on page 44.)

Ridgway, J. (2016). Computation of Gaussian orthant probabilities in high dimension. *Statistics and Computing*, 26(4):899–916. (Cited on page 75.)

Robbins, H. (1945). On the measure of a random set. II. *The Annals of Mathematical Statistics*, 16(4):342–347. (Cited on page 25.)

Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer. (Cited on page 82.)

Rohmer, J. and Idier, D. (2012). A meta-modelling strategy to identify the critical offshore conditions for coastal flooding. *Natural Hazards and Earth System Sciences*, 12(9):2943–2955. (Cited on pages 1, 117, and 118.)

Roustant, O., Ginsbourger, D., and Deville, Y. (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodelling and optimization. *Journal of Statistical Software*, 51(1):1–55. (Cited on pages 15, 44, 54, 57, 62, and 114.)

Sacks, J. and Schiller, S. (1988). Spatial designs. In *Statistical decision theory and related topics IV*, volume 2, pages 385–399. Springer New York, NY. (Cited on page 55.)

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statististical Science*, 4(4):409–435. (Cited on pages 1, 17, 41, 50, 55, and 56.)

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York. (Cited on page 1.)

Schervish, M. J. (1984). Algorithm AS 195: Multivariate normal probabilities with error bound. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(1):81–94. (Cited on page 75.)

Scheuerer, M. (2009). *A comparison of models and methods for spatial interpolation in statistics and numerical analysis*. PhD thesis, University of Göttingen. (Cited on pages 11, 13, 14, and 69.)

Schonlau, M. (1997). *Computer Experiments and Global Optimization*. PhD thesis, University of Waterloo. (Cited on page 42.)

Seikel, M., Clarkson, C., and Smith, M. (2012). Reconstruction of dark energy and expansion dynamics using gaussian processes. *Journal of Cosmology and Astroparticle Physics*, 2012(06):036. (Cited on page 1.)

Simpson, T. W., Poplinski, J., Koch, P. N., and Allen, J. K. (2001). Metamodels for computer-based engineering design: survey and recommendations. *Engineering with computers*, 17(2):129–150. (Cited on page 17.)

Singer, I. M. and Thorpe, J. A. (2015). *Lecture notes on elementary topology and geometry*. Springer. (Cited on page 20.)

Snelson, E. and Ghahramani, Z. (2006). Sparse Gaussian processes using pseudo-inputs. In *Advances in Neural Information Processing Systems 18*. (Cited on page 137.)

Stein, M. L. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151. (Cited on page 57.)

Stein, M. L. (1999). *Interpolation of spatial data, some theory for kriging*. Springer, New York. (Cited on pages 12, 13, and 57.)

Stoyan, D. and Stoyan, H. (1994). *Fractals, random shapes and point fields: methods of geometrical statistics*. Wiley. (Cited on page 24.)

Tong, Y. L. (2012). *The multivariate normal distribution*. Springer. (Cited on pages 10 and 75.)

Vazquez, E. and Bect, J. (2010). Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. *Journal of Statistical Planning and Inference*, 140(11):3088–3095. (Cited on page 54.)

Vazquez, E. and Martinez, M. P. (2006). Estimation of the volume of an excursion set of a Gaussian process using intrinsic kriging. *arXiv preprint math/0611273*. (Cited on page 68.)

Villemonteix, J., Vazquez, E., and Walter, E. (2009). An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534. (Cited on page 44.)

Vorob'ev, O. Y. (1984). Srednemernoje modelirovanie (mean-measure modelling). *Nauka, Moscow*, In Russian. (Cited on pages 2, 24, 25, 36, 48, and 108.)

Wang, G. G. and Shan, S. (2007). Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical design*, 129(4):370–380. (Cited on page 17.)

Zhang, J., Craigmile, P. F., and Cressie, N. (2008). Loss function approaches to predict a spatial quantile and its exceedance region. *Technometrics*, 50(2):216–227. (Cited on page 39.)

Zhao, X., Stein, A., and Chen, X.-L. (2011). Monitoring the dynamics of wetland inundation by random sets on multi-temporal images. *Remote Sensing of Environment*, 115(9):2390–2401. (Cited on page 39.)

Declaration of consent

on the basis of Article 28 para. 2 of the RSL05 phil.-nat.

Name/First Name: Azzimonti Dario Filippo

Matriculation Number: 13-122-809

Study program: Ph.D. in Statistics

Bachelor ☐   Master ☐   Dissertation ◉

Title of the thesis: Contributions to Bayesian set estimation relying on random field priors

Supervisor: PD Dr. David Ginsbourger
Prof. Dr. Ilya Molchanov

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 para. 1 lit. r of the University Act of 5 September, 1996 is authorised to revoke the title awarded on the basis of this thesis. I allow herewith inspection in this thesis.

Place/Date

MARTIGNY 27/09/2016

Signature

Dario F. Azzimonti

Dario Filippo Azzimonti
LinkedIn • GitHub • dario.azzimonti@gmail.com

Work Experience

Idiap Research Institute   **Martigny, Switzerland**
Visiting Research Assistant   *September 2015 - Present*

- Uncertainty Quantification and Optimal Design group

IMSV, University of Bern   **Bern, Switzerland**
Ph.D. student   *September 2013 - Present*

- Research on Bayesian set estimation with Gaussian process priors.
- Teaching assistant for Statistical methods for Climate Science I and II.
- Contributed in statistical consultancy projects.

Sisal   **Milan, Italy**
Intern and Game Designer   *December 2012 - July 2013*

- Product & Business Innovation. Mathematical consultancy in the development of new games, development in Matlab of a new predictive tool for game revenues.

Mines ParisTech, Center of Mathematical Morphology   **Fontainebleau, France**
Research internship   *February 2012 - June 2012*

- Theoretical study on Boolean models and simulation study in Python.

Education

University of Bern   **Bern, Switzerland**
Ph.D. student in Statistics   *September 2013 - Present*

- Ph.D. supervised by PD Dr. David Ginsbourger and Prof. Ilya Molchanov.
- Thesis: "Contributions to Bayesian set estimation relying on random field priors".

Università degli Studi   **Milano, Italy**
ECMI Postgraduate Programme in Mathematics for Industry   *February 2013*

- Postgraduate certificate for MSc with prominent industrial problem solving.
- Participation in European Summer School in Industrial Mathematics 2011

Università degli Studi   **Milano, Italy**
MSc in Mathematics, ECMI program   *October 2010 - October 2012*

- Master thesis: "Effective Properties of Disordered Media: Applications to the Optical Properties of Automotive Paint", Thesis Advisors: Prof. V. Capasso and Prof. D. Jeulin. 110/110 cum laude.

Mines ParisTech   **Paris, France**
Visiting student   *September 2011 - June 2012*

- Erasmus visit in the framework of ECMI postgraduate program.

Università degli Studi   **Milano, Italy**
BA in Applied Mathematics   *September 2007 - October 2010*

- Bachelor thesis: "On the Theory of Fisher's Reproductive Value. A Mathematical Approach", Thesis Advisor: Prof. V. Capasso. 110/110.

Pre-prints and publications

D. F. Azzimonti, D. Ginsbourger, C. Chevalier, J. Bect, Y. Richet. Adaptive design of experiments for conservative estimation of excursion sets. Preprint at hal-01379642

D. F. Azzimonti, D. Ginsbourger. Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation. In review, 2016, preprint at hal-01289126.

D. F. Azzimonti, J. Bect, C. Chevalier, D. Ginsbourger. Quantifying uncertainties on excursion sets under a Gaussian random field prior. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):850–874, 2016.

D. F. Azzimonti, D. Ginsbourger, C. Chevalier, Y. Richet. Conservative estimates of excursion sets in reliability engineering. *Proceedings of 47 Journées de Statistique de la SFdS*, 2015.

D. F. Azzimonti, D. Jeulin, and F. Willot. Optical properties of deposit models for paints: full-fields FFT computations and representative volume element. *Journal of Modern Optics*, 60(7):519–528, 2013.

Teaching Duties

September 2016 - December 2016	University of Bern, Fall semester 2016, TA for *Linear Models and Regression I*. Lecturer: PD Dr. David Ginsbourger
September 2014 - June 2016	University of Bern, academic years 2014/2015 and 2015/2016, TA for *Statistical Methods for Climate Sciences I/II*. Lecturer: Dr. Michel Piot

---

Selected talks and presentations

September 2016	*Bayesian excursion set estimation with GPs*. Uncertainty Quantification Workshop, Sheffield, UK.
August 2016	*Estimating Orthant Probabilities of High Dimensional Gaussian Vectors with an Application to Set Estimation*. 12th International Conference on Monte Carlo and Quasi Monte Carlo methods in Scientific computing, Stanford University, USA.
August 2016	*Bayesian excursion set estimation under Gaussian random field priors*. Invited talk at UC Davis Statistics Summer Seminar, Davis, USA.
April 2016	*Bayesian estimation of excursion sets under Gaussian random field priors*. Invited talk at Ulm University. Ulm, Germany.
April 2016	*Adaptive Design of Experiments for Conservative Estimation of Excursion Sets*. SIAM UQ Conference. Lausanne, Switzerland.
March 2016	*A new method to compute conservative estimates for excursion sets of functions*. Ph.D. day presentation at GdR Mascot-Num Annual Meeting. Toulouse, France

November 2015   *Conservative estimate for the excursion set of a deterministic function under a Gaussian random field model.* Invited talk at the 1st PlanetS workshop on Bayesian Statistics. Bern, Switzerland.

August 2015   *Quantifying uncertainties on excursion sets under a Gaussian random field prior.* 8th International Congress on Industrial and Applied Mathematics. Beijing, China.

June 2015   *Bayesian set estimation with safety engineering application.* Short presentation within group talk Selected Topics in Uncertainty Quantification and Optimal Design at ETH. Zurich, Switzerland.

June 2015   *Conservative estimates of excursion sets in reliability engineering.* 47éme Journées de Statistique de la SFdS. Lille, France.

April 2015   *Simulating excursion sets under a GRF prior, how to choose simulation points?* Poster presentation at GdR Mascot-Num Annual Meeting. Saint-Étienne, France.

February 2015   *Conservative estimates of excursion sets.* ReDice consortium. IRSN research center, Cadarache, France.

November 2014   *Simulating excursion sets under a GRF prior, how to choose simulation points?* GRF-Sim workshop: Simulation of Gaussian and related Random Fields. Bern, Switzerland.

March 2014   *Workshop on MCMC: Reversible Jump algorithm.* Two hours tutorial. University of Neuchâtel, Switzerland.

October 2012   *Effective Properties of Disordered Media: Applications to the Optical Properties of Automotive Paints.* Master thesis presentation. Dep. of Mathematics, Università degli Studi, Milano, Italy.