Gaussian random field models for function approximation under structural priors and adaptive design of experiments

David Ginsbourger

Gaussian random field models for function approximation under structural priors and adaptive design of experiments

Habilitation thesis

by

David Ginsbourger

December 2013

Submitted to the faculty of science of the University of Bern

in partial fulfillment of the requirements for the Venia Docendi

Acknowledgments

This work would not have been possible without the support of the Institute of Mathematical Statistics and Actuarial Science of the University of Bern, where I have been offered the privilege of working as senior assistant since April 2010. Let me thank particularly thank my mentors Lutz Dümbgen and Ilya Molchanov for their trust and the scientific interactions I have had with them during these years. A warmly thank goes to all researchers, students, and staff members I have had the chance to know at the IMSV, with a special mention for the secretaries of the Institute Andrea Fraefel, Barbla Fraefel, Andrea Hitz and Pia Weingart, who have supported me on a number of occasions in project management and related duties, always in a good mood. Master and PhD students (but also senior members who will recognize themselves) also participate in the good ambiance, notably through devilish Töggele games or unhealthy lunch breaks.

Looking back to the articles appended to the present document and that have contributing to keep me active between my Ph.D. defense (in 2009) and today, I feel an urgent necessity to thank my closest research partners, and especially the PhD and master students I have had the pleasure of co-advising or supervising during these years. They include in the first place Nicolas Durrande (Ph.D. in Applied Mathematics, 2011) and Clément Chevalier (Ph.D. in Statistics, 2013). Some of our jointly authored papers constitute part of the research presented here (respectively in Chapters 1 and 2). They also include Bastien Rosspopoff (Ecole des Mines student, 2010) and Nicolas Lenz (Master in Mathematics, 2013), who have through their thesis directly or indirectly contributed to Chapter 3 (Aquifer modelling application) and Chapter 1 (Outlook section on invariances). Additionally, let me greet Mickaël Binois (PhD student in Applied Mathematics since January 2013), Dario Azzimonti (PhD student in Statistics since September 2013) and Robert Komara (Master student in Statistics) who have recently engaged in a thesis under my (co-)supervision.

A special thank goes to my colleagues and co-authors with whom I have interacted through the two main projects I have been involved in, ReDICE and ENSEMBLE, and also through further kinds of collaborations. Although of completely different natures, the ReDICE and ENSEMBLE projects have been motivating a number of the contributions presented here, be they applied or more theoretical. It would be really long to thank all ReDICE

partners individually (and I would take the risk of unfortunately forgetting some colleagues!) but I would like to stress that working with motivated colleagues from academia and industry has been an enriching experience, and a chance to think out of the box. A special thank to Clément Chevalier who has been coordinating the project with me, for better or for worse, and to Olivier Roustant and Yves Deville for regular "AlpesKriging" meetings in Saint-Etienne, Chambéry, and Bern. Let me finally stress that the success of such consortium relies in good part on the quality of the talks given at the plenary and satellite project meetings, so that all partners and students who have contributed should be thanked for their input.

The same applies to the ENSEMBLE project, where the excellent atmosphere seems to catalyse multidisciplinary work. I am really proud of being part of this team, and I thank everybody for bringing and sharing his own vision, and contributing to a collective piece of work. It is really impressive to see how, almost three years after the project kick-off, everyone has deepen his own approach while the common knowledge has developed, and the whole group has become a multi-task entity working towards shared challenges.

In a different context, let me thank my colleague Victor Picheny for the good continuing joint work over the years, Julien Bect who has visited us a couple of times in Bern for short or longer research stays, and all the other researchers I have had the pleasure of collaborating with, whether in collaborations leading to joint papers or code, or through diverse scientific discussions. In particular, I would like to thank the mODa community for having given me a chance to participate in the (successive) 2010 and 2013 conferences, the UCM community for having associated me to the 2012 and 2014 conferences in one way or the other, to the MascotNum network for keeping organizing high-quality events and having entrusted Bruno Sudret and me with the organization of the 2014 meeting, and to my colleagues at TU Dortmund, the Ecole des Mines de Saint-Etienne, and Toulouse Institute of Mathematics for numerous fruitful research stays and visits in both directions. Many thanks again to Victor Picheny and Olivier Roustant, who have kindly accepted to proofread the present document, and also to Thibault Espinasse for useful comments.

Let me finally thank my family and friends for continuous support despite my scandalous lack of availability in the last years, and my beloved wife for everything!

Contents

Introduction

Approximating, optimizing, and more generally studying multivariate functions under very limited evaluation budget are non-trivial problems that have to be tackled in a number of practical situations. A notable example is the case where a parametric complex system is modelled by a set of equations, and where the evaluation of the system response for any given instance of the input parameter is done by time-consuming numerical simulations. In mathematical terms, how to analyse a function $f : D \subset \mathbb{R}^d \to \mathbb{R}$ relying on a finite number of evaluation results $\{(\mathbf{x}_i, f(\mathbf{x}_i)), 1 \leq i \leq n\}$? Obviously, the problem is ill-posed without further hypotheses on f. While this problem is of deterministic nature, and there exist deterministic approaches to treat it, we are focusing here on stochastic approaches, where prior hypotheses on f are done trough probabilistic assumptions. Stochastic concepts are introduced to model epistemic uncertainty, namely that we do not know the values of f outside of $\mathbf{X} := \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. The work presented in this habilitation thesis essentially deals with Gaussian random field models, and to a lesser extent with their deterministic counterpart, reproducing kernel Hilbert spaces.

The document is structured around three chapters, each of them referring to selected publications included in appendix.

Chapter 1 mainly focuses on contributions related to (positive definite) kernels, and to the way "structural" prior information on f may be incorporated in Gaussian field modelling through the covariance kernel. Four accompanying papers (3 journal articles and a proceedings article) are sent in Appendix A:

- "Argumentwise invariant kernels for the approximation of invariant functions" Ginsbourger et al. (2012) (published in 2012 in **Annales de la Faculté de Sciences de Toulouse**)
- "Additive covariance kernels for high-dimensional Gaussian process

modeling" Durrande et al. (2012) (published in 2012 in **Annales de la Faculté de Sciences de Toulouse**)

- "ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis" Durrande et al. (2013) (published in 2013 in **Journal of Multivariate Analysis**)

- "Kernels and designs for modelling invariant functions: From group invariance to additivity" Ginsbourger et al. (2013b) (published in 2013 in **mODa 10 Advances in Model-Oriented Design and Analysis**)

The second and third articles are contributions realized in the framework of Nicolas Durrande's Ph.D. thesis (Ecole Nationale Supérieure des Mines de Saint-Etienne, 2008-2011), which I had the pleasure of officially co-advising from 2009 to 2011. While the first paper deals with the incorporation of invariances under group actions into Gaussian field modelling and related function approximation approaches (including kriging and conditional simulations), the second one deals with the incorporation of a prior hypothesis of additivity into them, and is intended to make a connection between an approach that has been successful in high-dimensional modelling (generalized additive models) and kriging. The third paper is about another kernel class that we have studied for "high-dimensional" Gaussian field modelling (which in this context means of the order of 10 to 100, i.e. would be moderate for scientists dealing with genomic data for instance), with an emphasis on model-based sensitivity analysis. The fourth paper is a conference article (presented in mODa 10) about ongoing work on invariant kernels (resumed in the recent preprint Ginsbourger et al. (2013e)), that makes a synthesis between results obtained in the first and second papers.

Chapter 2 focuses on contributions related to sequential evaluation strategies (also known as *adaptive design*) relying on Gaussian random field models, with applications in finite time optimization, noisy optimization, probability of failure estimation, and excursion set estimation. Four accompanying papers (two journal articles and two proceedings articles) are sent to Appendix B:

- "Towards Gaussian process-based optimization with finite time horizon" Ginsbourger and Le Riche (2010) (published in 2010 in **mODa 9 Advances in Model-Oriented Design and Analysis**)

- "Quantile-based optimization of noisy computer experiments with tunable precision (with discussion)" Picheny et al. (2013a) (published in 2013 in **Technometrics**)

- "Fast kriging-based stepwise uncertainty reduction with application to the identification of an excursion set" Chevalier et al. (2014a) (accepted to **Technometrics**)

- "Estimating and quantifying uncertainties on level sets using the Vorob'ev expectation and deviation with Gaussian process models" Chevalier et al. (2013a) (published in 2013 in **mODa 10 Advances in Model-Oriented Design and Analysis**)

The first article deals with finite time optimization based on Gaussian random field models, and discusses some limitations of the most popular infill sampling criterion, the so-called Expected Improvement criterion. It prepares the ground for the following papers, where we consider the future distribution of a Gaussian random field if a point is added. The second article is the fruit of a long-standing collaboration with Victor Picheny (now at INRA) and colleagues from the French Nuclear Safety and Radioprotection Institute (Yann Richet and Grégory Caplin, IRSN) on strategies to optimize costly-to-evaluate functions relying on noisy evaluations, in cases where the noise variance is tunable and the computational budget is limited. The third and fourth articles are contributions realized in the framework of Clément Chevalier's Ph.D. thesis, which I had the pleasure of supervising at the University of Bern from 2010 to 2013 (Ph.D. project funded by IRSN and the ReDICE Consortium). The third publication aims at providing fast and rigorous approaches for the Sequential Uncertainty Reduction (SUR) strategies proposed in Bect et al. (2012). The fourth one emphasizes on the use of the Vorob'ev expectation and deviation to quantify uncertainty in the context of Bayesian set estimation with Gaussian field models.

Chapter 3 focuses on contributions related to applications and software implementation, with a special focus on a new inversion method proposed in aquifer modelling, and on R packages developed along the years on Gaussian field regression and related sequential strategies. The chapter also contains some informal thoughts about interdisciplinary projects, as well as a number of references to application (and more theoretical) papers by other

researchers, that are meant to provide the reader with a few selected perspectives beyond the work of the author. Two accompanying papers (two journal articles) are sent to Appendix C:

- "Distance-based kriging relying on proxy simulations for inverse conditioning" Ginsbourger et al. (2013d) (published in 2013 in **Advances in Water Resources**)
- "DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodelling and optimization" Roustant et al. (2012) (published in 2012 in **Journal of Statistical Software**)

A global conclusion closes the main body of the article. Appendix E gives some complementary proofs regarding a piece of ongoing work mentioned at the end of Chapter 1 (Section 1.3). Besides, proofs of some properties given in Chapter 1 are recalled in Appendix D, where the notations of Ginsbourger et al. (2012) are slightly revisited in order to keep some homogeneity throughout the present document.

Finally, as required in attachment to any habilitation dossier at the University of Bern, a recent CV with copy of the Ph.D. diploma, a publication list, and a teaching statement are given in Appendix F.

Nomenclature

$(\Omega, \mathcal{F}, \mathbb{P})$	Probability triplet
D	Index set (typically a compact subset of $\mathbb{R}^d$ here)
$\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_n$	Points in D
$(Z_{\mathbf{x}})_{\mathbf{x} \in D}$, or simply Z	Random field indexed by D (defined over $(\Omega, \mathcal{F}, \mathbb{P})$)
$Z_{\mathbf{x}}$ or $Z(\mathbf{x})$	Value of Z at the point $\mathbf{x}$
$Z_{\mathbf{x}}(\omega)$ or $Z(\mathbf{x}; \omega)$	Realization of $Z_{\mathbf{x}}$ corresponding to the outcome $\omega \in \Omega$
$Z_{\cdot}(\omega)$ or $Z(\cdot; \omega)$	Path (also known as *trajectory* or *realization*) of Z corresponding to the outcome $\omega \in \Omega$
$f(\cdot)$ or $z(\cdot)$	Objective function $D \to \mathbb{R}$ (the notation z is often used when f is assumed to be some realization of a random field Z)
GPR	Gaussian Process Regression
GRF	Gaussian Random Field
RKHS	Reproducing Kernel Hilbert Space
$k(\cdot, \cdot)$	Positive (semi)-definite kernel on $D \times D$
$\mathbf{k}_n(\mathbf{x})$ or $\mathbf{k}(\mathbf{x})$	Covariance vector $(k(\mathbf{x}, \mathbf{x}_1), \ldots, k(\mathbf{x}, \mathbf{x}_n))'$
K	Covariance matrix $(k(\mathbf{x}_i, \mathbf{x}_j))_{i,j \in \{1,\ldots,n\}}$
$\mathbf{z}$	Response vector $(z_1, \ldots, z_n)' \in \mathbb{R}^n$

$m(\cdot)$ or $m_n(\cdot)$	Kriging mean function $D \to \mathbb{R}$ (n denotes the step in case of a sequential strategy)
$s^2(\cdot)$ or $s_n^2(\cdot)$	Kriging variance function $D \to \mathbb{R}$ (n denotes the step in case of a sequential strategy)
$(G, *)$ or simply G	A group (typically assumed finite here)
$\Phi(\cdot, \cdot)$	A measurable group action $G \times D \to D$
$g.\mathbf{x}$	Alternative notation for $\Phi(g, \mathbf{x})$ ($g \in G, \mathbf{x} \in D$)
T	A linear operator (function spaces and properties such as boundedness precised upon usage)
T_v	Composition operator ($\mathbb{R}^D \to \mathbb{R}^D$) with symbol $v : D \to D$
ν, ν_i	Measures on D, D_i
$\nu_1 \otimes \cdots \otimes \nu_d$	Tensor product measure over $D_1 \times \cdots \times D_d$
$L^2(D, \nu)$	Space of equivalence classes of square-integrable functions on the measured space (D, ν)
FANOVA	Functional ANalysis Of VAriance
$L_1^2(D_i, \nu_i) \overset{\perp}{\oplus} L_0^2(D_i, \nu_i)$	Direct orthogonal sum of $L_1^2(D_i, \nu_i)$ and $L_0^2(D_i, \nu_i)$
I	Set of indices (Typically, $I \subset \{1, \ldots, n\}$)
D_{-I}, ν_{-I}	$\prod_{i \notin I} D_i$, $\bigotimes_{i \notin I} \nu_i$
$\mathcal{H}_k$ or simply $\mathcal{H}$	RKHS with kernel k
$\bigodot_{i \in I} \mathbf{v}_i$ or $\bigodot_{i \in I} \Gamma_i$	Term-wise product between vectors or matrices
$\mathcal{B}$	A Banach space (typically, continuous functions $D \to \mathbb{R}$ equipped with the sup-norm)
$\mathcal{B}^*$	Topological dual of $\mathcal{B}$
μ, η	Gaussian measures on $\mathcal{B}$

C_μ	Covariance operator of μ ($\mathcal{B}^* \times \mathcal{B}^* \to \mathbb{R}$ or $\mathcal{B}^* \to \mathcal{B}^*$ depending on the context)
$\widehat{\mu}$	Fourier transform of the measure μ
$\ell_*\mu, T_*\mu$	Pushforward measures of μ by ℓ, T
T^*	Adjoint of the operator T
$\mathbb{E}$	Expectation symbol
$\mathrm{EI}(\cdot)$	Expected Improvement function
A_n	Event, such as $\{Z(\mathbf{x}_1) = z_1, \ldots, Z(\mathbf{x}_n) = z_n\}$, summarizing the n first evaluation results in a sequential strategy
$\mathrm{I}_n(\cdot)$	Improvement field at step n (of a sequential strategy)
$\mathrm{EI}_n(\cdot)$	Expected Improvement criterion at step n
$\mathrm{QI}_n(\cdot)$	Quantile Improvement field at step n
$\mathrm{EQI}_n(\cdot)$	Expected Quantile Improvement criterion at step n
$\mathbb{1}_{\{Z(\mathbf{x}) \geq t\}}$	Characteristic function of the event $\{Z(\mathbf{x}) \geq t\}$
$J_n^{(\alpha)}, J_n^{(\Gamma)}$	Two SUR sampling criteria related to the variance of the excursion volume

Chapter 1

Some new classes of kernels for spatial interpolation and Gaussian field regression

Positive definite[1] (p.d.) kernels play a central role in several contemporary functional approximation methods, ranging from regularization techniques within the theory of *Reproducing Kernel Hilbert Spaces* (RKHS, cf. Berlinet and Thomas-Agnan (2004)) to *Gaussian Process Regression* (GPR) in machine learning Rasmussen and Williams (2006). One of the reason for that is presumably the following particularly elegant predictor, common solution to approximation problems in both frameworks. Indeed, if scalar responses $\mathbf{z} := (z_1, \ldots, z_n) \in \mathbb{R}^n$ are observed for n instances $\mathbf{x}_1, \ldots, \mathbf{x}_n \in D$ of a d-dimensional input variable (D is here assumed to be a compact subset of $\mathbb{R}^d$), the function

$$m : \mathbf{x} \in D \longrightarrow m(\mathbf{x}) = \mathbf{k}(\mathbf{x})' K^{-1} \mathbf{z}, \tag{1.1}$$

is at the same time the best approximation of any function f in the RKHS of kernel k subject to $f(\mathbf{x}_i) = z_i$ $(1 \leq i \leq n)$, and the GPR ("simple kriging") predictor of any squared-integrable centred random field $(Z_{\mathbf{x}})_{\mathbf{x} \in D}$ of covariance kernel k subject to $Z_{\mathbf{x}_i} = z_i$ $(1 \leq i \leq n)$. Here $k : D \times D \longrightarrow \mathbb{R}$ stands for an arbitrary p.d. kernel, with $\mathbf{k}(\mathbf{x}) := (k(\mathbf{x}, \mathbf{x}_1), \ldots, k(\mathbf{x}, \mathbf{x}_n))$ and $K := (k(\mathbf{x}_i, \mathbf{x}_j))_{1, i \leq j \leq n}$ (assumed invertible here and in the sequel).

[1] Following Stein (1999), we use here the term p.d. for what some authors also call "non-negative definite" or "positive semi-definite".

In practical situations (e.g., when the z_i's stem from the output of an expensive-to-evaluate deterministic function, say $f : D \to \mathbb{R}$), the choice of k is generally far from being trivial. Unless there is a strong prior in favour of a specific kernel or parametric family of kernels, the usual *modus operandi* to choose k in GPR (when d is too high and/or n too low for a geostatistical variogram estimation Cressie (1993)) is to rely on well-known families of kernels, and to perform classical Maximum Likelihood Sweeting (1980); Mardia and Marshall (1984), cross-validation Dubrule (1983), or Bayesian inference of the underlying parameters based on data Omre (1987); Handcock and Stein (1993). For example, most GPR or kriging softwares offer various options for the underlying kernel, often restricted to stationary but anisotropic correlations like the generalized exponential or Matérn kernels, allowing the user to choose between different levels of regularity. This is in fact based on mathematical results concerning the link between the regularity of covariance kernels and the mean square properties of squared integrable random fields (or even a.s. properties, both in the case of Gaussian random fields and in more general settings, see for instance Cramér and Leadbetter (1967) or Scheuerer (2009, 2010) to get an overview).

A weak point of such an approach, however, is that not all phenomena can reasonably be approximated by stationary random fields, even with a well-chosen level of regularity and a successful estimation of the kernel parameters. In order to circumvent that limitation, several non-stationary approaches have been proposed in the recent literature, including convolution kernels (see Paciorek (2003) or Lee et al. (2005)), kernels incorporating non-linear transformations of the input space (Guttorp and Sampson, 1992; Anderes, 2005; Xiong et al., 2007), or treed Gaussian processes (Gramacy and Lee, 2008), to cite an excerpt of some of the most popular approaches.

In this chapter, we present classes of covariance kernels that allow incorporating so-called "structural" priors withing GPR and RKHS function approximation techniques, encompassing priors of invariance under group actions and priors of additivity (Cf. Section 1.1), a prior of centredness leading to a convenient type of ANOVA kernels for sensitivity analysis (cf. Section 1.2), and perspectives about further kinds of priors that may be embedded in GPR through the covariance kernel (cf. Section 1.3).

1.1 On group-invariant and additive kernels

Our intent here is to address a specific question related to the choice of k: assuming a known geometric or algebraic invariance of the phenomenon under study, is it possible to incorporate it directly in a kernel-based approximation method like GPR or RKHS regularization? More precisely, given a function z invariant under a measurable action Φ of some finite group G on D, is it possible to construct a metamodel of z respecting that invariance?

Here we investigate classes of kernels leading to a (meta)model[2] of f inheriting known invariances. In the particular case of a GRF interpretation, the proposed kernels enable a deeper embedding of the prescribed invariance in the metamodel since the obtained random fields have invariant paths (up to a modification). Note that the proposed approach is complementary to the non-stationary kernels evocated above, rather than in competition with them. Our main goals are indeed to understand to what extent kernel methods are compatible with invariance assumptions, what kind of kernels are suitable to model invariant functions, and how to construct such kernels based on existing (stationary or already non-stationary) kernels.

1.1.1 Argumentwise group-invariant kernels

Various classical properties of kernels such as stationarity or isotropy can in fact be seen as particular cases (with natural actions of groups of translations or isometries, respectively) of the following definition of invariance given in Parthasarathy and Schmidt (1972):

Definition 1.1.1. *k is said invariant under the action of G on D when*

$$\forall g \in G, \forall \mathbf{x}, \mathbf{x}' \in D,\ k(g.\mathbf{x}, g.\mathbf{x}') = k(\mathbf{x}, \mathbf{x}') \tag{1.2}$$

Here we consider a different notion of invariance:

Definition 1.1.2. *A kernel k is said argumentwise Φ-invariant when*

$$\forall g, g' \in G, \forall \mathbf{x}, \mathbf{x}' \in D,\ k(g.\mathbf{x}, g'.\mathbf{x}') = k(\mathbf{x}, \mathbf{x}') \tag{1.3}$$

[2] i.e., an approximation of f. Depending on the context (deterministic or stochastic modelling), the word metamodel may refer to m only or to the whole conditional distribution of the GRF Z knowing the observations $Z_{\mathbf{x}_i} = z_i$ $(1 \leq i \leq n)$.

One can notice that Eq. (1.2) corresponds to the particular case of Eq. (1.3) where $g = g'$. As discussed next, this second kind of kernels corresponds to much stronger invariance properties of the associated random fields.

PROPOSITION 1.1.3. *(Kernels characterizing invariant fields) Let G be a finite group acting measurably on D via Φ, and Z be a centred squared-integrable random field over D. Then Z has Φ-invariant paths (up to a modification) if and only if its covariance kernel is argumentwise Φ-invariant.*

See Ginsbourger et al. (2012) (In Appendix A) or Appendix D for a proof with notations adapted to those of the present document. The same holds for the two following propositions.
Assuming that we dispose of n noiseless observations $Z_{\mathbf{x}_i} = z_i$ $(1 \leq i \leq n)$ of a square-integrable centred random field $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ with covariance kernel k, we get furthermore:

PROPOSITION 1.1.4. *If k is argumentwise Φ-invariant, then the kriging mean m and the kriging variance s^2 are Φ-invariant. In particular, $\forall g \in G$, $m(g.\mathbf{x}_i) = z_i$ and $s^2(g.\mathbf{x}_i) = 0$ $(1 \leq i \leq n)$.*

In order to generalize to the conditional distribution of Z knowing $Z_{\mathbf{x}_i} = z_i$ $(1 \leq i \leq n)$, we start by looking at its conditional covariance:

$$\mathrm{Cov}(Z_{\mathbf{x}}, Z_{\mathbf{x}'}|Z_{\mathbf{X}} = \mathbf{z}) = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})'K^{-1}\mathbf{k}(\mathbf{x}'). \tag{1.4}$$

In the case where Z is assumed Gaussian, the simple kriging mean and variance at $\mathbf{x}$ coincide respectively with the conditional expectation and variance of $Z_{\mathbf{x}}$ knowing the observations. In addition, the Gaussian assumption makes it possible to get conditional simulations of Z, relying only on the conditional mean function and covariance kernel.

PROPOSITION 1.1.5. *(Properties of the conditional distribution of a Gaussian Random Field with argumentwise Φ-invariant kernel)*

1. *The conditional covariance is argumentwise Φ-invariant*

2. *Conditional simulations are Φ-invariant*

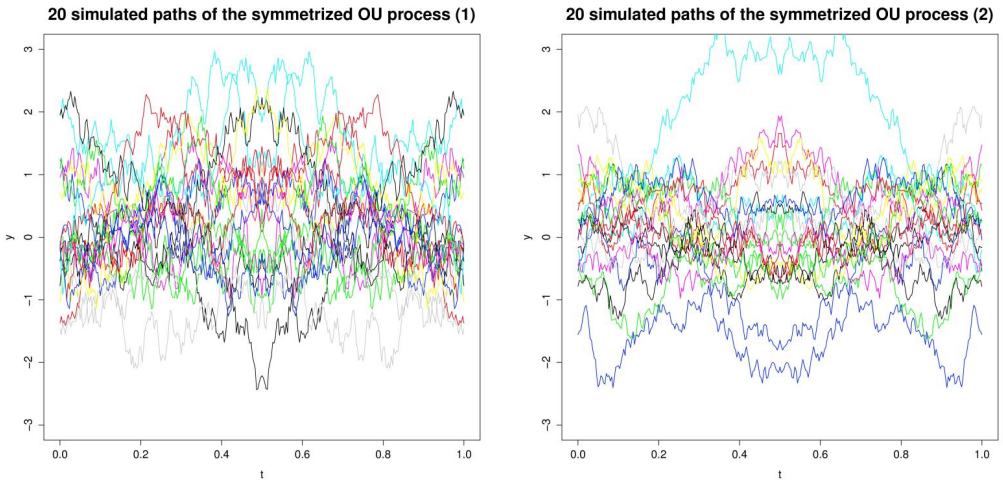

Figure 1.1: Symmetrization of the OU process relying on the kernels respectively defined by Eqs. 1.6 and 1.8. Left: by projection on a fundamental domain (Eq. 1.6). Right: by averaging over the orbits (Eq. 1.8).

Example 1.1.6. *Let us consider an Orstein-Uhlenbeck (OU) process* $(Y_x)_{x\in D}$ *restricted to* $D := [0,1]$, *and* $s : x \in D \longrightarrow 1-x \in D$ *the symmetry with respect to* $\frac{1}{2}$. *A first symmetrized OU process is obtained as follows*

$$Z_x^{(1)} = Y_{\min(x,s(x))} = Y_{\min(x,1-x)} \tag{1.5}$$

This centred Gaussian process is then characterized by the kernel

$$\begin{aligned} k_{Z^{(1)}}(x,x') &:= \mathrm{Cov}(Y_{\min(x,1-x)}, Y_{\min(x',1-x')}) \\ &= \exp\left(-|\min(x,1-x) - \min(x',1-x')|\right) \end{aligned} \tag{1.6}$$

On the other hand, a second symmetrized OU process is obtained by averaging over the orbits of the considered group action:

$$Z_x^{(2)} = \frac{1}{2}(Y_x + Y_{s(x)}) = \frac{1}{2}(Y_x + Y_{1-x}), \tag{1.7}$$

and possesses the following covariance kernel:

$$\begin{aligned} k_{Z^{(2)}}(x,x') &= \frac{1}{4}\mathrm{Cov}(Y_x + Y_{1-x}, Y_{x'} + Y_{1-x'}) \\ &= \frac{1}{2}\exp\left(-|x-x'|\right) + \frac{1}{2}\exp\left(-|1-x-x'|\right) \end{aligned} \tag{1.8}$$

Simulated paths of centred Gaussian processes with argumentwise invariant kernels respectively defined by Eqs 1.6 and 1.8 are represented on figure 1.1.

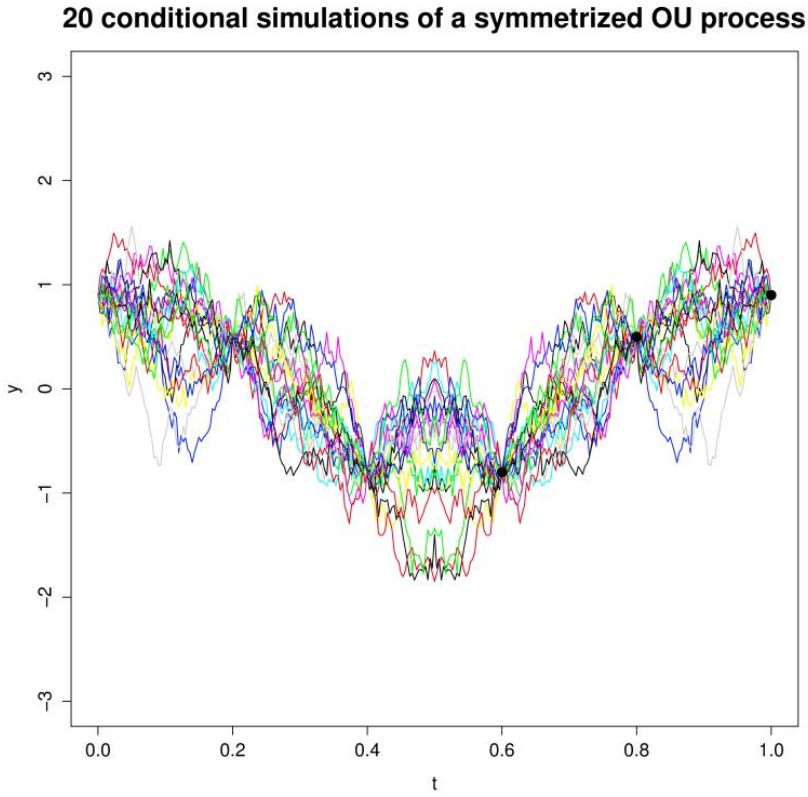

Figure 1.2: Conditional simulations of the symmetrized OU process with the kernel of Eq. 1.8. The black points stand for the conditioning data.

We now assume that the invariant process $Z^{(2)}$ was observed at the three points $x_1 = 0.6, x_2 = 0.8, x_3 = 1$, with response values $z_1 = -0.8, z_2 = 0.5, z_3 = 0.9$. The covariance kernel of Eq. 1.8 is used for performing simulations of $Z^{(2)}$ conditionally on the latter observations. Twenty such conditional simulations are represented on Figure 1.2. As can be seen on Figure 1.2, all paths are simultaneously Φ-invariant and interpolating the conditioning data, hence illustrating Property 1.1.5 on the conditional distribution of Gaussian random fields with argumentwise invariant kernel.

1.1.2 Additive kernels

After the work on Φ-invariant kernels presented in the previous section, I have been involved in research on additive kriging. This collaborative work was part of Nicolas Durrande's Ph.D. Durrande (2011) (co-advised in 2010 – 2012). Nicolas Durrande, Olivier Roustant (Co-author and also co-advisor of Nicolas), and I published this work (Durrande et al., 2012) in the same issue of the *Annales Scientifiques de la Faculté de Toulouse* where appeared (Ginsbourger et al., 2012).

The main motivation for studying additive kernels stemmed from the lack of suitability of usual kriging models (often relying on so-called *separable* or *product* kernels) for problems involving a relatively large number of variables (exceeding ten or so), and with a linear budget in the dimension. Having read about generalized additive models Hastie and Tibshirani (1995) and their success in high-dimensional smoothing (provided the phenomenon under study or a simple transformation of it is close to additive), it seemed sensible to us to investigate variants of kriging where usual kernels would be replaced by additive kernels such as defined below and in Durrande et al. (2012).

Before tackling the notion of additivity for kernels, let us define what we mean by additive in the setting of functions. For all $i \in \{1, \ldots, d\}$ $(d \geq 2)$, let $D_i \subset \mathbb{R}$ and $D := D_1 \times \cdots \times D_d$.

DEFINITION 1.1.7. *A function $f : D \to \mathbb{R}$ is additive whenever there exist d univariate functions $f_i : D_i \subset \mathbb{R} \to \mathbb{R}$ $(1 \leq i \leq d)$ such that*

$$\forall \mathbf{x} = (x_1, \ldots, x_d) \in D, \;\; f(\mathbf{x}) = \sum_{i=1}^{d} f_i(x_i). \tag{1.9}$$

What we called *additive kernels* in Durrande et al. (2012) can be seen as a direct extension of Definition 1.1.7 to p.d. kernels:

DEFINITION 1.1.8. *A p.d. kernel $k : D \times D \to \mathbb{R}$ is said additive whenever there exist d p.d. kernels $k_i : D_i \times D_i \to \mathbb{R}$ $(1 \leq i \leq d)$ such that*

$$\forall \mathbf{x}, \mathbf{x}' \in D, \;\; k(\mathbf{x}) = \sum_{i=1}^{d} k_i(x_i, x_i'). \tag{1.10}$$

Although not commonly encountered in practice, it is easily seen that such non-negative combination of p.d. kernels (p.d. kernels on $D_i \times D_i$ can always be seen as p.d. kernels on $D \times D$ – in the wide but not in the strict sense) is also a p.d. kernel Rasmussen and Williams (2006); Gaetan and Guyon (2009). Note that a sum of uncorrelated processes on the D_i's with respective covariance kernels k_i's admits such an additive kernel as covariance. We claim in Durrande et al. (2012) that for centred random fields, possessing an additive kernel reciprocally implies some kind of pathwise additivity:

PROPOSITION 1.1.9. *Any centred square-integrable random field with additive kernel is a modification of a random field with additive paths. In other words, for such a random field Z, there exists a random field A with additive paths such that* $\forall \mathbf{x} \in D,\ \mathrm{P}(Z_{\mathbf{x}} = A_{\mathbf{x}}) = 1$.

A direct proof of this property is given in appendix of Durrande et al. (2012) in the bi-dimensional case. For illustration, two realizations of a centred Gaussian random field with additive kernel considered in Nicolas Durrande's Ph.D. thesis are represented on Fig. 1.3 below.

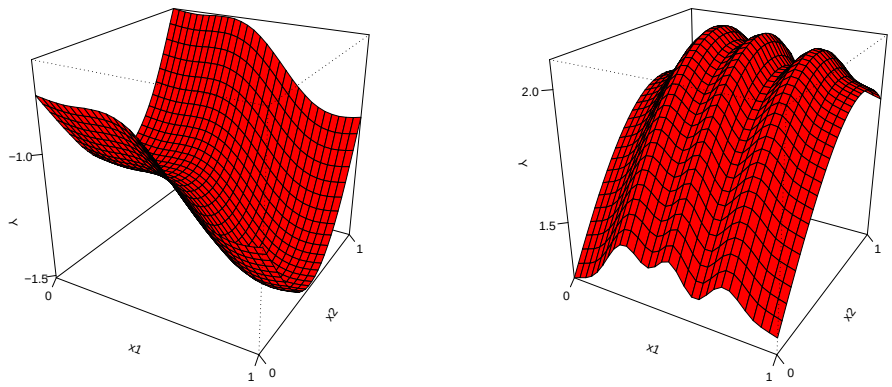

Figure 1.3: Two realizations of a centred Gaussian random field with additive kernel (in the sense of Def. 1.1.8).

We generalize Prop. 1.1.9 in the next section, where a necessary and sufficient condition on the covariance kernel is given for a centred square-integrable random field to have additive paths up to a modification.

Coming back to Durrande et al. (2012), let us finally remark that kriging with an additive kernel leads to an additive kriging mean. For example, we obtain in dimension two:

$$m(\mathbf{x}) = \mathbf{k_1}(x_1)'(\mathrm{K}_1 + \mathrm{K}_2)^{-1}\mathbf{z} + \mathbf{k_2}(x_2)'(\mathrm{K}_1 + \mathrm{K}_2)^{-1}\mathbf{z} \tag{1.11}$$

More will be said in the next sections on properties that can be shared between random field paths, covariance kernels, and kriging models.

1.1.3 A unifying framework: kernels invariant under a finite combination of composition operators

Following the characterizations of centred square-integrable random fields with group invariant or additive paths obtained respectively in Ginsbourger et al. (2012) and Durrande et al. (2012) (See Propositions 1.1.9 and 1.1.3 above), I felt that there should exist a more general framework in which group invariance and additivity would appear as particular cases. This framework turned out to be the one of combinations of composition operators, as summarized below. For more detail and proofs, see Ginsbourger et al. (2013b) (Appended in Chapter A).

Definition 1.1.10. *Let us consider an arbitrary function $v : x \in D \longrightarrow v(x) \in D$. The composition operator T_v with symbol v is defined as follows:*

$$T_v : f \in \mathbb{R}^D \longrightarrow T_v(f) := f \circ v \in \mathbb{R}^D.$$

Definition 1.1.11. *We call combination of composition operators with symbols $v_i \in D^D$ and weights $\alpha_i \in \mathbb{R}$ $(1 \leq i \leq q)$ the operator*

$$T := \sum_{i=1}^{q} \alpha_i T_{v_i}.$$

Proposition 1.1.12. *Let Z be a centred RF with kernel k. Then k is T-invariant, i.e.*

$$\forall \mathbf{x}' \in D,\ T(k(.,\mathbf{x}')) = k(.,\mathbf{x}') \tag{1.12}$$

if and only if Z equals $T(Z)$ up to a modification, i.e.

$$\forall \mathbf{x} \in D,\ \mathbb{P}\left(Z_{\mathbf{x}} = T(Z)_{\mathbf{x}}\right) = 1.$$

As pointed out by the following example, group invariances (under the action of a finite group) such as considered previously boil down to an invariance under a specific combination of composition operators. Prop. 1.1.3 hence appears as a particular case of Prop.1.1.12.

Example 1.1.13 (Case of group-invariance). $T(f)(\mathbf{x}) = \sum_{i=1}^{\#G} \frac{1}{\#G} f(v_i(\mathbf{x}))$ *with* $v_i(\mathbf{x}) := g_i.\mathbf{x}$ $(1 \leq i \leq \#G)$ *leads to* Z Φ*-invariant if and only if* k *is argumentwise invariant.*

Similarly, additivity can be characterized by a property of invariance under some combination of composition operators, as detailed below.

PROPOSITION 1.1.14. *Assuming* $\mathbf{a} \in D$*, a function* $f : D \to \mathbb{R}$ *is additive if and only if* f *is invariant under the following operator:*

$$\forall x \in D,\ T(f)(\mathbf{x}) = \sum_{i=1}^{d} f(v_i(\mathbf{x})) - (d-1) f(v_{d+1}(\mathbf{x})), \tag{1.13}$$

where $v_i(\mathbf{x}) := (a_1, \ldots, a_{i-1}, \underbrace{x_i}_{ith\ coordinate}, a_{i+1}, \ldots, a_d)$*, and* $v_{d+1}(\mathbf{x}) := \mathbf{a}$.

Applying Prop.1.1.12 in the case of the operator given in Prop. 1.1.14, we finally obtain a characterization of kernels leading to centred random fields with additive paths (up to a modification):

COROLLARY 1.1.15. *A centred squared-integrable random field* Z *with covariance kernel* k *has additive paths if and only if* k *is of the form*

$$k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{d} \sum_{j=1}^{d} k_{ij}(x_i, x'_j). \tag{1.14}$$

Hence, the kernels given by Def. 1.1.8 appear as a special case. The general model of covariance kernels leading to additive paths do in fact coincide with kernels of (tensor) sums of correlated univariate processes.

1.2 A new class of ANOVA kernels for kriging-based sensitivity analysis

In this section, we consider again some work done in the framework of Nicolas Durrande's Ph.D., that follow the previously presented results on additive

kernels. Instead of imposing a strong structural prior such as additivity, we considered a very flexible class of kernels referred to as "ANOVA kernels". The main goal was to construct a special class of ANOVA kernels for kriging-based global sensitivity analysis. The resulting class of so-called $K^{\star}_{\text{ANOVA}}$ kernels is presented at the end of this section (See also Durrande et al. (2013) for more detail).

In a nutshell, the purpose of global sensitivity analysis is to quantify the influence of each variable or group of variables on a function f Saltelli et al. (2008). In case $f \in L^2(D, \nu)$, where $D = D_1 \times \cdots \times D_d$ and $\nu = \nu_1 \otimes \cdots \otimes \nu_d$ is a tensor product of probability measures over D_i (corresponding to a hypothesis of independent inputs), a traditional approach in global sensitivity analysis is to calculate and study the following objects:

1. The FANOVA decomposition of f with respect to ν
2. The corresponding Sobol' indices

The first point leads to a decomposition of f as a sum of 2^d functions. Of course, attention may be restricted to those of the terms judged the most important. Sobol' indices precisely quantify the relative importance of each variable of group of variables on the response. In both cases, an analytical knowledge of f or a very large number of evaluations are necessary. Before explaining how $K^{\star}_{\text{ANOVA}}$ can help estimating FANOVA decompositions and Sobol' indices, let us recall some basics about these notions.

1.2.1 FANOVA decompositions and Sobol' indices

The starting point of FANOVA are the one-dimensional decompositions

$$L^2(D_i, \nu_i) = L^2_1(D_i, \nu_i) \overset{\perp}{\oplus} L^2_0(D_i, \nu_i), \tag{1.15}$$

where the $L^2_0(D_i, \nu_i)$'s are the subspaces of zero mean elements of the $L^2(D_i, \nu_i)$'s, i.e. $L^2_0(D_i, \nu_i) = \{g \in L^2(D_i, \nu_i), \int_{D_i} g(s)\mathrm{d}\nu_i(s) = 0\}$, and the $L^2_1(D_i, \nu_i)$'s are corresponding subspaces of almost everywhere constant elements. Using the tensor product structureof $L^2(D, \nu) = \bigotimes_{i=1}^{d} L^2(D_i, \nu_i)$, we then get

$$L^2(D, \nu) = \bigotimes_{i=1}^{d} \left(L^2_1(D_i, \nu_i) \overset{\perp}{\oplus} L^2_0(D_i, \nu_i) \right) = \overset{\perp}{\bigoplus_{B \in \{0,1\}^d}} L^2_B(D, \nu) \tag{1.16}$$

where $L^2_B(D,\nu) := \bigotimes_{i=1}^d L^2_{B_i}(D_i,\nu_i)$ for any $B \in \{0,1\}^d$. Given an arbitrary element $f \in L^2(D,\nu)$, the orthogonal projections of f onto those subspaces (See Durrande et al. (2013) for details) lead to the functional ANOVA representation Hoeffding (1948); Efron and Stein (1981); Antoniadis (1984); Sobol (2001) (or *Hoeffding-Sobol decomposition*) of f into main effects and interactions:

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{i<j} f_{i,j}(x_i,x_j) + \cdots + f_{1,\ldots,d}(\mathbf{x}). \tag{1.17}$$

This decomposition gives an insight on the influence of each set of variables $\mathbf{x}_I = \{x_i, i \in I\}$ on f. For the constant term, the main effects, and the two-factor interactions, one gets the classical expressions (Gu, 2002):

$$\begin{aligned} f_0 &= \int_D f(\mathbf{x})\mathrm{d}\nu(\mathbf{x}), \\ f_i(x_i) &= \int_{D_{-i}} f(\mathbf{x})\mathrm{d}\nu_{-i}(\mathbf{x}_{-i}) - f_0, \\ f_{i,j}(x_i,x_j) &= \int_{D_{-\{i,j\}}} f(\mathbf{x})\mathrm{d}\nu_{-\{i,j\}}(\mathbf{x}_{-\{i,j\}}) - f_i(x_i) - f_j(x_j) - f_0, \end{aligned} \tag{1.18}$$

with the notations $D_{-I} = \prod_{i\notin I} D_i$ and $\nu_{-I} = \bigotimes_{i\notin I}\nu_i$.

Considering now a random vector V with distribution ν, the $L^2(D,\nu)$ orthogonality between any two terms of the decomposition implies that the variance of the random variable $f(V)$ can be decomposed as

$$\mathrm{Var}(f(V)) = \sum_{i=1}^{d} \mathrm{Var}(f_i(V_i)) + \sum_{i<j} \mathrm{Var}(f_{i,j}(V_{i,j})) + \cdots + \mathrm{Var}(f_{1,\ldots,d}(V)). \tag{1.19}$$

The global sensitivity indices (also known as Sobol' indices) S_I are then defined as

$$S_I = \frac{\mathrm{Var}(f_I(V_I))}{\mathrm{Var}(f(V))} \quad (I \subset \{1,\ldots,d\}) \tag{1.20}$$

For any subset of indices I, S_I represents the proportion of variance of $f(V)$ explained by the interaction between the variables indexed by I. The knowledge of the indices S_I is very helpful for understanding the influence of the inputs on f, but the computation of the f_I's is cumbersome when the evaluation of f is costly since they rely on the computation of the integrals of

Eq. 1.18. Following Marrel et al. (2009), it can then be advantageous to perform the sensitivity analysis on a surrogate model m approximating f.

Let us now discuss some links between the FANOVA decompositions of f and m for the general case and some special case of ANOVA kernel.

1.2.2 ANOVA kernels

ANOVA kernels have been introduced in the literature of machine learning by Stitson et al. (1997) in the late 1990's. These kernels allow to control the order of interaction in a model and to enhance their interpretability Gunn and Kandola (2002); Berlinet and Thomas-Agnan (2004). For $D = D_1 \times \cdots \times D_d$, they are constructed as a product of univariate kernels $1 + k^i$, where 1 stands for a *bias term* and the k^i's are arbitrary p.d. kernels on $D_i \times D_i$ $(1 \leq i \leq d)$:

$$K_{ANOVA}(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{d}(1 + k^i(x_i, y_i)) = 1 + \sum_{I \subset \{1,\ldots,d\}} \prod_{i \in I} k^i(x_i, y_i). \tag{1.21}$$

As it appears on the right hand side of Eq. 1.21, ANOVA kernels can also be seen as a sum of separable kernels with increasing interaction orders Duvenaud et al. (2011).

Denoting by $\mathbb{1}^i$ and $\mathcal{H}^i$ the Reproducing Kernel Hilbert Space (RKHS) of functions defined over D_i with respective kernels 1 and k^i, K_{ANOVA} is the reproducing kernel of the space Aronszajn (1950); Berlinet and Thomas-Agnan (2004)

$$\mathcal{H}_{ANOVA} = \bigotimes_{i=1}^{d}(\mathbb{1}^i + \mathcal{H}^i) = \mathbb{1} + \sum_{I \subset \{1,\ldots,d\}} \mathcal{H}_I, \tag{1.22}$$

with $\mathcal{H}_I = \bigotimes_{i \in I} \mathcal{H}^i \otimes \bigotimes_{i \notin I} \mathbb{1}^i$.

Now, back to Eq. 1.21, the particular structure of K_{ANOVA} allows us to develop the covariance vector $\mathbf{k}(\mathbf{x})$ as $\mathbf{k}(\mathbf{x}) = \mathbf{1} + \sum_{I \subset \{1,\ldots,d\}} \bigodot_{i \in I} \mathbf{k}^i(x_i)$, where $\odot$ denotes a term-wise product (i.e. $\left(\bigodot_{i \in I} \mathbf{k}^i(x_i)\right)_j = \prod_{i \in I} k^i(x_i, x_j)$). From this relation, we can get the decomposition of the kriging mean m onto the subspaces $\mathcal{H}_I$:

$$m(\mathbf{x}) = \mathbf{1}^t \mathrm{K}^{-1}\mathbf{F} + \sum_{I \subset \{1,\ldots,d\}} \left(\bigodot_{i \in I} \mathbf{k}^i(x_i)\right)' \mathrm{K}^{-1}\mathbf{F} \tag{1.23}$$

Noting $m_0 = \mathbf{1}^t \mathrm{K}^{-1}\mathbf{F}$ and $m_I(\mathbf{x}) = \left(\bigodot_{i\in I} \mathbf{k}^i(x_i)\right)^t \mathrm{K}^{-1}\mathbf{F}$, we obtain an expression for m which looks similar to its FANOVA representation:

$$m(\mathbf{x}) = m_0 + \sum_{i=1}^{d} m_i(x_i) + \sum_{i<j} m_{i,j}(\mathbf{x}_{i,j}) + \cdots + m_{1,\ldots,d}(\mathbf{x}_{1,\ldots,d}). \tag{1.24}$$

In this expression, the m_I's have the nice feature of not requiring any recursive computation of integrals. However, Eq. 1.24 differs from the ANOVA decomposition of m since the properties of Eq. 1.17 are not respected. Indeed, the m_I's of Eq. 1.24 are not necessarily zero-mean functions and any two terms of the decomposition are generally not orthogonal in $L^2(D,\nu)$.

1.2.3 $K^*_{\mathbf{ANOVA}}$ kernels: construction and properties

An alternative to get round the previous issue is to consider RKHS such that for all i, $\mathcal{H}^i$ is L^2-orthogonal to the space of constant functions $\mathbb{1}^i$ (i.e. the $\mathcal{H}^i$ are spaces of zero mean functions for ν_i). This construction ensures that the decomposition of Eq. 1.23 has the properties required in Eq. 1.17, so that we benefit from the advantages of the two equations: the meaning of Eq. 1.17 for the analysis of variance and the easiness of computation of the m_I's from Eq. 1.23.

Let $\mathcal{H}$ be a RKHS of functions defined over a set $D \subset \mathbb{R}$ and ν a finite Borel measure over D. Furthermore, we consider the hypothesis:

H 1. *(i)* $k : D \times D \to \mathbb{R}$ *is* $\nu \otimes \nu$*-measurable,*
(ii) $\int_D \sqrt{k(s,s)}\mathrm{d}\nu(s) < \infty$.

Proposition 1.2.1. *Under H1,* $\mathcal{H}$ *can be decomposed as a sum of two orthogonal sub-RKHS,* $\mathcal{H} = \mathcal{H}_0 \overset{\perp}{\oplus} \mathcal{H}_1$*, where* $\mathcal{H}_0$ *is a RKHS of zero-mean functions for* ν*, and its orthogonal* $\mathcal{H}_1$ *is at most 1-dimensional.*

The reproducing kernels k_0, k_1 of $\mathcal{H}_0$ and $\mathcal{H}_1$ satisfy $k = k_0 + k_1$. Let π denote the orthogonal projection onto $\mathcal{H}_1$. Following Berlinet and Thomas-Agnan (2004) (theorem 11),

$$k_0(x,y) = k(x,y) - \frac{\int_D k(x,s)\mathrm{d}\nu(s) \int_D k(y,s)\mathrm{d}\nu(s)}{\iint_{D\times D} k(s,t)\mathrm{d}\nu(s)\mathrm{d}\nu(t)} \tag{1.25}$$

The $K^{\star}_{\text{ANOVA}}$ class is obtained by combining Eqs. 1.21 and 1.25:

$$K^{\star}_{\text{ANOVA}}(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{d}(1+k_0^i(x_i,y_i)) = 1 + \sum_{I\subset\{1,\dots,d\}}\prod_{i\in I} k_0^i(x_i,y_i). \tag{1.26}$$

Let us finally give two properties of $K^{\star}_{\text{ANOVA}}$ summarizing the computational benefits of using of such kernels for the FANOVA decomposition of the kriging mean and for the computation of global sensitivity indices.

PROPOSITION 1.2.2. *If m is a best predictor based on K^{*}_{ANOVA},*

$$m_I = \left(\bigodot_{i\in I}\mathbf{k}_0^i(x_i)^t\right)\mathrm{K}^{-1}\mathbf{F} \tag{1.27}$$

is the term of the functional ANOVA representation of m indexed by I. Hence, the decomposition of m given by Eq. 1.24 coincides with its functional ANOVA representation (Eq. 1.17).

PROPOSITION 1.2.3. *The sensitivity indices S_I of m are given by:*

$$S_I = \frac{\mathrm{Var}(m_I(V_I))}{\mathrm{Var}(m(V))} = \frac{\mathbf{F}^t\mathrm{K}^{-1}\left(\bigodot_{i\in I}\Gamma_i\right)\mathrm{K}^{-1}\mathbf{F}}{\mathbf{F}^t\mathrm{K}^{-1}\left(\bigodot_{i=1}^{d}(1_{n\times n}+\Gamma_i)-1_{n\times n}\right)\mathrm{K}^{-1}\mathbf{F}}, \tag{1.28}$$

where Γ_i is the $n\times n$ matrix $\Gamma_i = \int_{D_i}\mathbf{k}_0^i(x_i)\mathbf{k}_0^i(x_i)^t\mathrm{d}\nu_i(x_i)$ and $1_{k\times l}$ is the $k\times l$ matrix of ones.

1.3 Ongoing work and outlook

Current research directions include the extension of the invariance results obtained in Section 1.1.3 to a broader class of operators in the case of Gaussian random fields (GRF).

Let me briefly present a few unpublished ideas and intermediate results, that range from characterization of pathwise invariance through the Loève isometry to Gaussian measure interpretations and applications to the design of kernels leading to sparse GRF paths.

1.3.1 Further invariances in the GRF case

In the framework of a collaboration with Olivier Roustant and Nicolas Durrande, we have recently proposed an approach relying on the Loève isometry Berlinet and Thomas-Agnan (2004) in order to extend the invariance results to operators beyond the combinations of compositions. A first important point is to notice that given a linear operator $T : \mathbb{R}^D \to \mathbb{R}^D$ and a second-order random-field Z such that $T(Z)$ is second order, it is still possible to characterize pathwise invariances of Z by T relying on second-order properties of the joint process $(Z_{\mathbf{x}}, T(Z)_{\mathbf{x}})_{\mathbf{x}\in D}$, without any additional assumption concerning Z's probability distribution:

PROPOSITION 1.3.1. *$Z = T(Z)$ up to a modification if and only if*

$$k(\mathbf{x}, \mathbf{x}) = 2\,\mathrm{cov}(T(Z)_{\mathbf{x}}, Z_{\mathbf{x}}) - \mathrm{var}(T(Z)_{\mathbf{x}}) \qquad (\mathbf{x} \in D). \tag{1.29}$$

In the case of a combination of composition operators, we already saw that $\mathrm{cov}(T(Z)_{\mathbf{x}}, Z_{\mathbf{x}})$ simplifies to $T(k(\cdot, \mathbf{x}))(\mathbf{x})$. In more general cases, however, it is not straightforward how to deal with this term.

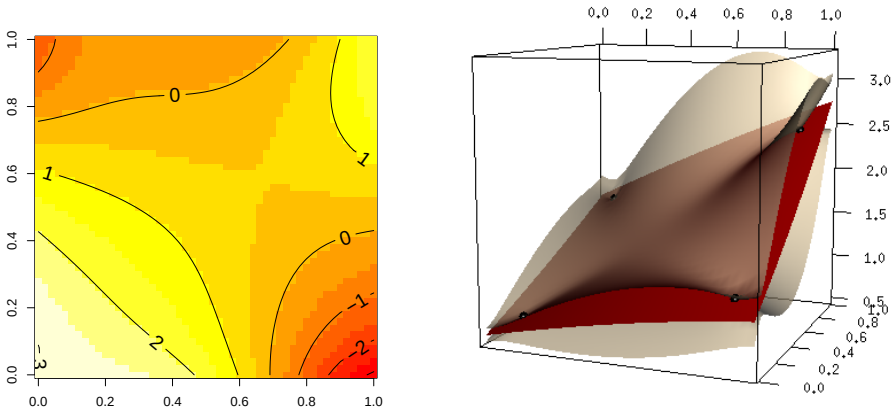

Figure 1.4: GRF Simulation and kriging relying on a bi-harmonic kernel.

The route followed in the preprint Ginsbourger et al. (2013e) is to restrict the scope to operators T such that $T(Z)_{\mathbf{x}} \in \mathcal{L}(Z)$ (the Hilbert space generated by the GRF Z), so that $\mathrm{cov}(T(Z)_{\mathbf{x}}, Z'_{\mathbf{x}})$ turns out to coincide with $\mathcal{T}(k(\cdot, \mathbf{x}'))(\mathbf{x})$ where $\mathcal{T} : \mathcal{H} \to \mathbb{R}^D$ is an restriction of T to the RKHS $\mathcal{H}$.

In such settings, the T-invariance of Z up to a modification may be characterized in terms of $\mathcal{T}$ being the identity of $\mathcal{H}$. Examples in Gaussian process regression are provided in (Ginsbourger et al., 2013e, Section 4), where Gaussian random field models with paths solution to ordinary equations are considered. An ongoing example of a GRF model with harmonic paths (relying on a bi-harmonic kernel recently introduced in Schaback (2009)) and its incoporation in kriging are illustrated on Fig. 1.4.

1.3.2 A Gaussian measure perspective

Following a question asked to me by Prof. Ilya Molchanov during an institute seminar, another theoretical set-up I am currently considering for extending the invariance results is the one of Gaussian measures on (separable) Banach spaces Kuo (1975); Ledoux and Talagrand (1991); Bogachev (1998).

Bridges between GRFs and Gaussian measures on function spaces have been studied at least since Rajput and Cambanis (1972), and it is known in particular that GRFs with continuous paths defined on a compact set D define Gaussian measures on the Banach space of continuous functions from D to $\mathbb{R}$ equipped with the uniform norm.

Starting from this, studying pathwise properties of Gaussian random fields (under technical conditions ensuring that the paths live in a prescribed Banach space, notably discussed in Scheuerer (2009)) may be done through measure-theoretic notions such as Fourier transforms and covariance operators of measures.

Compared to the approach of Section 1.1, where the obtained results were "up to a modification", working with covariance operators of Gaussian measures enables to directly get "almost sure" results. A result of that kind is given in Section E, where the argumentwise invariance (under the adjoint of T) of a centred Gaussian measure's covariance operator is shown to imply that the measure in question puts full mass on invariant elements (the reciprocal being immediate).

The proof scheme is certainly not the fastest, but I found it worth being displayed in appendix as it tackles the invariance by considering the product of the considered measure with its push-forward by T. Note that working directly with $(I-T)_*\mu$ would lead to a faster proof, as it is known that it is equivalent for a centred Gaussian measure to put full mass on null elements or to have a null covariance operator Tarieladze and Vakhania (2007). A paper summarizing these results and studying the link to Cameron-Martin spaces

(See Hairer (2009) for a tutorial, but also Van der Vaart and Van Zanten (2008) for a discussion on their relation to RKHS) is in preparation.

Besides, from a more applied point of view, relations between the FANOVA decomposition of GRF paths and the covariance operator are considered in current work. Preliminary results on the subject have been obtained in Nicolas Lenz's master thesis (Lenz, 2013, jointly advised with Dominic Schuhmacher), including in particular the explicit calculation of covariance structures leading to GRFs with paths orthogonal to additive functions (w.r.t. the usual L^2 scalar product). Perspectives include the extension of these preliminary results to the whole FANOVA decomposition, in the spirit of Kuo et al. (2010).

Chapter 2

Contributions to adaptive design strategies relying on random field models

2.1 Bayesian global optimization

Bayesian global optimization refers to algorithms and methods for optimizing expensive-to-evaluate functions using a Bayesian approach. The objective function is modelled relying on a prior distribution in some function space and on available evaluation results. Often, the function space and the corresponding prior distribution are implicitly defined by a Gaussian Random Field (GRF) assumption on f. In other words, $f : D \to \mathbb{R}$ is assumed to be one path of a GRF $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$, where Z is typically assumed to have a mean linear with respect to fixed basis functions, and a known covariance kernel. Of course, in applications, some covariance (hyper-)parameters are usually considered as unknown, and have to be estimated, e.g., by Maximum Likelihood, or inferred in a fully Bayesian way. Bayesian global optimization dates back to Kushner (1964); Mockus et al. (1978); Mockus (1989); Zilinskas (1992), but seems to have been strongly popularized in the last 15 years following the Efficient Global Algorithm (EGO) Jones et al. (1998), even though the Bayesian nature of approach is not much emphasized in the latter paper, where the constant mean and the parameters of a generalized exponential covariance kernel are estimated by Maximum Likelihood. Later on, Jones et al. (1998) has been revisited in many ways Forrester et al. (2006); Fra-

zier et al. (2008); Gramacy and Lee (2008); Osborne et al. (2009), and now appears as a seminal paper, especially concerning the use of the *Expected Improvement* (EI) criterion, the basic brick of EGO.

In this chapter, we first present some work related to the EI criterion (in Section 2.1.1) and a proposed extension for the case where observations are corrupted by a Gaussian noise with tunable variance (in Section 2.1.2). Then, Section 2.2 deals with criteria, originally inspired by EI but substantially departing from it, that aim at estimating a probability of failure, learning contour lines, and estimating excursion sets of a function above a fixed threshold, based on a random field prior. Finally, ongoing work and perspectives following Sections 2.1 and 2.2 are discussed in Section 2.3.

2.1.1 Variations on the Expected Improvement

In Bayesian optimization, it is common to sequentially enrich the current Design of Experiments (DoE) $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in D^n$ ($n \in \mathbb{N}^*$) ($n = n_0$ denotes the initial state and $\mathbf{X}_{n_0}$ the initial DoE) by evaluating the objective function at a point maximizing some so-called infill sampling criterion, update the GRF model, and iterate. EI is now one of the most popular infill sampling criteria in Bayesian optimization, and is defined (in minimization) as:

$$\mathrm{EI}_n(\mathbf{x}) = \mathbb{E}\left[(\min(Z(\mathbf{X}_n)) - Z(\mathbf{x}))^+ \,|Z(\mathbf{X}_n) = z(\mathbf{X}_n)\right] = \mathbb{E}\left[\mathrm{I}_n(\mathbf{x})|A_n\right] \quad (2.1)$$

where $\min(Z(\mathbf{X}_n)) := \min_{1\leq i\leq n}(Z(\mathbf{x}_i))$ is the best (i.e. minimal) response observed so far, $\mathrm{I}_n(\mathbf{x}) := \max(0, \min(Z(\mathbf{X}_n)) - Z(\mathbf{x}))$ is the improvement (which is itself a random field) beyond $\min(Z(\mathbf{X}_n))$, and the event $A_n := \{(Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)) = (z(\mathbf{x}_1), \ldots, z(\mathbf{x}_n))\}$ sums up all available observations at step n. EI is appreciated for trading off sampling intensification in promising regions (exploitation) and exploration of unvisited search space areas. Furthermore, EI is known in closed form Jones et al. (1998), which allows fast evaluations, and calculation of its derivatives (see calculations in Roustant et al. (2012), appended to the present document). Such a criterion, though updated by integrating new data, is typically considered at each iteration without structural change (with exception of Schonlau et al. (1998), where a generalized EI criterion with a varying exponent is considered). In fact, in EI algorithms, the point visited at the j^{th} iteration is determined by maximizing a conditional expectation (cf. Alg. 1, where unicity of EI's global maximizer at each step is assumed for simplicity).

Algorithm 1 EI algorithm with known parameters and $r \in \mathbb{N}^*$ iterations

1: **for** $j \leftarrow 1, \ldots, r$ **do**
2: Set $A_{n_0+j-1} = \{Z(\mathbf{x}_1) = z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_{n_0+j-1}) = z(\mathbf{x}_{n_0+j-1})\}$
3: Find $\mathbf{x}_{n_0+j} \in \arg\max_{\mathbf{x}\in D} \{\mathbb{E}\left[\mathrm{I}_{n_0+j-1}(\mathbf{x})|A_{n_0+j-1}\right]\}$
4: Evaluate z at $\mathbf{x}_{n_0+j}$
5: **end for**

In Ginsbourger and Le Riche (2010), we call *deterministic strategy with horizon* r ($r \in \mathbb{N}^*$) any finite sequence $\mathcal{S} = (s_j)_{j\in[1,r]}$ of measurable[1] functions $s_j(.) : (D \times \mathbb{R})^{n_0+j-1} \longrightarrow D$ ($j \in \{1, \ldots, r\}$), and denote by $\mathbb{S}_r$ the space of such strategies. For example, in Algorithm 1, the $s_j's$ are implicitly taken as $\arg\max_{\mathbf{x}\in D} \mathbb{E}[\mathrm{I}_{n_0+j-1}(\mathbf{x})|Z(\mathbf{X}_{j-1})]$ for all $j \in \{2, \ldots, r\}$, where $\mathbf{X}_{j-1} = \mathbf{X}_{n_0} \cup \{\mathbf{x}_{n_0+1}, \ldots, \mathbf{x}_{n_0+j-1}\}$ denotes the augmented DoE.

The purpose of this contribution is to consider more general strategies, where the $s_j's$ may be subject to structural changes at each iteration, and to show and illustrate that under the hypothesis of known finite horizon such strategies can beat Algorithm 1 in expectation (cf. the example on the decomposition of the EI of a horizon-2 strategy Ginsbourger and Le Riche (2010), and the main results given by Theorem 1 and Corollary 1). Corollary 1 is now recalled, with slightly adapted notations:

The optimal finite-time strategy satisfies the following recursion:

$$\begin{cases} s_r^*(\mathbf{X}_{n_0+r-1}, \mathbf{Z}_{r-1}) \in \arg\max_{\mathbf{x}\in D} \mathbb{E}[I_{n_0+r-1}(\mathbf{x})|A_{n_0+r-1}] \\ s_{r-1}^*(\mathbf{X}_{n_0+r-2}, \mathbf{Z}_{n_0+r-2}) \\ \quad \in \arg\max_{\mathbf{x}\in D} \mathbb{E}[I_{n_0+r-2}(\mathbf{x}, s_r^*(\mathbb{X}_{n_0+r-1}(\mathbf{x}), \mathbb{Z}_{n_0+r-1}(\mathbf{x})))|A_{n_0+r-2}], \\ \ldots, \\ s_1^*(\mathbf{X}_{n_0}, \mathbf{Z}_{n_0}) \\ \quad \in \arg\max_{\mathbf{x}\in D} \mathbb{E}[I_{n_0}(\mathbf{x}, s_2^*(\mathbb{X}_1(\mathbf{x}), \mathbb{Z}^1(\mathbf{x})), \ldots, s_r^*(\mathbb{X}_{n_0+r-1}(\mathbf{x}), \mathbb{Z}^{n_0+r-1}(\mathbf{x})))|A_{n_0}] \end{cases}$$

The finite time solution to the "strategic" EI maximization turns out to be recursive as recalled above, and as such with rather limited potential for applications (Benassi, 2013, Section 2.2.2). However, some of the modelling involved in this work prepared the ground for infill sampling criteria dedicated to further research questions, as detailed in the forthcoming sections.

[1] The source and goal spaces being equipped with their respective Borel σ-fields

2.1.2 Extensions to noisy optimization

In subsequent research articles such as Picheny et al. (2013a,b); Picheny and Ginsbourger (2014), a focus has been put on kriging-based strategies for noisy optimization problems. Such methods are needed when searching for the minimum of a function that can be only evaluated in noise, such as PDE solvers for neutronics based on Monte-Carlo Fernex et al. (2005). In a long-standing collaboration with a number of co-authors (See notably Picheny and Ginsbourger (2010); Ginsbourger et al. (2013c); Picheny et al. (2013a,b); Picheny and Ginsbourger (2014)), we have mainly considered the case where each evaluation is corrupted by a centred additive noise with variance known up to a constant, the variances being depending on the computational effort invested at each evaluation point. Typical assumptions write

$$\begin{aligned} &\forall i \in \{1,\ldots,n\},\ \widetilde{z}_i = z(\mathbf{x}_i) + \epsilon_i \quad (\varepsilon_i \sim \mathcal{N}(0,\tau_i^2) \text{ independently}) \\ &\widetilde{\mathbf{z}_n} := (\widetilde{z}_1,\ldots,\widetilde{z}_n) \qquad \Delta_n := \mathrm{diag}(\tau_1^2,\ldots,\tau_n^2) \quad (\tau_i^2\text{'s controllable}) \end{aligned}$$

Although adaptation of kriging to such framework turned out to be easily done (cf. for instance Picheny et al., 2013a), adapting the EI lead to less straightforward questions and developments. Here we put a tilde on events to stress the presence of noise:

$$\widetilde{A_n} := \{(Z(\mathbf{x}_1)+\varepsilon_1,\ldots,Z(\mathbf{x}_n)+\varepsilon_n) = \widetilde{\mathbf{z}_n}\}\,.$$

$\widetilde{A_n}$ represents the information at step n in a noisy optimization context, i.e. summarizes the set of noisy evaluation results that have been collected so far. Contrarily to what happened in the noiseless case, the distribution of $Z(\mathbf{x}_i)$ $(1 \leq i \leq n)$ conditionally on $\widetilde{A_n}$ is not necessarily a Dirac (provided for instance that $\tau_i \neq 0$). We sketch below how the introduction of noise makes it necessary to revisit the EI criterion. Plainly applying EI's definition, we find indeed in the noisy case

$$\mathrm{EI}_n(\mathbf{x}) = \mathbb{E}\left[\left(\underbrace{\min(Z(\mathbf{X}_n))}_{\text{unknown}} - \underbrace{Z(\mathbf{x})}_{\text{unreachable}}\right)^+ \middle| \widetilde{A_n}\right] \tag{2.2}$$

As indicated in the equation above, two major issues arise [2]:

[2] Assuming that the τ_i's are not all zero and the noise variance at $\mathbf{x}$ is positive

1. $\min_{\mathbf{x}\in\mathbf{X}_n}(Z(\mathbf{x}))$ is not known conditional on $\widetilde{A_n}$,

2. $Z(\mathbf{x})$ will remain non-exactly known after an evaluation at $\mathbf{x}$.

The first point prevents one from using the closed form formula known for EI in the noiseless case. Note that plugging in the minimum of noisy observations is only reasonable in cases where the noise variances are kept very small, as larger fluctuations caused by noise may lead to severe changes in the EI landscape. While EI may be estimated by conditional simulations (i.e. sampling from the joint distribution of $(\min(Z(\mathbf{X}_n)), Z(\mathbf{x}))$, this option is not really practical when it comes to optimizing the criterion (even though some clever approach relying on a stochastic gradient algorithm has recently been proposed for a similar issue in parallel optimization Frazier (2012)).

Besides (point 2), another difficulty in the noisy case is that the outcome of an evaluation at $\mathbf{x}$ will not be the value $z(\mathbf{x})$, as in the noiseless case, but rather $z(\mathbf{x}) + \epsilon$ with ϵ a realization of a noise variable with variance τ^2. Now, EI as defined above does not take into account the fact that a noisy observation is being performed; in particular, it does not depend on τ^2. In the context of Monte Carlo simulations, where the number of draws is controllable, it would seem more sensible to have an infill sampling criterion depending on the computational effort invested at $\mathbf{x}$. In Picheny and Ginsbourger (2010); Picheny et al. (2013a), we approached the problem by trying to transpose the notion of *current minimum* (and hence of improvement, which is a difference between minima before and after one or several evaluations) to the noisy case, as illustrated in Fig. 2.1.

A first question to be addressed to do so is of decision-theoretic nature: what does the term "improvement" mean when comparing two sets of noisy observations? What criterion should be used to judge that a set of noisy observations, or the associated kriging model, is better (in terms of minimization) after the $(n+1)^{\text{th}}$ measurement than before it? Relying only on the noisy observations $\widetilde{\mathbf{z}_n}$ and $\widetilde{z}_{n+1}$ is risky, as noise may introduce errors in the ranking. In Picheny and Ginsbourger (2010); Picheny et al. (2013a), we proposed to use the β-quantiles given by the kriging conditional distribution, for a given level $\beta \in [0.5, 1)$: a point is declared "best" over a set of candidates $\mathbf{X}_n$ whenever it has the lowest kriging β-quantile:

$$\mathbf{x}_* \in \arg\min_{\mathbf{x}\in\mathbf{X}_n} \left[q_n(\mathbf{x})\right] = \arg\min_{\mathbf{x}\in\mathbf{X}_n} \left[m_n(\mathbf{x}) + \Phi^{-1}(\beta) s_n(\mathbf{x})\right], \tag{2.3}$$

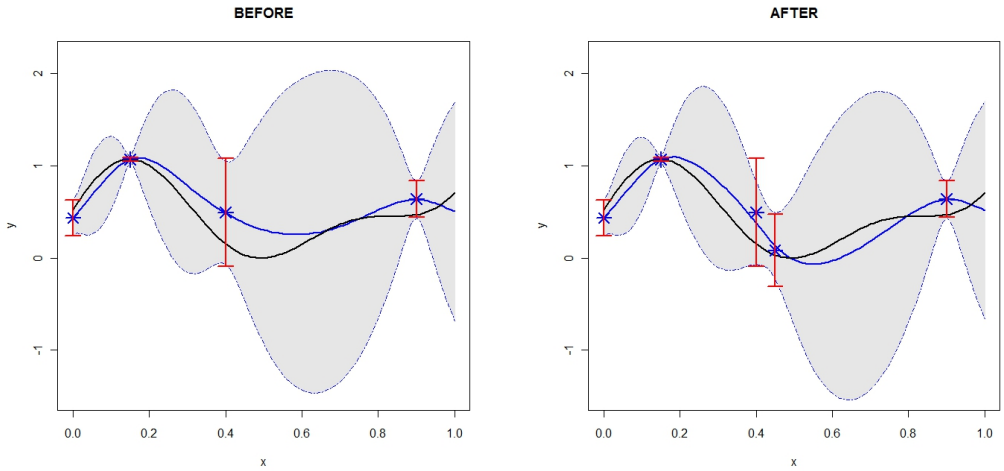

Figure 2.1: What does "improvement" mean when starting from a noisy kriging model and finishing with another one? Here a deterministic function (in black) is approximated using a Gaussian process regression model (mean in blue, pointwise prediction intervals in gray) based on noisy observations (blue stars, with uncertainty bars in red). On the right panel, a new noisy measurement is incorporated. Several options are possible to quantify to what extent the model of the right panel, containing more information, improves our knowledge about the objective function's global minimum/minimizer.

where Φ stands for the cumulative distribution function of the standard Gaussian distribution. From this, we defined a notion of quantile improvement[3] that is *consistent* with our decision criterion: Denoting by $Q_i(\mathbf{x})$ the kriging quantile $q_i(\mathbf{x})$ $(i \leq n+1)$ where the measurements are still in their random form, we define QI_n to be the decrease of the lowest kriging β-quantile between the present step n and the forthcoming step $n+1$:

$$\mathrm{QI}_n(\mathbf{x}_{n+1}) = \left(\min\left(Q_n\left(\mathbf{X}_n\right)\right) - Q_{n+1}\left(\mathbf{x}_{n+1}\right)\right)^+ \tag{2.4}$$

Of course, like in the noiseless case, this quantile improvement cannot be known in advance, because $Q_{n+1}\left(\mathbf{x}_{n+1}\right)$ depends on the future observation $Z(\mathbf{x}_{n+1}) + \varepsilon_{n+1}$. However, thanks to the linearity of the kriging predictor with respect to the observations, the future kriging β-quantile Q_{n+1} turns

[3]Note that QI is denoted by I and called *improvement* in Picheny et al. (2013a).

out to be a GRF (seen from step n), and its expectation and covariance can be analytically calculated based on the GRF model at step n.

Note that in our quantile improvement (2.4), we restrict attention to the observed points ($\mathbf{X}_n$ and $\mathbf{x}_{n+1}$), even though a similar criterion could be defined over the entire design space. However, such a restriction allows simplification, yielding a criterion in closed form.

The Expected Quantile Improvement (EQI) is defined as the following function of $\mathbf{x}_{n+1}$ and τ^2_{n+1}:

$$\mathrm{EQI}_n(\mathbf{x}_{n+1}, \tau^2_{n+1}) := \mathbb{E}\left[\left(\min_{i\leq n}(Q_n(\mathbf{x}_i)) - Q_{n+1}(\mathbf{x}_{n+1})\right)^+ \middle| \widetilde{A_n}\right], \tag{2.5}$$

where the dependence on the future noise τ^2_{n+1} appears through $Q_{n+1}(\mathbf{x})$'s distribution. The randomness of $Q_{n+1}(\mathbf{x})$ conditional on $\widetilde{A_n}$ is a consequence of $\widetilde{Z_{n+1}} := Z(\mathbf{x}_{n+1}) + \varepsilon_{n+1}$ having not been observed yet at step n. However, following the fact that $\widetilde{Z_{n+1}}|\widetilde{A_n}$ is Gaussian with known mean and variance, one can show that $Q_{n+1}(\cdot)$ is a GRF conditional on $\widetilde{A_n}$ (see proof and details in the appendix of Picheny et al. (2013a), itself appended to the present document). Furthermore, $\min_{i\leq n}(Q_n(\mathbf{x}_i))$ is known conditional on $\widetilde{A_n}$. As a result, the *EQI* criterion is analytically tractable, with a closed form formula given in (Picheny et al., 2013a, Eq. 10).

An original feature of EQI is that it allows defining strategies where both the evaluation point and the computational effort vary at each iteration. In Picheny et al. (2013a), such strategies are referred to as *online allocation algorithms*, and a heuristic is proposed for dynamically choosing the computational resource allocated to every successive noisy "measurement" (or function evaluation). The whole procedure is made practical by appealing to update formulas and aggregation of noisy measurements at the same point using so-called equivalent measurements (cf. Section 4.3 and Supplementary material of Picheny et al. (2013a) for details). Another interesting point, discussed in (Picheny et al., 2013a, Section 5), is that EQI gives a surprising and elegant way to address the finite time (or "finite computational budget") issue, by letting the parameter τ^2_{n+1} increase along a sequence of measurements (as the budget decreases), leading to a strategy being relatively more exploratory at the beginning and spending the last iterations locally refining the search in the vicinity of the "best" points found so far.

Experimental results reported in (Picheny et al., 2013a, Section 7) suggest good overall performances of the criterion and the proposed approaches, and show its applicability to a realistic test-case from nuclear criticality safety. Further results and comparisons in the "constant allocation" set up Picheny et al. (2013b) revealed that EQI is not always the best criterion (especially when compared to AEI Huang et al. (2006) or AKG Frazier et al. (2008)). However, the "online allocation" feature enabled by EQI makes it so far an unchallenged criterion in its category, to the best of our knowledge. A natural perspective for improving over existing approaches in noisy kriging-based optimization with constant and online allocation would be to follow the "Stepwise Uncertainty Reduction" approach (discussed in the next section), extending works such as Villemonteix et al. (2009) and working towards an enhanced applicability of this class of algorithms. From a more theoretical point of view, convergence properties and rates of convergence of kriging-based noisy optimization strategies owe to be studied, which may require new proof schemes compared to the results available in the noiseless case Vazquez and Bect (2010); Bull (2011); Srinivas et al. (2012).

2.2 On contour lines, probability of failure, and excursion set estimation

2.2.1 Contributions to the definition and the analysis of new infill sampling criteria

In recent research on kriging for computer experiments, several approaches have been proposed about adaptively sampling f for learning a contour line $\{\mathbf{x} \in D : f(\mathbf{x}) = t\}$, estimating a "volume of exceedance over t" (in both cases $t \in \mathbb{R}$ is a prescribed threshold) given some measure on D, and for related problems such as excursion set estimation.

The approach of Ranjan et al. (2008) for contour line estimation consists in sequentially evaluating f at points maximizing a variant of the expected improvement criterion, $EI(\mathbf{x}) = \mathbb{E}[(\alpha^2 s_n^2(\mathbf{x}) - (Z(\mathbf{x}) - t)^2)^+ \,|A_n]$, where $\alpha > 0$ is a tunable parameter. An alternative approach for tackling this problem, suggested in Picheny et al. (2010), consists in introducing a weight inside of the IMSE criterion (*Integrated Mean Squared Error*, cf. Sacks et al. (1989)), in order to favour regions with responses likely to be close to t. The obtained

"targeted IMSE" criterion, parametrized by $\epsilon > 0$, writes

$$\begin{aligned} \mathrm{IMSE}_{t,\epsilon} &= \mathbb{E}\left[\int_D \mathbf{1}_{Z(\mathbf{x})\in(t-\epsilon,t+\epsilon)} s^2_{n+1}(\mathbf{x})\mathrm{d}\mathbf{x}\middle| A_n\right] \\ &= \int_D \mathbb{P}(Z(\mathbf{x}) \in (t-\epsilon, t+\epsilon)|A_n) s^2_{n+1}(\mathbf{x})\mathrm{d}\mathbf{x}. \end{aligned} \tag{2.6}$$

The criterion obtained in the limit case where $\epsilon \to 0$ was also studied in Picheny et al. (2010). In Bect et al. (2012), we have summarized the strategies above and some other variants from the point of view of excursion volume estimation, and proposed a unified *Sequential Uncertainty Reduction* (SUR) framework for this specific problem. The principle of SUR strategies is to define an index of uncertainty on the quantity to estimate, and to iteratively evaluate Z (i.e. f, in practice) at a well-chosen point such that the expected uncertainty after evaluating Z at that point be minimized. The main index of variability considered in the latter article being the variance of the excursion volume, $\mathrm{Var}(\mathrm{Leb}_d(\Gamma))$ (where $\Gamma := \{\mathbf{x} \in D : Z(\mathbf{x}) \geq t\}$ and Leb_d is the Lebesgue measure on $\mathbb{R}^2$), the criterion for which several surrogate quantities are proposed is

$$J_n^{(\alpha)}(\mathbf{x}) := \mathbb{E}\left[\mathrm{Var}_{n+1}(\mathrm{Leb}_d(\Gamma); \mathbf{x}_{n+1} = \mathbf{x})|A_n\right], \tag{2.7}$$

where $\mathrm{Var}_{n+1}(\mathrm{Leb}_d(\Gamma); \mathbf{x}_{n+1} = \mathbf{x})$ is the variance of $\mathrm{Leb}_d(S)$ at "step $n+1$", assuming that $\mathbf{x}_{n+1} = \mathbf{x}$ but that $Z_{x_{n+1}}$ has not been observed yet (so that $\mathrm{Var}_{n+1}(\mathrm{Leb}_d(\Gamma); \mathbf{x}_{n+1} = \mathbf{x})$ is a $Z_{x_{n+1}}$-measurable random variable). The letter α in exponent symbolizes the fact that this criterion focuses on the volume of excursion, denoted by α in Bect et al. (2012).

As the criterion defined in Eq. 2.7 was considered intractable, an effort has been paid to find a similar criterion that would be computable. It turned out that the alternative criterion $J_n^{(\Gamma)}$ (taken from Bect et al. (2012) as well) recalled below is an upper bound of $J_n^{(\alpha)}$, so that a convergence of $J_n^{(\Gamma)}$ to zero guarantees that $J_n^{(\alpha)}$ converges to zero as well (both quantities being non-negative). The basic brick underlying $J_n^{(\Gamma)}$ is the binary field $\mathbb{1}_{\{Z(\mathbf{x})\geq t\}}$ together with the associated conditional mean and variance functions. In particular, the conditional expectation of $\mathbb{1}_{\{Z(\mathbf{x})\geq t\}}$ is the probability of excursion

$$p_n(x) := \mathbb{E}\left[\mathbb{1}_{\{Z(\mathbf{x})\geq t\}}|A_n\right] = \mathbb{P}(Z(\mathbf{x}) \geq t|A_n) = \Phi\left(\frac{m_n(\mathbf{x}) - t}{s_n(\mathbf{x})}\right).$$

Furthermore, $\mathbb{1}_{\{Z(\mathbf{x})\geq t\}}$ has conditional variance $p_n(\mathbf{x})(1-p_n(\mathbf{x}))$, so that $\int_D p_n(1-p_n)\mathrm{d}\nu$ can serve as a measure of global uncertainty about the target excursion volume, where ν is a finite measure (a probability, say) on D. The corresponding SUR sampling criterion writes

$$J_n^{(\Gamma)}(\mathbf{x}_{n+1}) := \mathbb{E}_n\left[\int_D p_{n+1}(1-p_{n+1})\mathrm{d}\nu \;\middle|\; X_{n+1}=\mathbf{x}_{n+1}\right], \tag{2.8}$$

where $\mathbb{E}_n$ refers to the conditional expectation knowing A_n and "$\big|\, X_{n+1}=\mathbf{x}_{n+1}$" is not really a probabilistic conditioning but rather means that the $(n+1)^{\text{th}}$ point in p_{n+1} is $\mathbf{x}_{n+1}$. This criterion has proven efficient on the test cases covered in (Bect et al., 2012, Sections 5.2 and 5.3), with comparable or better performances (in terms of excursion volume estimation) than the best performers among those taken from Picheny et al. (2010); Ranjan et al. (2008); Bichon et al. (2008); Echard et al. (2010). However, the considered SUR criteria appear to be more cumbersome than the other considered ones as they require integrating both over the input space and the range of the response value at the candidate point, while the targeted IMSE requires only a spatial integral, and the other criteria are "point-wise", so that they can be calculated based on explicit formulas involving solely the kriging mean and variance at the candidate point (see Bect et al. (2012), Appendix B for more detail). Besides, of all the considered criteria, none enables a batch-sequential approach, i.e. choosing points r by r ($r>1$) instead of one by one. Both limitations have been tackled during the Ph.D. of Clément Chevalier, resulting in novel formulas for SUR criteria and parallelizations thereof, presented in Chevalier et al. (2014a) and summarized in the next section.

2.2.2 Efficient calculations of SUR criteria

In Chevalier et al. (2014a) (appended to the present document), we have simultaneously speeded up the computation of $J_n^{(\Gamma)}$ (thanks to a decrease in computational complexity enabled by kriging update formulas presented in Emery (2009) and Chevalier et al. (2013b)) and produced a so-called "parallel" version of this criterion for distributed computing. The main result is recalled below:

Proposition 2.2.1.

$$J_n^{(\Gamma)}\left(\mathbf{x}_{n+1},\ldots,\mathbf{x}_{n+r}\right) = \int_D \Phi_2\left(\begin{bmatrix} a(\mathbf{x}) \\ -a(\mathbf{x}) \end{bmatrix}, \begin{bmatrix} c(\mathbf{x}) & 1-c(\mathbf{x}) \\ 1-c(\mathbf{x}) & c(\mathbf{x}) \end{bmatrix}\right)\nu(\mathrm{d}\mathbf{x}),$$

where:

- $\Phi_2(\cdot, M)$ *is the cumulative distribution function of the centred bivariate Gaussian with covariance matrix* M
- $a(x) := (m_n(\mathbf{x}) - t)/s_{n+r}(\mathbf{x})$,
- $\mathbf{b}(\mathbf{x}) := \frac{1}{s_{n+r}(\mathbf{x})}\boldsymbol{\Sigma}^{-1}\left(k_n(\mathbf{x}, \mathbf{x}_{n+1}), \ldots, k_n(\mathbf{x}, \mathbf{x}_{n+r})\right)'$ *where* k_n *is the kriging covariance kernel at step* n *(See Chevalier et al. (2014a) for detail),*
- $c(\mathbf{x}) := 1 + \mathbf{b}(\mathbf{x})'\boldsymbol{\Sigma}\mathbf{b}(\mathbf{x}) = s_n^2(\mathbf{x})/s_{n+r}^2(\mathbf{x})$,
- $\boldsymbol{\Sigma}$ *is the covariance matrix of* $(Z_{\mathbf{x}_{n+1}}, \ldots, Z_{\mathbf{x}_{n+r}})'$ *conditional on* A_n.

Proposition 2.2.1 lead to significant computational speed-ups, as discussed in Chevalier et al. (2014a). Additionally, numerical experiments presented in Section 4 of the latter article even suggested that there is almost no loss in parallelizing runs four by four compared to the purely sequential approach; In other terms, parallelizing leads to dividing the "wall-clock time" by a factor of almost four, at the cost of running four Central Processing Units (CPUs) instead of one. Future tests are needed to study whether such impressive speed-up continues to hold in further experimental conditions. However, it is already clear that the approach makes sense for a number of application settings where diminishing wall-clock time is more crucial than saving computational resource (e.g., in engineering applications where studies need to be handed in at a fixed date while computing units can always more easily be rented –e.g., through clouds– when necessary).

Another nice by-product of the calculations done in Chevalier et al. (2014a) is that the $J_n^{(\alpha)}$ criterion (considered so far intractable) finally turned out to be tractable. We obtained indeed the following formula, directly in the parallel set up:

Proposition 2.2.2.

$$J_n^{(\alpha)}(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+r}) =$$
$$\gamma_n - \int_{D\times D} \Phi_2\left(\begin{bmatrix} a(\mathbf{x}) \\ a(\mathbf{x}') \end{bmatrix}, \begin{bmatrix} c(\mathbf{x}) & d(\mathbf{x}, \mathbf{x}') \\ d(\mathbf{x}, \mathbf{x}') & c(\mathbf{x}') \end{bmatrix}\right) \nu(\mathrm{d}\mathbf{x})\nu(\mathrm{d}\mathbf{x}'),$$

where

- $\Phi_2, a, \mathbf{b}$, c *and* $\boldsymbol{\Sigma}$ *are defined as in Proposition 2.2.1,*

- $d(\mathbf{x}, \mathbf{x}') := \mathbf{b}(\mathbf{x})'\boldsymbol{\Sigma}\mathbf{b}(\mathbf{x}')$,
- γ_n *is a constant, in the sense that it does not depend on* $(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+r})$.

While the integrand (perhaps surprisingly) has the same complexity as the one of Proposition 2.2.1, it is noticeable that the integration domain now is a tensor product of the input domain by itself, hence making it more cumbersome to compute $J_n^{(\alpha)}$. Numerical results obtained in Chevalier et al. (2014a), Section 4.1 suggest that selecting points based on $J_n^{(\alpha)}$ instead of $J_n^{(\Gamma)}$ does indeed slightly improve performances in terms of volume of excursion estimation. Let us finally remark that both criteria are available in the KrigInv R package Chevalier et al. (2012, 2014b).

2.3 Ongoing work and perspectives

2.3.1 On Expected Improvement criteria and beyond

Coming back to infill sampling criteria for Bayesian global optimization, the idea of parallelizing EI has been considered and studied in Schonlau (1997), Taddy et al. (2009), and Ginsbourger et al. (2010b) where the so-called *Multipoint(s) Expected Improvement* (q-EI) could be estimated by Monte-Carlo but not calculated in closed form, making its maximization delicate. Recently, an approach based on a stochastic gradient algorithm was proposed in Frazier (2012), but to the best of the author's knowledge, the potential of this approach has not been investigated in practice yet. On the other hand, following classical but not so well-known results from the literature Tallis (1961) recently put to the fore by Da Veiga and Marrel (2012), we have proposed in Chevalier and Ginsbourger (2013) a closed form formula for q-EI relying on the Gaussian multivariate CDF, for which various efficient evaluation algorithms are available Genz and Bretz (2009); Genz et al. (2012).

In the continuation of parallel (also known as *batch-sequential*) kriging-based optimization and the work on the q-EI criterion, another question I have started to work on during my post-doctoral years (inspired by research questions that arose in a long-standing collaboration with Rodolphe Le Riche and his team) is the suitability of kriging to adapt to asynchronous parallel optimization problems. In such framework, a number of processors (say q) are available, but if calculations are run at q points, the responses are not

coming back altogether but in sequence, potentially with long pauses between successive results. Following an objective of wall-clock time reduction, it then makes sense to re-allocate new calculations to the free processors while the slowest ones are still running. The issue of choosing the new points is then not trivial: indeed, at the busy nodes, evaluations are running at known points but the corresponding responses are unknown. First attempts to tackle this issue by adapting the EI and q-EI criteria have been presented in project reports Ginsbourger et al. (2010a); Girdziusas et al. (2012) and in a conference paper Janusevskis et al. (2012), but there is still room for more in-depth research on the subject. While this may be a theme for future work, the newest results obtained in Chevalier and Ginsbourger (2013) may bring considerable simplifications, at least for cases where the number of CPUs (be they available or busy) remains moderate.

Other perspectives concerning EI criteria and related work include new approaches in multi-objective optimization , variants of EI for various specific problems such as curve (ridgelines, profile optima, filaments, etc.) estimation, or inversion with distance-based kriging Ginsbourger et al. (2013d) such as presented in Chapter 3. Concerning the two first points, let us finally mention two recent ongoing projects on the subject. On kriging-based multi-objective optimization (where a number of variants of EI already exist, see for example Emmerich et al. (2006)), let us mention the recent paper Binois et al. (2013) where a new approach relying on GRF and Pareto front simulations is proposed for monitoring the convergence of multi-objective EI algorithms. To finish with EI criteria and strategies, let us briefly mention Ginsbourger et al. (2013a), where a so-called "Profile-EI" is proposed for sequentially estimating curves of profile optima, i.e. how the optimum of a (say) bivariate function with respect to the second variable behaves as a function of the first variable.

2.3.2 Estimating excursion sets

Another topic of ongoing work is the estimation of sets (such as excursion sets) using GRF models and SUR strategies. This may seem redundant with what was already covered in Section 2.2. Actually, in the SUR strategies studied in Bect et al. (2012) and Chevalier et al. (2014a), the excursion set appears only through its volume, which makes it possible to speak about a variance in the proposed SUR criteria. However, even though the experimental results obtained on the test cases covered in these papers suggest that the

proposed strategies perform well in excursion set estimation, nothing guarantees *a priori* that reducing the variance of the volume of excursion will lead to accurately estimate the excursion set itself (Consider the pathological case of an excursion set with fixed volume but random position). This is the subject of Chevalier et al. (2013a) (appended to the present document), where concepts from random set theory (namely the Vorob'ev expectation and deviation) are considered in the framework of excursion set estimation with GRF priors. Ongoing work includes SUR strategies based on the Vorob'ev deviation (a first implementation is available in Chevalier et al. (2012)), and further comparisons of this new approach to existing ones.

Besides, the GRF approach to excursion set estimation may be enriched by considering further notions of random set expectations and variability Molchanov (2005). As most considered notions of expectations and variability seem analytically intractable, estimating them relying on GRF (conditional) simulations appears as a sensible option, which I am looking forward to work on in the framework of an ongoing collaboration with Julien Bect, Clément Chevalier, and Ilya Molchanov. In particular, original questions will arise such as the optimal choice of simulation points (to be distinguished from evaluation points) for an optimal estimation of candidate expectations and uncertainty measures based on stochastic simulation.

Finally, let me mention that some more theoretical questions such as consistency properties of the proposed approaches are considered as well in the recently granted "Bayesian Set Estimation" Swiss National Science Foundation project, notably through the work of Dario Azzimonti, who recently started his Ph.D. within this project, under joint supervision with Ilya Molchanov.

Chapter 3

Applications and software implementation

3.1 An overview of some applications

3.1.1 On applications, theories, and project life

Without any ambition of generality, these few lines are more intended to be taken as a personal story giving an insight on some phenomena going on at the interface between targeted research and fundamental research. I found being at the crossroad between two such worlds always exciting. Sometimes rewarding, sometimes difficult to sustain, but never boring!

During my postdoctoral years so far, I have had indeed the true privilege of sharing my time between a pure mathematics institute, a stochastic hydrogeology team in a hydrology and geothermics department, and a mathematical statistics institute. Besides, I have been actively involved in targeted research projects with several industry partners from various horizons, in a hybrid (targeted/fundamental) research project on aquifer modelling, and more recently on a fundamental research project in mathematical statistics.

It is impossible for me to summarize in a few lines the numerous interesting experiences that arose from this journey, but I would like to share a bit of it through some selected thoughts.

A first point, that may seem trivial but took me years to understand, is that a true diversity of profiles in research projects and institutions really is an asset. Of course, some deep abstract works may be the result of long-

standing individual research, and it is doubtlessly necessary that researchers with adequate profiles be sponsored on the long run to tackle hard theoretical issues (be it in sciences or in other fields). But, on the other hand, there are some kinds of achievements, especially when it comes to interface, interdisciplinary research, that can only be successfully addressed by a group of people with different backgrounds and visions, and last but not least with the ability to work together. This last point may sound easy but is actually not! The more I see researchers involved in such projects, the more I understand that interpersonal relations, informal brainstormings, and sharing knowledge beyond disciplinary boundaries truly are key ingredients.

Obviously, preserving and cultivating a solid anchor in one's home discipline(s) seems also crucial in order to keep an equilibrium and not completely scatter. I am very indebted to the institutes where I have been working during these years, and especially to the IMSV where I hold a senior assistant position (A kind of non-tenured assistant professor position with a main focus on research) as I write these lines, for having enabled me to sustain a number of external collaborations while benefiting from a home institute with strong theoretical standards and an institute life with statistics seminars and colloquia, high-level visitors and graduate students, etc. I am really striving after working in such a scientific environment, that enables me to keep making progresses in science and hard core academic research. Regularly chatting with experienced peers is a source of inspiration and challenges. Being confronted to the curious and fresh minds of students through teaching activities and master/PhD advising is another such source, of great quality.

To come back to project life and *clichés*, I would like to stress that if theory is immensely useful for applications, applications should not be seen only as a final product in the eco-system of research, but also as a chance for theories to be stimulated and rejuvenated. To give a concrete example, the theory of Reproducing Kernel Hilbert Spaces has started (to the best of my knowledge) with Aronszajn (1950). After Schwartz (1964) and others in the following years, the subject may have seemed to be already covered in full from a theoretical point of view. But who could predict at that time that so-called kernel methods would become central subject in statistics and computer science (See, e.g., Schölkopf and Smola, 2002) with numerous applications ranging from spatial statistics to computational biology (Schölkopf et al., 2004) and from engineering (Santner et al., 2003) to finance (Chalup and Mitschele, 2008) a few decades later? And even better: that the rise of

kernel methods would not only benefit from the foundations laid a generation ago, but would also inspire a number of new problems that would call for pure mathematics research, notably about positive definite kernels beyond the standard context of Euclidean spaces. Research topics on which recent advances have been done include (but are far from being limited to) isotropic p.d. kernels on spheres (Ziegel, 2013), p.d. kernels on measure spaces (Cuturi et al., 2005; Sriperumbudur, 2010), p.d. kernels on graphs (Espinasse, 2011), and p.d. operator valued kernels (Kadri et al., 2012). Future applications will call for further similarity measures between complicated objects, and it seems predictable that new mathematical advances will be inspired by the search for p.d. kernels on corresponding mathematical structures.

3.1.2 Distance-based kriging and the ProKSI algorithm

The ideas summarized now are the result of a long-standing collaboration with water scientists and applied mathematicians. It all began with discussions with Philippe Renard, when I realized that applying kriging-based optimization techniques to problems involving very high-dimensional parameter fields as input variables would not be straightforward. During the master's internship of Nicolas Durrande at Neuchâtel University (co-supervized by Philippe Renard and me), we started looking at the distance methods developed at Stanford by Céline Scheidt and Jef Caers (Scheidt and Caers, 2009) with the aim to apply them to a CO_2 sequestration problem. Following Nicolas' internship, we started thinking about incorporating the distance/kernel approach of Scheidt and Caers (2009) within a kriging model. By lack of time, the idea was temporarily abandoned. It is only at the start of the ENSEMBLE project (and notably through the implementation work done by Bastien Rosspopoff in summer 2011 during his master internship in Bern), that we had a chance to dig deeper and propose an original inversion method out of this distance-based kriging idea: the ProKSI algorithm (Ginsbourger et al., 2013d).

The main idea, following the foundations laid by Scheidt and Caers (2009) in the context of expensive flow simulations, is to endow a set of high-dimensional parameter fields with a pseudo-distance obtained by taking (usual) distances in the response space, by substituting the expensive code by quick-to-evaluate simplified flow simulations (where the simplification can be done, e.g., by neglecting diffusion/dispersion effects in the equations or by coarsen-

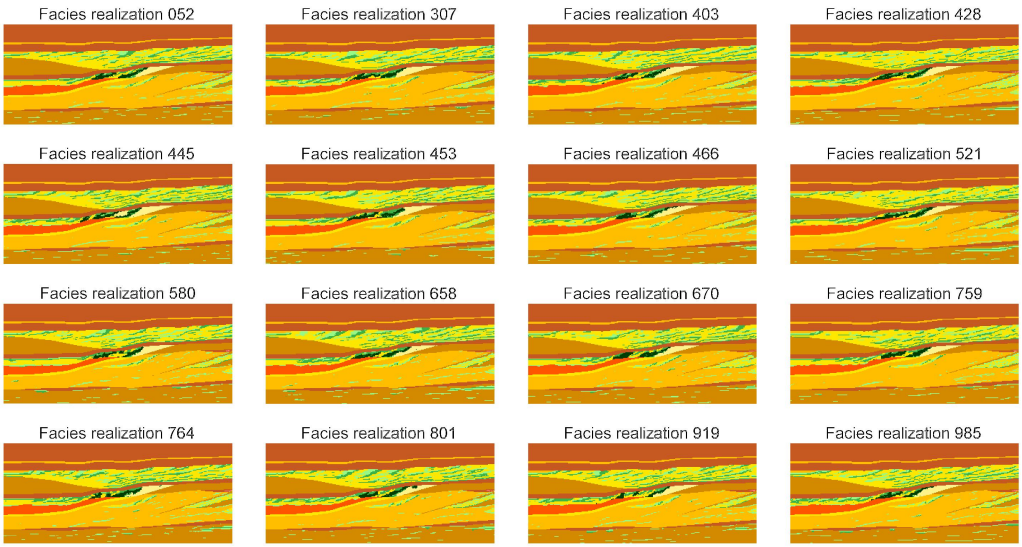

Figure 3.1: 16 among 1000 parameter fields generated by multipoints simulation, considered as candidates in the ProKSI algorithm of Ginsbourger et al. (2013d) [Article appended to the present document]

ing the time discretisation – See the appended article for more detail). Then, given N candidate parameter fields $\mathbf{x}_1, \ldots, \mathbf{x}_N \in E$ (cf. Fig. 3.1) and the associated *proxy* responses (i.e. breakthrough curves obtained from simplified flow simulations), a distance matrix can be formed, and transformed into a valid covariance matrix by applying a suitable kernel function entry-wise. Given a response of interest depending on the parameter field, the $N \times N$ covariance matrix obtained from the proxy distances is all we need to create a kriging model. In Ginsbourger et al. (2013d), the response of interest is the (transformed) misfit between a reference curve and the expensive-to-evaluate breakthrough curve $f_{\mathbf{x}}(\cdot)$ (cf. Fig. 3.2), and the kriging model is used to predict the misfit associated with $N - n$ remaining parameter fields when the actual misfit is known at n of them.

Two conditions are essential for the approach to be legitimate: That the proposed covariance matrix is indeed positive definite, and that the hypothesis that dissimilarities in the response of interest can be explained by dissimilarities in the proxy responses is tenable. Concerning the first point, the kind

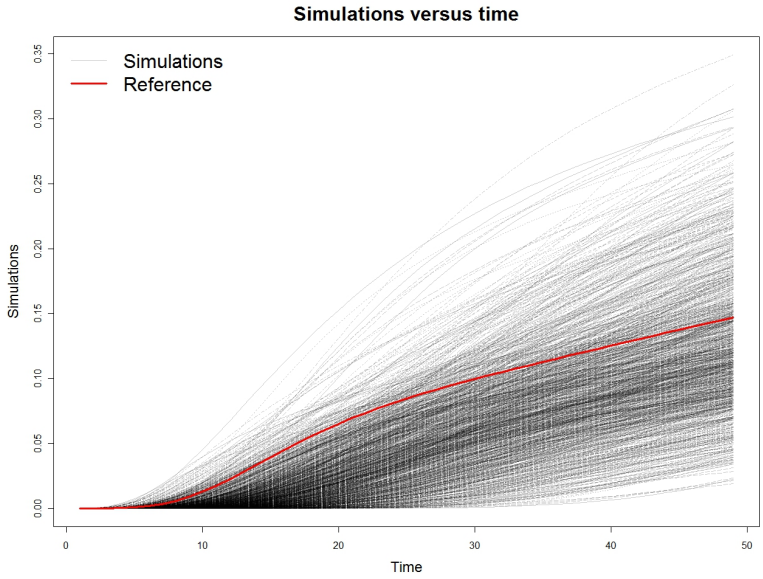

Figure 3.2: 1000 accurate breakthrough curves corresponding to the candidates parameter fields, one of them (in red) serving as reference. The goal of the ProKSI algorithm is to find, in a small number of sequential evaluations of the accurate solver, a subset of parameter fields leading to a breakthrough curve with a good fit to the reference.

of kernel used in the ProKSI article turned out to be admissible as composition of a change of reference space (each parameter field $\mathbf{x}$ being labelled by a discretized breakthrough curve $p_{\mathbf{x}}$) with a p.d. kernel (chosen Gaussian there, but it could have been any kernel from the $\mathcal{D}_\infty$ class of Stein (1999, p.44)) defined over the space of discretized breakthrough curves, as easily proven in the appended article. A nugget effect (with respect to the original space of parameter fields) was added, leading to a strictly p.d. kernel, and reflecting the intuition that two parameter fields with the exact same proxy response should not necessarily lead to identical misfits (calculated from the accurate breakthrough curves, not from the proxy). For the second point, the prediction results using this *ad hoc* kernel were surprisingly good, whether in cross-validation or in external validation. Besides, the tested Expected Improvement algorithms lead to competitive results for the two different kinds of proxy considered. Of course, the success of the method heavily relies on the use of informative proxys. More than the fidelity of the proxy curves

to the accurate curves, what seems important here is that dissimilarities in the proxy curves reflect dissimilarities in the response of interest (here the misfit). In particular, assuming that the dissimilarities in the accurate curves account well for the dissimilarities in the misfit, adding a bias to a high fidelity proxy would not affect the method as it is solely based on pairwise distance.

Perspectives include developing diagnostics for proxies (given a response of interest) and investigating automatic methods (in the framework of distance-based kriging) for extracting relevant dissimilarity measures from proxies. Of course, the choice of the kernel is another issue of interest, but it seems closer to classical statistical problems. An ideal approach would actually treat both issues at once. Let us finally remark that if the pseudo-distance matrix constructed in Ginsbourger et al. (2013d) is actually Euclidean (by construction), the approach can be generalized to an arbitrary similarity measure by first transforming the matrix into a Euclidean one, e.g., through multi-dimensional scaling (MDS). A sensible research question would then be to study how the prediction quality is affected by the non-Euclidean nature of the matrix, and how it depends on the dimension chosen within MDS.

3.1.3 Further applications

External collaborations, and notably with industrial partners through the ReDICE Consortium, have been a continuing source of inspiration for investigating statistical methods, adapting them to specific problems, and sometimes develop original approaches. Distance methods, for instance, have been used for a number of applications in psychology, biology, and beyond Hastie et al. (2001). Perhaps more surprisingly, such methods turned out to be useful in a mechanical engineering application encountered through ReDICE with colleagues from the French Nuclear Safety Institute (IRSN LIMAR, Cadarache), where macroscopic mechanical properties of an heterogeneous (two-phase) material are investigated in function of the position of inclusions in a homogeneous matrix (details are omitted to avoid any confidentiality issue). In that context, distances between materials are not straightforward to define, and several candidate distances are currently being considered for classifying materials. The ideal distance would reflect to what extent the mechanical behaviour of two materials will be similar, and would hence enable reducing a large set of randomly generated materials to a small sample of

representative ones. This is all very similar to the set up of the ProKSI algorithm presented above, except that the distance is not given a priori. Current work on this application includes the investigation of distances that respect the invariance of the mechanical behaviour by some prescribed symmetries, as well as the search of positive definite kernels based on such distances.

Other applications treated with ReDICE include car crash-test simulations. A vehicle being parametrized by a set of design parameters, and the responses of interest being the mass of the vehicle and an index characterizing the intrusion of the car body into the passenger compartment (both to be minimized), the goal is to explore the Pareto set and the Pareto front corresponding to this multi-objective problem. Kriging models (See also Bayarri et al., 2009, for related works in the same application field) are used as a base to sequential strategies using Expected Improvement criteria for model-based multi-objective optimization (Wagner et al., 2011; Emmerich et al., 2011; Parr, 2012). In ReDICE, this work is conducted in the framework of Mickaël Binois' Ph.D. thesis (funded by Renault).

A number of other ongoing applications with ReDICE partners, that are not presented here for the sake of brevity, will appear on the longer term first as project deliverables and then for some of them as peer-reviewed research articles, as in the case of Richet et al. (2013), to take a recent example.

Beyond the projects I am involved in, the class of techniques covered in the present document have been found useful in various application fields. To give a quick overview of this diversity, let us mention that Gaussian processes have been used for cosmic calibration (Habib et al., 2007) or exploring dark energy dynamics (Nair et al., 2013), for designing an aircraft wing (Sóbester et al., 2012), and for performing a functional analysis of variance analysis on climate models (Kaufman and Sain, 2010).

Climate sciences offer a number of potential applications for *ad hoc* variants of the methods discussed in the two last chapters. Following a first collaboration (together with Clément Chevalier) with Olivia Romppainen-Martius and Irina Mahlstein (Mahlstein et al., 2012), I am looking forward to contributing more to this field.

3.2 R packages

This section is about some R packages for kriging and kriging-based strategies that have been produced in relation to the methods and research results presented in the previous chapters. These packages are the outcome of years of collective work (not full time, of course!). They were originally created (first within the DICE Consortium concerning the DiceKriging and DiceOptim (Ginsbourger et al., 2013c) but also the DiceDesign (Franco et al., 2013), DiceEval (Dupuy and Helbert, 2013), and DiceView (Richet et al., 2012) packages) pursuing several overlapping goals:

- To dispose of a reasonably efficient prototype software gathering kriging and further methods for computer experiment applications,
- To participate in open source and reproducible research,
- And to facilitate the transfer between new approaches developed in research and practitioners from the industry or other

I have been mainly involved, together with Olivier Roustant and Yves Deville in the design and the development of DiceKriging and DiceOptim (later Victor Picheny, Tobias Wagner, and Clément Chevalier have also contributed to DiceOptim). The next section is dedicated to these packages and to a tutorial paper (appended) that was published in Journal of Statistical Software.

3.2.1 A quick tour of DiceKriging and DiceOptim

As detailed in the introduction of Roustant et al. (2012), DiceKriging complements existing R kriging programs as it allows kriging in arbitrary dimensions, offers universal kriging with arbitrary trend basis functions and a standard formula system, conditional simulations, and an enhanced likelihood maximization through a genetic algorithm (Mebane and Sekhon, 2011) using analytically calculated gradients. For the time being, the covariance kernels implemented in DiceKriging are anisotropic separable Matérn (with $\nu = 3/2$ or $5/2$), Gaussian, Exponential, and Power-exponential (with tunable exponent). Besides, short after the writing of Roustant et al. (2012), we have implemented a class of non-stationary kernels directly inspired from Xiong et al. (2007), that bases on a non-linear coordinate-wise input transformation ("scaling" option in `km`). Another option that has been recently

added to DiceKriging is to specify an arbitrary covariance kernel (stationary or not, that can be given as "kernel" functional input in `km`). Parameter estimation for arbitrary kernels is not proposed yet but under development for the newest (still at the experimental stage) DKlab package mentioned in the next section.

The appended paper Roustant et al. (2012) also presents the DiceOptim (Ginsbourger et al., 2013c) package, that was originally conceived as a satellite of DiceKriging for EI functions (cf. previous chapter) and their maximization. Again, one of the strengths of this package is that the analytical gradient of EI is coded (and intermediate calculations common to EI and its gradient are passed from one calculation to the other), making EI's maximization efficient (although some specific approaches potentially enabling further computational savings mentioned in Jones et al. (1998) or Franey et al. (2011) are not implemented at present). Since the version of DiceOptim presented in Roustant et al. (2012) noisy kriging-based optimization approaches have been added (thanks in particular to contributions of Victor Picheny, who has been the driving force behind the "noisy.optimizer" function). The corresponding criteria and their implementation are presented in the recent paper Picheny and Ginsbourger (2014). The code has notably been used for a numerical benchmark of kriging-based noisy optimization algorithms, resulting in Picheny et al. (2013b) (the collaboration with Tobias Wagner being a follow-up to numerous discussions in Dortmund and to the Noisy Kriging-based Optimization workshop –organized in Bern in 2010– where we started designing the benchmark with Victor Picheny).

3.2.2 Ongoing developments and perspectives

Started within a collaboration with the French Nuclear Safety Institute (IRSN), considerably extended with Victor Picheny, an finally revisited by Clément Chevalier (who is now first author and maintainer) during his Ph.D., the KrigInv package (Chevalier et al., 2012) is another satellite of DiceKriging for kriging-based probability of excursion and excursion set estimation. All the criteria, strategies, and methods presented in Chevalier et al. (2014a) and Chevalier et al. (2013a) are implemented in KrigInv, and ongoing work on kriging-based set estimation may be coded in R as well (within KrigInv or other).

As for the DiceOptim package, future developments include the addition

of a fast implementation of the multi-point EI criterion, such as proposed in Chevalier and Ginsbourger (2013). Additional work is needed for the efficient maximization of the multi-point EI criterion, with potential applications in synchronous and asynchronous parallel optimization (Janusevskis et al., 2012; Girdziusas et al., 2012). Another very promising direction is the use of Gaussian random field models for multi-objective optimization and Pareto front learning (Binois et al., 2013), that is planned to be developed in a separate dedicated package within the ReDICE Consortium.

Finally, a straightfoward perspective for extending the scope of kriging methods and surpassing what has been achieved with DiceKriging would be to create a package gathering a number of basis kernels (in the spirit of what has been done with the `kernlab` package (Karatzoglou et al., 2004)) and accepting user-defined kernels, with the possibility to combine all those kernels in different ways (tensor products and sums, weighted sums, block-wise operations, "centering" kernels as in $K^{\star}_{\mathrm{ANOVA}}$, symmetrizing and extracting invariant parts of a kernel, etc.) and to estimate underlying structures and parameters (as is already proposed, e.g., in the fanovaGraph approach of Muehlenstaedt et al. (2012); Fruth et al. (2013, 2014)). With Yves Deville and Olivier Roustant, we recently started to implement a new kriging package, simpler than DiceKriging in terms of functionalities, but with much more freedom concerning the classes of covariance kernels used. For the moment, the approach is to let the user enter a kernel (and optionally its gradient with respect to related parameters) in R or in C/C++, and to perform both Maximum Likelihood Estimation for the covariance parameters and kriging prediction with this arbitrary kernel. The working version of this new package is called `DKlab`, and medium term perspectives for it include the implementation of kernel classes such as the additive ones of Durrande et al. (2012) and the $K^{\star}_{\mathrm{ANOVA}}$ ones of Durrande et al. (2013). On the longer term, it would be ideal to have a sufficiently versatile package so that it can be used directly for research purposes, e.g. for testing new ideas of kernels on the fly.

Conclusion

Gaussian random field models have proven to be efficient for a number of approximation, optimization, and further classes of problems related to the study of real-valued functions (usually with moderate-dimensional source space) under severely limited evaluation budget.

The contributions presented in this habilitation thesis are structured around three complementary directions.

The first direction (Chapter 1) is the incorporation of "structural" prior information into Gaussian field models through their covariance kernel. The two first kinds of structural priors considered, namely the invariance under the action of a finite group and the additivity property, turned out to be interpretable as particular cases of invariances under combinations of composition operators. While the characterization results obtained in that setting apply in a general squared-integrable set-up, extensions to further operators in the Gaussian case (notably to differential operators) open exciting perspectives such as Gaussian field models incorporating, e.g., priors of divergence-freeness of curl-freeness Scheuerer and Schlather (2012) (in the case of a Gaussian random vector fields), of harmonicity Ginsbourger et al. (2013c), and beyond.

The second direction (Chapter 2) is the definition of new sequential policies and infill sampling criteria (still in the Gaussian field modelling set-up) for a variety of goals including deterministic global optimization, contour line estimation, noisy optimization, and excursion set estimation. Starting from a study on the *Expected Improvement* criterion, which is shown to be sub-optimal in finite time settings, other criteria such as the Expected Quantile Improvement for noisy optimization and further criteria for probability of excursion or set estimation are introduced and/or studied (notably those given in Bect et al., 2012), and the efficiency of related evaluation strategies is illustrated on toy functions and industrial test cases.

The third direction (Chapter 3) is the creation of generic tools for the

implementation of those approaches in various contexts, and also their application to large-scale problems, whether in terms of input (or output) dimensionality, or in terms of number of observations to handle, number of computational nodes to manage, etc.

Perspectives of future research for pushing further the limits of Gaussian field modelling and related sequential strategies abound, as the proposed solutions open new questions, but also because the kind of problems tackled here has already inspired contributions in a variety of neighbouring research fields in the past few years.

First, much has been recently done in the machine learning community that might be useful to transfer and adapt to other fields of applied statistics (whether machine learning should be seen as a field of applied statistics, computer science, or as an independent field is an open question left to the reader!). This includes use-inspired work on handling a large number of observations (Snelson and Ghahramani, 2006, and following works), enhanced hyper-parameter estimation techniques and McMC procedures for Bayesian inference (de Baar et al., 2013; Wang and Neal, 2013), but also theoretical work on the convergence of "bandit" optimization algorithms relying on pure random process and information theory (Srinivas et al., 2012; Contal and Vayatis, 2013)

Second, considering that (centred) Gaussian field models are one-to-one with positive definite kernels, investigating original families of p.d. kernels and related efficient estimation techniques offer promising avenues of research. On this question again, results from the machine learning community owe to be considered. In particular, p.d. kernels have been intensively studied in the framework of Support Vector Machines (cf. notably Genton, 2001; Schölkopf and Smola, 2002, and numerous more recent works basing on them), and Gaussian field modelling may benefit from a transfer of ideas that seems not completely achieved yet (even though some connections have already been made, e.g. in Archambeau and Bach (2011); Urtasun (2011); Duvenaud et al. (2011); Durrande et al. (2013)). Note however that on the subject of covariance kernels, the links to existing work is very far from restricting to machine learning, as positive definite kernels have been a broadly studied object across many fields of pure and applied mathematics (See Schoenberg, 1938; Berg et al., 1984; Sriperumbudur, 2010; Fasshauer, 2011; Cohen and Lifshits, 2012). Modelling problems involving complicated structures (not only arising from intellectual speculations but also from con-

crete applications) may catalyse future investigations about p.d. kernels, whether in terms of revisiting existing theories or in terms of posing new challenges to the mathematicians!

Keeping in mind that they are potentially compatible with any of the covariance kernels mentioned so far (and beyond), infill sampling criteria and sequential evaluation strategies relying on Gaussian random field models seem to have a high potential for extension and further developments. To take a first example, we have seen how to incorporate noise with tunable fidelity in kriging-based global optimization, but the "unit" noise variance was assumed constant over space. Finding realistic and efficient models to incoporate the estimation of a noise variance function into noisy kriging-based optimizers seems a reasonable and useful research perspective. On a different question, sequential strategies for Bayesian set estimation with a Gaussian field prior are just in their infancy, and quantities such as the Vorob'ev deviation proposed in Chevalier et al. (2013b) may be considered as "measures of uncertainty" in the Sequential Uncertainty Reduction framework of Bect et al. (2012), in order to define and calculate new infill sampling criteria for this specific problem. If the objective is reasonably ambitious, studying theoretical properties of such criteria and strategies is more to be seen as a long-term goal, culminating with sufficient conditions of consistency, and also ideally with convergence rates (cf. Bull, 2011). From a more applied point of view, tools are needed to describe and visualize distributions of sets in dimension larger than three, that would summarize their variability and help describing regions of high uncertainty.

In all, dimensionality and limitations on the number of observation points appear as the main bottlenecks of Gaussian field models and their use for optimization and related strategies. Approaches for solving one or both of those issues are needed for extending the scope of the considered methods, especially when it comes to the sequential framework. An original approach for high-dimensional global optimization based on Gaussian field models was recently proposed in Wang et al. (2013), which may open new research perspectives also for the other goals usually pursued with this class of models.

Efficient and versatile (and ideally open source) implementations of the newest Gaussian field modelling techniques will in my view play a significant role, not only for spreading such techniques to practitioners, but also for research purposes. A promising perspective on that topic would be to dispose of Gaussian random field software allowing the user to define new classes of covariance kernels and yet to benefit from fast parameter estimation algo-

rithms. Environments such as Rcpp (Eddelbuettel and Francois, 2011), an R package for encapsulating C++ code into R codes and packages, may be key to achieve such goals.

After so many perspectives concerning Gaussian random field models and their use in applied functional analysis, it seems worth to stress that they solely constitute one particular galaxy in a gigantic universe of models and methods. A number of approaches from numerical analysis have been found useful for function approximation, ranging from multi-resolution techniques (Daubechies, 1992; Candès and Donoho, 1999), interpolating and smoothing splines (Wahba, 1990; Friedman, 1991; Gu, 2002; Lin and Zhang, 2006), and many others (see Hastie et al., 2001). In optimization, let us plainly mention that derivative-free methods are well-represented in the evolutionary community (See for example Hansen (2006) on the CMA-ES algorithm, or Auger et al. (2013) and the references therein for a global overview). Similarly, problems such as probability of failure and excursion set estimation dispose of their own literature across diverse research fields, and it would be overly ambitious to try to summarize them here. The approaches considered in the present work do have a limited scope (in terms of dimensionality, number of evaluations, type of functions considered, etc.), and even if this scope might slightly enlarge at the cost of additional research, they owe to be seen as part of a class of methods among a myriad of other existing classes.

Another limitation of the kind of model considered here is the distributional assumption of Gaussianity, that constitutes a very special prior with a number of potentially detrimental consequences for uncertainty quantification. In particular, the Gaussianity of the associated conditional distributions makes such models very rough when it comes to quantifying probabilities of extreme events. Making it fully Bayesian is an option that enables to get thicker tails on predictive distributions. Another possibility would be to work directly with other kinds of random field models, e.g. with max-stable fields Schlather (2002); Spodarev et al. (2013). Of course, this would be at the price of loosing the computational simplicity of Gaussian random fields in terms of prediction and conditional simulations. Future research may tell if such alternative models are worth for the approximation of deterministic functions, and also to what extent max-stable fields and results from extreme value theory may lead to useful approaches in numerical optimization (Huessler, 2010).

Bibliography

Anderes, E. B. (2005). *Estimating deformations of isotropic Gaussian random fields*. PhD thesis, Univ. Chicago.

Antoniadis, A. (1984). Analysis of variance on function spaces. *Statistics*, 15:59–71.

Archambeau, C. and Bach, F. (2011). Multiple Gaussian process models. Technical report. http://arxiv.org/pdf/1110.5238.pdf.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transaction of the American Mathematical Society*, 68 (3):337 – 404.

Auger, A., Bischl, B., Brockhoff, D., Hansen, N., Mersmann, O., Pošìk, P., and H., T. (2013). Black-box optimization benchmarking (BBOB) 2013. Webiste of the 4th GECCO Workshop for Real-Parameter Optimization. http://coco.gforge.inria.fr/doku.php?id=bbob-2013.

Bayarri, M. J., Berger, J. O., Kennedy, M. C., Kottas, A., Paulo, R., Sacks, J., Cafeo, J., Lin, C.-H., and Tu, J. (2009). Predicting vehicle crashworthiness: Validation of computer models for functional and hierarchical data. *Journal of the American Statistical Association*, 104:929–943.

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.

Benassi, R. (2013). *Nouvel algorithme d'optimisation bayésien utilisant une approche Monte-Carlo séquentielle*. PhD thesis, Supélec.

Berg, C., Christensen, J., and Ressel, P. (1984). *Harmonic Analysis on Semigroups*. Springer.

Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing kernel Hilbert spaces in probability and statistics*. Kluwer Academic Publishers.

Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M. (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal*, 46(10):2459–2468.

Binois, M., Ginsbourger, D., and Roustant, O. (2013). Quantifying uncertainty on Pareto fronts with Gaussian Process conditional simulations, submitted.

Bogachev, V. (1998). *Gaussian Measures*. American Mathematical Society, Providence, RI.

Bull, A. (2011). Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 12:2879–2904.

Candès, E. and Donoho, D. (1999). Ridgelets: a key to higher-dimensional intermittency? *Phil. Trans. R. Soc. Lond. A.*, 357:24952509.

Chalup, S. and Mitschele, A. (2008). *Handbook on Information Technology in Finance. International Handbooks Information System*, chapter Kernel Methods in Finance, pages 655–687. Springer.

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 (Accepted on October 1st 2013). *Technometrics*.

Chevalier, C. and Ginsbourger, D. (2013). Fast computation of the multipoint expected improvement with applications in batch selection. *Proceedings of the LION7 conference, Lecture Notes in Computer Science (in press)*.

Chevalier, C., Ginsbourger, D., Bect, J., and Molchanov, I. (2013a). 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.

Chevalier, C., Ginsbourger, D., and Emery, X. (2013b). Corrected kriging update formulae for batch-sequential data assimilation. *Proceedings of the*

15th annual conference of the international association for mathematical geosciences (to appear in Springer Lectures Notes in Earth System Sciences).

Chevalier, C., Picheny, V., , and Ginsbourger, D. (2012). *KrigInv: Kriging-based Inversion for Deterministic and Noisy Computer Experiments*. R package version 1.3.

Chevalier, C., Picheny, V., and Ginsbourger, D. (2014b). The KrigInv package: An efficient and user-friendly R implementation of kriging-based inversion algorithms. *Computational Statistics and Data Analysis*, 71:1021–1034.

Cohen, S. and Lifshits, M. (2012). Stationary Gaussian random field on hyperbolic spaces and Euclidean spheres. *ESAIM: Probability and Statistics*, 16:165–221.

Contal, E. and Vayatis, N. (2013). Gaussian process optimization with mutual information. Technical report, Ecole Normale Supérieure de Cachan.

Cramér, H. and Leadbetter, M. R. (1967). *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*. Wiley, New York.

Cressie, N. (1993). *Statistics for spatial data*. Wiley.

Cuturi, M., Fukumizu, K., and Vert, J. (2005). Semigroup kernels on measures. *Journal of machine learning research*, 6:1169–1198.

Da Veiga, S. and Marrel, A. (2012). Gaussian process modeling with inequality constraints. *Annales de la Faculté des Sciences de Toulouse*, 21 (3):529–555.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics.

de Baar, J., Dwight, R., and Bijl, H. (2013). Speeding up kriging through fast estimation of the hyperparameters in the frequency-domain. *Computers and Geosciences*, 54:99–1006.

Dubrule, O. (1983). Cross validation of kriging in a unique neighborhood. *Journal of the International Association for Mathematical Geology*, 15 (6):687–699.

Dupuy, D. and Helbert, C. (2013). *DiceEval: Construction and evaluation of metamodels*. R package version 1.2.

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.

Durrande, N., Ginsbourger, D., and Roustant, O. (2012). Additive covariance kernels for high-dimensional Gaussian process modeling. *Annales de la Faculté des Sciences de Toulouse*, 21 (3):481–499.

Durrande, N., Ginsbourger, D., Roustant, O., and Carraro, L. (2013). ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis. *Journal of Multivariate Analysis*, 115:57–67.

Duvenaud, D., Nickisch, H., and Rasmussen, C. (2011). *Additive Gaussian Processes*. NIPS conference.

Echard, B., Gayton, N., and Lemaire, M. (2010). Kriging-based Monte Carlo simulation to compute the probability of failure efficiently: AK-MCS method. In *6èmes Journées Nationales de Fiabilité, 24–26 mars, Toulouse, France*.

Eddelbuettel, D. and Francois, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *The Annals of Statistics*, 9:586–596.

Emery, X. (2009). The kriging update equations and their application to the selection of neighboring data. *Computational Geosciences*, 13(1):211–219.

Emmerich, M., Giannakoglou, K., and Naujoks, B. (2006). Single-and multiobjective optimization assisted by Gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10 (4).

Emmerich, M. T., Deutz, A. H., and Klinkenberg, J. W. (2011). Hypervolume-based expected improvement: Monotonicity properties and exact computation. In *Evolutionary Computation (CEC), 2011 IEEE Congress on*, pages 2147–2154. IEEE.

Espinasse, T. (2011). *Modélisation et prédiction de champs et processus gaussiens indexés par des graphes*. PhD thesis, Institut de Mathématiques de Toulouse.

Fasshauer (2011). Positive definite kernels: past, present and future. *Dolomites Research Notes on Approximation*, 4:21–63.

Fernex, F., Heulers, L., Jacquet, O., Miss, J., and Richet, Y. (2005). The MORET 4B Monte Carlo code - New features to treat complex criticality systems. In *M&C International Conference on Mathematics and Computation Supercomputing, Reactor Physics and Nuclear and Biological Application, Avignon, France*.

Forrester, A. I. J., Bressloff, N. W., and Keane, A. J. (2006). Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proc. R. Soc. A*, 462:2177–2204.

Franco, J., Dupuy, D., Roustant, O., Damblin, G., and Iooss, B. (2013). *DiceDesign: Designs of Computer Experiments*. R package version 1.3.

Franey, M., Ranjan, P., and Chipman, H. (2011). Branch and bound algorithms for maximizing expected improvement functions. *Journal of Statistical Planning and Inference*, 141:42–55.

Frazier, P. I. (2012). Parallel global optimization using an improved multipoints expected improvement criterion. In *INFORMS Optimization Society Conference, Miami FL*.

Frazier, P. I., Powell, W. B., and Dayanik, S. (2008). A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439.

Friedman, J. (1991). Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, 19:1–141.

Fruth, J., Muehlenstaedt, T., Roustant, O., and Jastrow, M. (2013). *fanovaGraph: Building Kriging models from FANOVA graphs*. R package version 1.4.4.

Fruth, J., Roustant, O., and Kuhnt, S. (2014). Total interaction index: A variance-based sensitivity index for second-order interaction screening (Available online December 4th 2013). *Journal of Statistical Planning and Inference*.

Gaetan, C. and Guyon, X. (2009). *Spatial statistics and modeling*. Springer Verlag.

Genton, M. (2001). Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research*, 2:299–312.

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Springer-Verlag.

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.

Ginsbourger, D., Baccou, J., Chevalier, C., Garland, N., Perales, F., and Monerie, Y. (2013a). Bayesian adaptive reconstruction of profile optima and optimizers, in preparation.

Ginsbourger, D., Bay, X., Roustant, O., and Carraro, L. (2012). Argument-wise invariant kernels for the approximation of invariant functions. *Annales de la Faculté des Sciences de Toulouse*, 21 (3):501–527.

Ginsbourger, D., Durrande, D., and Roustant, O. (2013b). Kernels and designs for modelling invariant functions: From group invariance to additivity. In Uciński, D., Atkinson, A., and Patan, C., editors, *mODa 10 Advances in Model-Oriented Design and Analysis*. Physica-Verlag HD.

Ginsbourger, D., Janusevskis, J., and Le Riche, R. (2010a). Dealing with asynchronicity in parallel Gaussian process based global optimization. Technical report, OMD2 Project Technical report.

Ginsbourger, D. and Le Riche, R. (2010). Towards Gaussian process-based optimization with finite time horizon. In Giovagnoli, A., Atkinson, A. C.,

Torsney, B., and May, C., editors, *mODa 9 Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, pages 89–96. Physica-Verlag HD.

Ginsbourger, D., Le Riche, R., and Carraro, L. (2010b). Kriging is well-suited to parallelize optimization. In Hiot, L. M., Ong, Y. S., Tenne, Y., and Goh, C.-K., editors, *Computational Intelligence in Expensive Optimization Problems*, Adaptation Learning and Optimization, pages 131–162. Springer Berlin Heidelberg.

Ginsbourger, D., Picheny, V., Roustant, O., and with contributions by C. Chevalier and T. Wagner (2013c). *DiceOptim: Kriging-based optimization for computer experiments*. R package version 1.4.

Ginsbourger, D., Rosspopoff, B., Pirot, G., Durrande, N., and Renard, P. (2013d). Distance-based kriging relying on proxy simulations for inverse conditioning. *Advances in Water Resources*, 52:275–291.

Ginsbourger, D., Roustant, O., and Durrande, N. (2013e). Invariances of random fields paths, with applications in Gaussian process regression. arXiv:1308.1359.

Girdziusas, R., Le Riche, R., Viale, F., and Ginsbourger, D. (2012). Parallel budgeted optimization applied to the design of an air duct. Technical report, OMD2 Project Technical Report.

Gramacy, R. B. and Lee, H. K. H. (2008). Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103:1119–1130.

Gu, C. (2002). *Smoothing spline ANOVA models*. Springer Verlag.

Gunn, S. and Kandola, J. (2002). Structural modelling with sparse kernels. *Machine learning*, 48:137–163.

Guttorp, P. and Sampson, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87 (417):108–119.

Habib, S., Heitmann, K., Higdon, D., Nakhleh, C., and Williams, B. (2007). Cosmic calibration: Constraints from the matter power spectrum and the cosmic microwave background. *Phys. Rev.*, D76.

Hairer, M. (2009). An introduction to stochastic pdes. Lecture notes.

Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics*, 35(4):403–410.

Hansen, N. (2006). *Towards a new evolutionary computation. Advances in estimation of distribution algorithms*, chapter The CMA Evolution Strategy: A Comparing Review, pages 75–102. Springer.

Hastie, T. and Tibshirani, R. (1995). Generalized additive models. *Encyclopedia of Statistical Sciences*.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19:293–325.

Huang, D., Allen, T., Notz, W., and Zheng, N. (2006). Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization*, 34:441–466.

Huessler, J. (2010). *Experimental Methods for the Analysis of Optimization Algorithms*, chapter On Applications of Extreme Value Theory in Optimization, pages 185–207. Springer.

Janusevskis, J., Le Riche, R., and Ginsbourger, D. (2012). Expected improvements for the asynchronous parallel global optimization of expensive functions : potentials and challenges. In *Learning and Intelligent Optimization*, Lecture Notes in Computer Science, pages 413–418.

Jones, D., Schonlau, M., and Welch, W. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492.

Kadri, H., Rakotomamonjy, A., Bach, F., and Preux, P. (2012). Multiple operator-valued kernel learning. In *Advances in Neural Information Processing Systems (NIPS)*.

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 11 (9):1–20.

Kaufman, C. and Sain, S. (2010). Bayesian functional anova modeling using gaussian process prior distributions. *Bayesian Analysis*, 5:123–150.

Kuo, F. Y., Sloan, I. H., Wasilkowski, G. W., and Wozniakowski, H. (2010). On decompositions of multivariate functions. *Mathematics of Computation*, 79:953–966.

Kuo, H.-H. (1975). *Gaussian Measures in Banach Spaces*, volume 463 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg.

Kushner, H. J. (1964). A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *J. Basic Engineering*, 86:97–106.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces. Isometry and processes*. Springer.

Lee, H., Higdon, D., Calder, K., and Holloman, C. (2005). Spatial models via convolutions of intrinsic processes. *Statistical Modelling*, 5:1–21.

Lenz, N. (2013). Additivity and ortho-additivity in gaussian random fields. Master's thesis, Mathematical Institute, University of Bern.

Lin, Y. and Zhang, H. (2006). Component selection and smoothing in multivariate nonparametric regression. *Annals of Statistics*, 34:2272–2297.

Mahlstein, I., Martius, O., Chevalier, C., and Ginsbourger, D. (2012). Changes in the odds of extreme events in the atlantic basin depending on the position of the extratropical jet. *Geophysical Research Letters*, 39:L22805.

Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71:135–46.

Marrel, A., Iooss, B., Laurent, B., and Roustant, O. (2009). Calculations of sobol indices for the gaussian process metamodel. *Reliability Engineering & System Safety*, 94:742–751.

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

Mockus, J. (1989). *Bayesian Approach to Global Optimization. Theory and Applications*. Kluwer Academic Publisher, Dordrecht.

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.

Molchanov, I. (2005). *Theory of Random Sets*. Springer, London.

Muehlenstaedt, T., Roustant, O., Carraro, L., and Kuhnt, S. (2012). Data-driven Kriging models based on FANOVA-decomposition. *Statistics and Computing*, 22 (3):723–738.

Nair, R., Jhingan, S., and Jain, D. (2013). Exploring dark energy dynamics with Gaussian processes. arXiv:1306.0606.

Omre, H. (1987). Bayesian kriging – merging observations and qualified guesses in kriging. *Mathematical Geology*, 19:25–39.

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), online proceedings*, Trento, Italy.

Paciorek, C. (2003). *Nonstationary Gaussian Processes for Regression and Spatial Modelling*. PhD thesis, Carnegie Mellon University.

Parr, J. M. (2012). *Improvement Criteria for Constraint Handling and Multiobjective Optimization*. PhD thesis, University of Southampton.

Parthasarathy, K. and Schmidt, K. (1972). *Positive Definite Kernels, Continuous Tensor Products, and Central Limit Theorems of Probability Theory*. Lecture Notes in Mathematics, Springer.

Picheny, V. and Ginsbourger, D. (2010). Noisy expected improvement and on-line computation time allocation for the optimization of simulators with tunable fidelity. In *2nd International Conference on Engineering Optimization*.

Picheny, V. and Ginsbourger, D. (2014). Noisy kriging-based optimization methods: a unified implementation within the DiceOptim package (in press, available online March 26. 2013). *Computational Statistics and Data Analysis*, 71:1035–1053.

Picheny, V., Ginsbourger, D., Richet, Y., and Caplin, G. (2013a). Quantile-based optimization of noisy computer experiments with tunable precision (with discussion). *Technometrics*, 55(1):2–13.

Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R., and Kim, N.-H. (2010). Adaptive designs of experiments for accurate approximation of target regions. *Journal of Mechanical Design*, 132(7).

Picheny, V., Wagner, T., and Ginsbourger, D. (2013b). Benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626.

Rajput, B. and Cambanis, S. (1972). Gaussian processes and Gaussian measures. *The Annals of Mathematical Statistics*, 43 (6):1944–1952.

Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541.

Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. the MIT Press.

Richet, Y., Caplin, G., Crevel, J., Ginsbourger, D., and Picheny, V. (2013). Using the efficient global optimization algorithm to assist nuclear criticality safety assessment. *Nuclear Science Engineering (American Nuclear Society)*, 175 (1):1–18.

Richet, Y., Deville, Y., and Chevalier, C. (2012). *DiceView: Plot methods for computer experiments design and surrogates*. R package version 1.3-0.

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.

Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis: the primer*. Wiley Online Library.

Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.

Schaback, R. (2009). Solving the laplace equation by meshless collocation using harmonic kernels. *Advances in Computational Mathematics*, 31(4):457–470.

Scheidt, C. and Caers, J. (2009). Representing spatial uncertainty using distances and kernels. *Mathematical Geosciences*, 41(4):397–419.

Scheuerer, M. (2009). *A Comparison of Models and Methods for Spatial Interpolation in Statistics and Numerical Analysis*. PhD thesis, University of Göttingen.

Scheuerer, M. (2010). Regularity of the sample paths of a general second order random field. *Stochastic Processes and their Applications*, 120:1879–1897.

Scheuerer, M. and Schlather, M. (2012). Covariance models for divergence-free and curl-free random vector fields. *Stochastic Models*, 28(3):433–451.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5 (1):33–44.

Schoenberg, I. J. (1938). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):522–536.

Schölkopf, B. and Smola, A. (2002). *Learning with kernels*. MIT Press.

Schölkopf, B., Tsuda, K., and Vert, J. (2004). *Kernel Methods in Computational Biology (Computational Molecular Biology)*. MIT Press.

Schonlau, M. (1997). *Computer Experiments and Global Optimization*. PhD thesis, University of Waterloo.

Schonlau, M., Welch, W., and Jones (1998). Global versus local search in constrained optimization of computer models. In *New developments and applications in experimental design*, number 34 in IMS Lecture Notes - Monograph Series, pages 11–25.

Schwartz, L. (1964). Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *Journal d'Analyse Mathématique*, 13 (1):115–256.

Snelson, E. and Ghahramani, Z. (2006). Sparse Gaussian processes using pseudo-inputs. In *Advances in Neural Information Processing Systems 18*.

Sóbester, A., Forrester, A., Toal, D., Tresidder, E., and Tucker, S. (2012). Engineering design applications of surrogate-assisted optimization techniques. *Optimization and Engineering*, pages 1–22.

Sobol, I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280.

Spodarev, E., Shmileva, E., and Roth, S. (2013). Extrapolation of stationary random fields.

Srinivas, N., A., K., Kakade, S., and Seeger, M. (2012). Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58:3250–3265.

Sriperumbudur, B. (2010). *Reproducing kernel space embeddings and metrics on probability measures*. PhD thesis, UC San Diego.

Stein, M. (1999). *Interpolation of spatial data, some theory for kriging*. Springer.

Stitson, M. O., Gammerman, A., Vapnik, V., Vovk, V., Watkins, C., and Weston, J. (1997). Support vector regression with ANOVA decomposition kernels. Technical report, Royal Holloway , University of London.

Sweeting, T. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *The Annals of Statistics*, 8(6):1375–1381.

Taddy, M. A., Lee, H. K. H., Gray, G. A., and Griffin, J. D. (2009). Bayesian guided pattern search for robust local optimization. *Technometrics*, 51:389–401.

Tallis, G. (1961). The moment generating function of the truncated multinormal distribution. *J. Roy. Statist. Soc. Ser. B*, 23(1):223–229.

Tarieladze, V. and Vakhania, N. (2007). Disintegration of gaussian measures and average-case optimal algorithms. *Journal of Complexity*, 23:851–866.

Urtasun, R. (2011). A Gaussian process view on MKL. Video Lecture.

Van der Vaart, A. and Van Zanten, J. (2008). Reproducing kernel hilbert spaces of gaussian priors. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, pages 200–222. Institute of Mathematical Statistics Collect 3.

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.

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.

Wagner, T., Emmerich, M., Deutz, A., and Ponweiser, W. (2011). On expected-improvement criteria for model-based multi-objective optimization. *Parallel Problem Solving from Nature–PPSN XI*, pages 718–727.

Wahba, G. (1990). *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia.

Wang, C. and Neal, R. (2013). MCMC methods for Gaussian process models using fast approximations for the likelihood. Technical report, University of Toronto.

Wang, Z., Zoghi, M., Hutter, F., Matheson, D., and de Freitas, N. (2013). Bayesian optimization in high dimensions via random embeddings. In *International Joint Conferences on Artificial Intelligence*.

Xiong, Y., Chen, W., Apley, D., and Ding, X. (2007). A non-stationary covariance-based kriging method for metamodelling in engineering design. *International Journal of Numerical Methods in Engineering*, 71:733–756.

Ziegel, J. (2013). Convolution roots and differentiability of isotropic positive definite functions on spheres. *Proceedings of the American Mathematical Society, to appear*.

Zilinskas, A. (1992). A review of statistical models for global optimization. *J. Global Optim.*, 2:145–153.

Appendix A

Selected papers related to chapter 1

A.1 Argumentwise invariant kernels for the approximation of invariant functions

Original title: *Argumentwise invariant kernels for the approximation of invariant functions*

Date of publication: **2012**

Authors: D. Ginsbourger, X. Bay, L. Carraro, O. Roustant

Journal: *Annales de la Faculté des Sciences de Toulouse* (Tome 21, Number 3, pp. 501-527)

URL: http://afst.cedram.org/afst-bin/feuilleter?id=AFST_2012_6_21_3

Corresponding file in archive: A1_Appendix_Habilitation_Ginsbourger.pdf

A.2 Additive Covariance kernels for high-dimensional Gaussian Process modeling

Original title: *Additive Covariance kernels for high-dimensional Gaussian Process modeling*

Date of publication: **2012**

Authors: N. Durrande, D. Ginsbourger, O. Roustant

Journal: *Annales de la Faculté des Sciences de Toulouse* (Tome 21, Number 3, pp. 481-499)

URL: http://afst.cedram.org/afst-bin/feuilleter?id=AFST_2012_6_21_3

Corresponding file in archive: A2_Appendix_Habilitation_Ginsbourger.pdf

A.3 ANOVA kernels for model-based sensitivity analysis

Original title: *ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis*

Date of publication: **2013**

Authors: N. Durrande, D. Ginsbourger, O. Roustant, L. Carraro

Journal: *Journal of Multivariate Analysis* (Volume 115, pp. 57-67)

DOI: http://dx.doi.org/10.1016/j.jmva.2012.08.016

Corresponding file in archive: A3_Appendix_Habilitation_Ginsbourger.pdf

A.4 Kernels invariant under combination of composition operators

Original title: *Kernels and Designs for Modelling Invariant Functions: From Group Invariance to Additivity*

Date of publication: **2013**

Authors: D. Ginsbourger, O. Roustant, N. Durrande

Book: *mODa 10 – Advances in Model-Oriented Design and Analysis Contributions to Statistics* (pp. 107-115)

DOI: http://dx.doi.org/10.1007/978-3-319-00218-7_13

Corresponding file in archive: A4_Appendix_Habilitation_Ginsbourger.pdf

Appendix B

Selected papers related to chapter 2

B.1 Finite Time Optimization

Original title: *Towards Gaussian Process-based Optimization with Finite Time Horizon*

Date of publication: **2010**

Authors: D. Ginsbourger, R. Le Riche

Book: *mODa 9 Advances in Model-Oriented Design and Analysis Contributions to Statistics* (pp. 89-96)

DOI: http://dx.doi.org/10.1007/978-3-7908-2410-0_12

Corresponding file in archive: B1_Appendix_Habilitation_Ginsbourger.pdf

B.2 Quantile-based noisy optimization, with discussion

Original title: *Quantile-Based Optimization of Noisy Computer Experiments With Tunable Precision*

Date of publication: **2013**

Authors: V.Picheny, D. Ginsbourger, Y. Richet, G. Caplin

Journal: *Technometrics* (Volume 55, Issue 1, pp. 2-36)

DOI: http://dx.doi.org/10.1080/00401706.2012.707580

Corresponding files in archive: B2_Appendix_Habilitation_Ginsbourger.pdf and B2bis_Appendix_Habilitation_Ginsbourger.pdf

B.3 Fast parallel kriging-based SUR strategies

Original title: *Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set*

Date of publication: published online 21 Nov. 2013

Authors: C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, Y. Richet

Journal: *Technometrics* (in press)

DOI: http://dx.doi.org/10.1080/00401706.2013.860918

Corresponding files in archive: B3_Appendix_Habilitation_Ginsbourger.pdf

B.4 Vorob'ev expectation and deviation for excursion set estimation

Original title: *Estimating and quantifying uncertainties on level sets using the Vorobev expectation and deviation with Gaussian process models*

Date of publication: **2013**

Authors: C. Chevalier, D. Ginsbourger, J. Bect, I. Molchanov

Book: *mODa 10 – Advances in Model-Oriented Design and Analysis Contributions to Statistics* (pp. 35-43)

DOI: http://dx.doi.org/10.1007/978-3-319-00218-7_5

Corresponding file in archive: B4_Appendix_Habilitation_Ginsbourger.pdf

Appendix C

Selected papers related to chapter 3

C.1 Distance-based kriging and the ProKSI algorithm

Original title: *Distance-based Kriging based on proxy simulations for inverse conditioning*

Date of publication: **2013**

Authors: D. Ginsbourger, B. Rosspopoff, G. Pirot, N. Durrande, P. Renard

Journal: *Advances in Water Resources* (Volume 52, pp. 275-291)

DOI: http://dx.doi.org/10.1016/j.advwatres.2012.11.019

Corresponding file in archive: C1_Appendix_Habilitation_Ginsbourger.pdf

C.2 DiceKriging and DiceOptim

Original title: *DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodelling and optimization*

Date of publication: **2012**

Authors: O. Roustant, D. Ginsbourger, Y. Deville

Journal: *Journal of Statistical Software* (Volume 51, Issue 1, pp. 1-55)

URL: http://www.jstatsoft.org/v51/i01

Corresponding file in archive: C2_Appendix_Habilitation_Ginsbourger.pdf

Appendix D

Proofs of some properties of Chapter 1

The proofs are similar to those given in Ginsbourger et al. (2012) but have been adapted here to the notations used throughout the habilitation.

Proof of Prop. 1.1.3. Let us first assume that Z has Φ-invariant paths, up to a modification. Then, there exists a field $\widetilde{Z}$ with Φ-invariant paths and such that $\forall \mathbf{x} \in D,\ \mathbb{P}(Z_{\mathbf{x}} = \widetilde{Z}_{\mathbf{x}}) = 1$. This implies that $k_Z = k_{\widetilde{Z}}$ since the 2-dimensional distributions of Z and $\widetilde{Z}$ are the same. Now, by Φ-invariance of $\widetilde{Z}$'s paths, we have $\forall \mathbf{x} \in D\ \forall g \in G\ \forall \omega \in \Omega,\ \widetilde{Z}_{\mathbf{x}}(\omega) = \widetilde{Z}_{g.\mathbf{x}}(\omega)$, so that in particular, $\forall \mathbf{x} \in D\ \forall g, g' \in G$

$$k_{\widetilde{Z}}(g.\mathbf{x}, g'.\mathbf{x}') = \mathrm{Cov}[\widetilde{Z}_{g.\mathbf{x}}, \widetilde{Z}_{g'.\mathbf{x}'}] = \mathrm{Cov}[\widetilde{Z}_{\mathbf{x}}, \widetilde{Z}_{g'.\mathbf{x}'}] = \mathrm{Cov}[\widetilde{Z}_{\mathbf{x}}, \widetilde{Z}_{\mathbf{x}'}] = k_{\widetilde{Z}}(\mathbf{x}, \mathbf{x}').$$

Reciprocally, let us now assume that k_Z is argumentwise invariant under Φ. Let us denote by $A \subset D$ a fundamental domain for Φ, and by $\pi_A : D \longrightarrow A$ the projector mapping any $\mathbf{x} \in D$ to its representer $\pi_A(\mathbf{x}) \in A$, i.e. to the point of A being in the same orbit. We then define the random field $\widetilde{Z}$ by

$$\widetilde{Z}_{\mathbf{x}} := Z_{\pi_A(\mathbf{x})} \quad (\mathbf{x} \in D\ )$$

By construction, $\widetilde{Z}$ has paths invariant under Φ. Now, for any $\mathbf{x} \in D$, there exists $g \in G$ such that $\pi_A(\mathbf{x}) = g.\mathbf{x}$. Subsequently,

$$\mathrm{Var}[Z_{\mathbf{x}} - \widetilde{Z}_{\mathbf{x}}] = \mathrm{Var}[Z_{\mathbf{x}} - Z_{g.\mathbf{x}}] = k(\mathbf{x}, \mathbf{x}) + k(g.\mathbf{x}, g.\mathbf{x}) - 2k(\mathbf{x}, g.\mathbf{x}) = 0,$$

so that $\mathbb{P}(Z_{\mathbf{x}} = \widetilde{Z}_{\mathbf{x}}) = 1$, and Z is indeed a modification of a random field with Φ-invariant paths. □

Proof of Prop. 1.1.4. The covariance vector $\mathbf{k}(.)$ is Φ-invariant by argumentwise Φ-invariance of k. The results then follow by plugging in the equality $\mathbf{k}(g.\mathbf{x}) = \mathbf{k}(\mathbf{x})$ in Eq. 1.1 and in the kriging variance equation (cf. (Ginsbourger et al., 2012, Eq. 3.4)). □

Proof of Prop. 1.1.5. (1) follows from the Φ-invariance of $\mathbf{k}(.)$ applied to Eq.1.4. For (2), it is useful to recall that conditional simulations are paths drawn from the conditional distribution of the considered field. Now, conditionally on the observations, this field has a mean function (the kriging mean m) known to be Φ-invariant according to Prop. 1.1.4. Since the complement to this mean function is a centred GRF with argumentwise Φ-invariant kernel (from (1)), Prop. 1.1.3 implies that the conditional simulations are Φ-invariant, as sums of a Φ-invariant function plus Φ-invariant paths. □

Appendix E

Invariances in terms of Gaussian measures

Let us consider here a Gaussian measure μ on a separable Banach space $(\mathcal{B}, ||\cdot||_{\mathcal{B}})$. Let us denote by

$$C_\mu : (\ell, \ell') \in (\mathcal{B}^*)^2 \to C_\mu(\ell, \ell') := \int_{\mathcal{B}} \ell(x)\ell'(x)d\mu(x)$$

the covariance operator of μ, and by

$$\widehat{\mu} : \ell \in \mathcal{B}^* \to \widehat{\mu}(\ell) := \int_{\mathcal{B}} e^{i\ell(x)}d\mu(x)$$

the Fourier transform of μ. In case of a separable $\mathcal{B}$, it is well-known that μ is completely characterized through the push-forward measures $\ell_*\mu$ ($\ell \in \mathcal{B}^*$), and consequently that any of C_μ or $\widehat{\mu}$ characterizes μ. Let us also recall that both notions are notably connected through the identity $\widehat{\mu}(\ell) = e^{-\frac{1}{2}C_\mu(\ell,\ell)}$.

As pointed in Hairer (2009), it follows from the definition of a Gaussian measure and the expression for its Fourier transform that images of Gaussian measures by bounded operators are Gaussian measures. More precisely,

Proposition E.0.1. *If μ is a Gaussian measure on $\mathcal{B}$ and $T : \mathcal{B} \to \mathcal{B}_2$ is a bounded linear map for $\mathcal{B}_2$ some other Banach space, then $\eta = T_*\mu$ is a Gaussian measure on $\mathcal{B}_2$, with*

$$C_\eta(\ell, \ell') = C_\mu(T^*\ell, T^*\ell') \quad (\ell, \ell') \in (\mathcal{B}_2^*)^2$$

where $T^* : \mathcal{B}_2^* \to \mathcal{B}^*$ is the adjoint to T, i.e. the operator such that $(T^*\ell)(x) = \ell(T(x))$ for every $x \in \mathcal{B}$ and $\ell \in \mathcal{B}_2^*$.

In our context, $\mathcal{B}_2 = \mathcal{B}$, and the link between path invariances and argumentwise invariance of the covariance kernel can be reformulated as a consequence of Proposition E.0.1, relying on the concepts of product and diagonal measures. Before stating the main proposition, let us recall the definition of the diagonal extension of a measure to a product space.

DEFINITION E.0.2. *Let μ be a measure on $(\mathcal{B}, \mathcal{A})$ ($\mathcal{A}$ being the Borel sigma-algebra associated with $\mathcal{B}$ unless precised otherwise). The diagonal measure μ_Δ on $(\mathcal{B}^2, \mathcal{A}^2)$ associated with μ is defined by*

$$\mu_\Delta(S) := \mu(\{x \in \mathbb{B} : (x, x) \in S\}) \quad (S \in \mathcal{A}^2).$$

In other words, μ_Δ is the push-forward of μ on $\mathcal{B}^2$ by the map $(id_\mathcal{B}, id_\mathcal{B}) : x \in \mathcal{B} \longrightarrow (x, x) \in \mathcal{B}^2$. Similarly, let us denote by $(id_\mathcal{B}, T)$ the function mapping $x \in \mathcal{B}$ to $(x, T(x)) \in \mathcal{B}^2$. We now characterize the almost sure invariance of Gaussian random elements by giving a necessary and sufficient condition for the measures $(id_\mathcal{B}, T)_*\mu$ and μ_Δ to coincide.

PROPOSITION E.0.3. *Let μ be a Gaussian measure on $(\mathcal{B}, \mathcal{A})$, $T \in \mathcal{L}(\mathcal{B})$ be a bounded linear map from $\mathcal{B}$ to itself, and $T^* \in \mathcal{L}(\mathcal{B}^*)$ denote its adjoint. Then, $(id_\mathcal{B}, T)_*\mu = \mu_\Delta$ if and only if C_μ is argumentwise invariant by T^*, i.e.*

$$C_\mu(T^*\ell, \ell') = C_\mu(\ell, T^*\ell') = C_\mu(\ell, \ell') \quad ((\ell, \ell') \in \mathcal{B}^* \times \mathcal{B}^*).$$

Proof. Assume first that C_μ is argumentwise invariant by T^*. Consider an arbitrary bounded linear form $\ell \in (\mathcal{B} \times \mathcal{B})^*$, and let $\ell_i = \ell \circ \pi_i \in \mathcal{B}^*$ $(i = 1, 2)$, where π_1, π_2 are maps on $\mathcal{B}$ defined by $\pi_1(x) = (x, 0)$ and $\pi_1(x) = (0, y) \in \mathcal{B} \times \mathcal{B}$, so that $\ell = (\ell_1, \ell_2)$. Then,

$$\begin{aligned} \widehat{\mu_\Delta}(\ell) &= \int_{\mathcal{B}\times\mathcal{B}} e^{i\ell((x,y))} d\mu_\Delta((x, y)) \\ &= \int_\mathcal{B} e^{i\ell((x,x))} d\mu(x) = \int_\mathcal{B} e^{i(\ell_1+\ell_2)(x)} d\mu(x) = e^{-\frac{1}{2}C_\mu(\ell_1+\ell_2, \ell_1+\ell_2)}, \end{aligned} \tag{E.1}$$

whereof μ_Δ is a Gaussian measure on $\mathcal{B} \times \mathcal{B}$ with covariance operator:

$$C_{\mu_\Delta} : (\ell, \ell') \in ((\mathcal{B} \times \mathcal{B})^*)^2 \to C_{\mu_\Delta}(\ell, \ell') = C_\mu(\ell_1 + \ell_2, \ell'_1 + \ell'_2). \tag{E.2}$$

Now, using the assumption of argumentwise invariance, we get

$$\begin{aligned} C_{\mu_\Delta}(\ell, \ell') &= C_\mu(\ell_1 + T^*\ell_2, \ell'_1 + T^*\ell'_2) \\ &= \int_{\mathcal{B}} (\ell_1(x) + \ell_2(Tx))(\ell'_1(x) + \ell'_2(Tx))\mathrm{d}\mu(x) \\ &= \int_{\mathcal{B}} \ell((x, Tx))\ell'((x, Tx))\mathrm{d}\mu(x) \\ &= \int_{\mathcal{B}\times\mathcal{B}} \ell((x, y))\ell'((x, y))\mathrm{d}((id_{\mathcal{B}}, T)_*\mu)(x, y) \\ &= C_{(id_{\mathcal{B}},T)_*\mu}(\ell, \ell'), \end{aligned} \tag{E.3}$$

and we conclude that $\mu_\Delta = (id_{\mathcal{B}}, T)_*\mu$. Reciprocally, assuming $\mu_\Delta = (id_{\mathcal{B}}, T)_*\mu$ implies (using Eq E.3) that for all $\ell_1, \ell_2, \ell'_1, \ell'_2 \in \mathcal{B}^*$,

$$C_\mu(\ell_1, T^*\ell'_2)+C_\mu(T^*\ell_2, \ell'_1)+C_\mu(T^*\ell_2, T^*\ell'_2) = C_\mu(\ell_1, \ell'_2)+C_\mu(\ell_2, \ell'_1)+C_\mu(\ell_2, \ell'_2),$$

leading to $C_\mu(\ell_1, \ell'_2) = C_\mu(\ell_1, T_*\ell'_2)$ for $\ell_2 = 0$, which concludes the proof. □

Corollary E.0.4. *Under the conditions of Proposition E.0.3, i.e. if μ is a Gaussian measure on $(\mathcal{B}, \mathcal{A})$ and $T \in \mathcal{L}(\mathcal{B})$ is a bounded linear map with C_μ argumentwise invariant by T^*, then we have*

$$\mu(\{x \in \mathcal{B} : Tx = x\}) = 1.$$

Proof. Using $(id_{\mathcal{B}}, T)_*\mu = \mu_\Delta$, we directly obtain that

$$\begin{aligned} \mu(\{x \in \mathcal{B} : Tx = x\}) &= \mu_\Delta\left(\{(x, Tx), x \in \mathcal{B}\}\right) \\ &= \mu_\Delta\left((id_{\mathcal{B}}, T)(\mathcal{B})\right) \\ &= (id_{\mathcal{B}}, T)_*\mu\left((id_{\mathcal{B}}, T)(\mathcal{B})\right) \\ &= \mu\left((id_{\mathcal{B}}, T)^{-1}(id_{\mathcal{B}}, T)(\mathcal{B})\right) = \mu\left(\mathcal{B}\right) = 1. \end{aligned}$$

□

Appendix F

CV, publication list, and teaching statement