Argumentwise invariant kernels for the approximation of invariant functions

David Ginsbourger, Xavier Bay, Olivier Roustant, Laurent Carraro

ARGUMENTWISE INVARIANT KERNELS FOR THE APPROXIMATION OF INVARIANT FUNCTIONS

DAVID GINSBOURGER, XAVIER BAY, OLIVIER ROUSTANT, AND LAURENT CARRARO

Abstract. We consider the problem of designing adapted kernels for approximating functions invariant under a known finite group action. We introduce the class of argumentwise invariant kernels, and show that they characterize centered square-integrable random fields with invariant paths, as well as Reproducing Kernel Hilbert Spaces of invariant functions. Two subclasses of argumentwise kernels are considered, involving a fundamental domain or a double sum over orbits. We then derive invariance properties for Kriging and conditional simulation based on argumentwise invariant kernels. The applicability and advantages of argumentwise invariant kernels are demonstrated on several examples, including a symmetric function from the reliability literature.

Résumé. Nous considérons le problème d'approximation par méthodes à noyaux de fonctions invariantes sous l'action d'un groupe fini. Nous introduisons les noyaux doublement invariants, et montrons qu'ils caractérisent les champs aléatoires centrés de carré intégrable à trajectoires invariantes, ainsi que les espaces de Hilbert à noyau reproduisant de fonctions invariantes. Deux classes particulières de noyaux doublement invariants sont considérées, basées respectivement sur un domaine fondamental ou sur une double somme sur les orbites. Nous établissons ensuite des propriétés d'invariance pour les modèles de Krigeage et les simulations consitionnelles associés. L'applicabilité et les avantages de tels noyaux sont illustrés sur plusieurs exemples, incluant une fonction symétrique issue d'un problème de fiabilité des structures.

1. Introduction

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) to *Gaussian Process Regression* (GPR) in machine learning. 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 $\mathbf{y} := (y_1, \ldots, y_n) \in \mathbb{R}^n$ 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

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

is at the same time the best approximation of any function f in the RKHS of kernel k subject to $f(\mathbf{x}_i) = y_i$ $(1 \leq i \leq n)$, and the GPR ("Simple Kriging") predictor of any squared-integrable centered random field $(Y_{\mathbf{x}\in D})_{\mathbf{x}\in D}$ of covariance kernel k subject to $Y_{\mathbf{x}_i} = y_i$ $(1 \leq i \leq n)$. $k : D \times D \longrightarrow \mathbb{R}$ stands here 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).

In practical situations (e.g., when the y_i's steem from the output of an expensive-to-evaluate deterministic numerical simulator, say $y : D \to \mathbb{R}$), 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) is to rely on well-known families of kernels, and to perform classical Maximum Likelihood, Cross-Validation, or Bayesian inference of the underlying parameters based on data. For example, most GPR or Kriging softwares offer different 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 solid 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 in the case of Gaussian random fields, see [12]).

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 [32] or [27]), kernels incorporating non-linear transformations of the input space ([20, 4, 47]), or treed gaussian processes ([19]), to cite an excerpt of some of the most popular approaches.

1. We use here the term p.d. for what some authors also call "non-negative definite".

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 y invariant under a measurable action Φ of some finite group G on D, is it possible to construct a metamodel of y respecting that invariance?

Here we investigate classes of kernels leading to metamodels m inheriting known invariances from y. In the particular case of a GPR 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.

The paper is organized as follows. In Section 2, we recall some fundamental algebraic definitions (2.1) and random fields technical notions useful in the sequel (2.2), followed by an overview with discussion on existing work concerning invariant kernels and random fields. The main results are given in Section 3: A characterization of positive definite kernels leading to invariant random fields is given (3.1), and several properties of the corresponding metamodels are discussed (3.2). (3.3) is dedicated to the RKHS interpretation of such kernels. Application results are then presented in Section 4, first with illustrations on toy examples (4.1), with a test case from the reliability literature (4.2), and then by approximating simulated invariant Gaussian Fields (4.3). Finally, a few concluding remarks and a discussion on perspectives and forthcoming research questions are given in Section 5.

2. Definitions and classical results

2.1. Group actions and invariant functions.

Let $(G, *)$ be a group and D a set. We denote by e the neutral element of G. Let us recall some standard definitions from algebra [26].

Definition 1. *A (left) action of the group G on D is a map*

$$\begin{array}{rccc} \Phi & : & D \times E & \longrightarrow & E \\ & & (g, \mathbf{x}) & \longmapsto & g.\mathbf{x} := \Phi(g, \mathbf{x}) \end{array}$$

such that

- $\mathbf{x} \in D \longmapsto \Phi(e, \mathbf{x})$ *is the identitity of D, i.e.* $\forall \mathbf{x} \in D,\ e.\mathbf{x} = \mathbf{x}$,
- $\forall \mathbf{x} \in D,\ \forall g, g' \in G,\ (g * g').\mathbf{x} = g.(g'.\mathbf{x})$.

Definition 2. *The orbit of a point $\mathbf{x} \in D$ under the action Φ is the set*

$$\mathcal{O}(\mathbf{x}) := \{g.\mathbf{x},\ g \in G\}, \tag{2}$$

constituted of images of $\mathbf{x}$ by the action of G.

Definition 3. *$\mathbf{x} \in D$ is a fixed point of the action when $\forall g \in G,\ g.\mathbf{x} = \mathbf{x}$.*

Definition 4. *The fixator of a set $S \subset D$ in G is defined by*

$$Fix_{\Phi}(S) := \{g \in G \mid \forall \mathbf{x} \in S,\ g.\mathbf{x} = \mathbf{x}\}. \tag{3}$$

Definition 5. *The stabilizer of a set $S \subset D$ in G is defined by*

$$Stab_{\Phi}(S) := \{g \in G \mid \forall \mathbf{x} \in S,\ g.\mathbf{x} \in S\} \tag{4}$$

Definition 6. *A measurable set $A \subset D$ is said to be a fundamental domain of Φ if it is a system of representatives of Φ's orbits.*

Remark 1. *A fundamental domain is usually required to have further topological properties, for instance to be the symmetric difference between an open set and a set of measure zero.*

Definition 7. *Le F be an arbitrary set. A map $y : D \longrightarrow F$ is said invariant by Φ, or invariant under the action of the group G, when*

$$\forall \mathbf{x} \in D,\ \forall g \in G,\ y(g.\mathbf{x}) = y(\mathbf{x}) \tag{5}$$

Equivalently, y is said invariant whenever it is constant on the orbits of Φ.

2.2. **Random fields.** We borrow here a few definitions from the book [35], with a few minor changes in the notations.

Definition 8. *Two random fields Y and Y', respectively defined on probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega', \mathcal{F}', \mathbb{P}')$ and sharing a common measurable state space $(D, \mathcal{D})$, are said equivalent if for any finite sequence of points $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)} \in D$ and events $A_1, \dots, A_n \in \mathcal{D}$,*

$$\mathbb{P}\left(Y_{\mathbf{x}^{(1)}} \in A_1, \dots, Y_{\mathbf{x}^{(n)}} \in A_n\right) = \mathbb{P}'\left(Y'_{\mathbf{x}^{(1)}} \in A_1, \dots, Y'_{\mathbf{x}^{(n)}} \in A_n\right) \tag{6}$$

One also says in that case that each one of these random fields is a version of the other, or that both are versions of the same random field. In other words, two random fields are versions of each other whenever they have the same finite-dimensional distributions.

Definition 9. *Two random fields Y and Y' defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are said to be modifications of each other when for all $\mathbf{x} \in D$,*

$$\mathbb{P}(Y_{\mathbf{x}} = Y'_{\mathbf{x}}) = 1 \tag{7}$$

They are said indistinguishable when

$$\mathbb{P}\left(\forall \mathbf{x} \in D, Y_{\mathbf{x}} = Y'_{\mathbf{x}}\right) = 1 \tag{8}$$

As precised in ([35], p. 18), if Y and Y' are modifications of each other, they clearly are equivalent. A slightly less straightforward result is that if two random fields modifications of each other are almost surely continuous, then they are indistinguishable. Finally, let us add a definition which will play a central rôle in the sequel of the paper:

Definition 10. *Y is said to have all its paths Φ-invariant whenever*

$$\forall \omega \in \Omega,\ \forall \mathbf{x} \in D,\ \forall g \in G,\ Y_{\mathbf{x}}(\omega) = Y_{g.\mathbf{x}}(\omega) \tag{9}$$

2.3. Classical results about invariant kernels and random fields.

2.3.1. *Stationarity, isotropy: Invariance-related notions in geostatistics.* A very classical notion in spatial statistics, and more generally in the literature of random processes (including time series in the first place), is the one of *second order* or *weak stationarity*. A centered squared-integrable random field Y is said weakly stationary whenever $\mathrm{cov}(Y_{\mathbf{x}}, Y_{\mathbf{x}'})$ is a function of $\mathbf{x}-\mathbf{x}'$ (here $\mathbf{x}, \mathbf{x}' \in D$) or, in other words, when for any $\mathbf{x} \in D$ and $\mathbf{h}$ such that $\mathbf{x}+\mathbf{h} \in D$, $\mathrm{cov}(Y_{\mathbf{x}+\mathbf{h}}, Y_{\mathbf{x}})$ depends only on $\mathbf{h}$ and not on $\mathbf{x}$. Equivalently, the covariance kernel of Y is such that for any translation $T_{\mathbf{h}}(\mathbf{x}) := \mathbf{x}+\mathbf{h}$ and pair of points $\mathbf{x}, \mathbf{x}' \in D$ with $T_{\mathbf{h}}(\mathbf{x}), T_{\mathbf{h}}(\mathbf{x}') \in D$,

$$k(T_{\mathbf{h}}(\mathbf{x}), T_{\mathbf{h}}(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}') \tag{10}$$

Additionally, a centered weakly stationary random field Y defined over some subset D of a Euclidean space is said to be *isotropic* whenever $\mathrm{cov}(Y_{\mathbf{x}}, Y_{\mathbf{x}'})$ depends only on the norm-induced distance between $\mathbf{x}$ and $\mathbf{x}'$, i.e. $k(\mathbf{x}, \mathbf{x}')$ is a function of $||\mathbf{x}-\mathbf{x}'||$. Again, this may be written as an invariance of the kernel under the simultaneous transformation of both arguments:

$$k(R(\mathbf{x}), R(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}') \tag{11}$$

where R belongs this time to the more general class of isometries. Both latter invariances can in fact be seen as particular cases (with natural actions of groups of translations or isometries, respectively) of the following definition given by Parthasarathy and Schmidt in [33]:

Definition 11. *k is said invariant under the action of G on D when*

$$\forall g \in G, \forall \mathbf{x}, \mathbf{x}' \in D,\ \ k(g.\mathbf{x}, g.\mathbf{x}') = k(\mathbf{x}, \mathbf{x}') \tag{12}$$

3. Main results

3.1. A characterization of kernels leading to invariant fields.

Before stating the main result of the paper, we need to introduce a new notion, generalizing the notion of invariant kernel presented in the last section.

Definition 12. *A kernel k is said argumentwise invariant under Φ when*

$$\forall g, g' \in G, \forall \mathbf{x}, \mathbf{x}' \in D,\ \ k(g.\mathbf{x}, g'.\mathbf{x}') = k(\mathbf{x}, \mathbf{x}') \tag{13}$$

One can notice that eq.(12) corresponds to the particular case of eq.(13) where $g = g'$. As discussed next, this second kind of kernels corresponds to much stronger invariance properties of the associated random fields.

Remark 2. For real-valued symmetric kernels such as considered here, it is equivalent to be argumentwise invariant, left invariant, or right invariant. Indeed, assuming that k is left invariant, we get for $g, g' \in G$ and $\mathbf{x}, \mathbf{x}' \in D$:

$$k(g.\mathbf{x}, g'.\mathbf{x}') = k(\mathbf{x}, g'.\mathbf{x}') = k(g'.\mathbf{x}', \mathbf{x}) = k(\mathbf{x}', \mathbf{x}) = k(\mathbf{x}, \mathbf{x}') \tag{14}$$

Property 3.1. *(Kernels characterizing invariant fields) Let G be a finite group acting on D via the action Φ, and Y be a centered squared-integrable random field over D. Y has its paths Φ-invariant (up to a modification) if and only if its covariance kernel k is argumentwise invariant under Φ.*

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

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

Reciprocally, let us now assume that k_Y 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 $\widetilde{Y}$ by

$$\forall \mathbf{x} \in D\ \widetilde{Y}_{\mathbf{x}} := Y_{\pi_A(\mathbf{x})}$$

By construction, $\widetilde{Y}$ has all its 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,

$$\begin{aligned}\mathrm{var}[Y_{\mathbf{x}} - \widetilde{Y}_{\mathbf{x}}] &= \mathrm{var}[Y_{\mathbf{x}} - Y_{g.\mathbf{x}}] \\ &= k(\mathbf{x}, \mathbf{x}) + k(g.\mathbf{x}, g.\mathbf{x}) - 2k(\mathbf{x}, g.\mathbf{x}) = 0,\end{aligned}$$

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

Remark 3. $\widetilde{Y}_{\mathbf{x}} := \frac{1}{\#G} \sum_{g \in G} Y_{g.\mathbf{x}}$ *would have led to the same conclusion.*

Remark 4. *A fundamental domain A is such that every orbit has a unique representer in A, and $\bigcup_{g \in G} g.A = D$. However, the $g.A$'s ($g \in G$) are not necessarily disjoints. For example, if $G = \mathbb{Z}/2\mathbb{Z}$, $D = \mathbb{R}$, and $\Phi : (g, x) \in (\mathbb{Z}/2\mathbb{Z}) \times \mathbb{R} \longrightarrow \mathbb{R}$ is the action defined by $\Phi(\overline{1}, x) = -x$, $A = [0, +\infty)$ is a fundamental domain containing* 0, *but* $0 \in \overline{1}.A = (-\infty, 0]$ *too. Consequently, when decomposing an invariant process over the orbits of A, one must account for the points appearing in several $g.A$'s by dividing by the number of appearances, characterized by the cardinal of their stabilizers:*

$$\forall \mathbf{x} \in D,\ Y_{\mathbf{x}} = \sum_{g \in G} Y_{\mathbf{x}} \frac{\mathbf{1}_{g.A}(\mathbf{x})}{\#Stab_{\Phi}(\{\mathbf{x}\})} = \sum_{g \in G} Z_{g.\mathbf{x}} \tag{15}$$

where $Z_{\mathbf{x}} := Y_{\mathbf{x}} \frac{\mathbf{1}_A(\mathbf{x})}{\#Stab_{\Phi}(\{\mathbf{x}\})}$. *Denoting Z's kernel by k_Z, we get in particular*

$$\forall \mathbf{x}, \mathbf{x}' \in D\ k_Y(\mathbf{x}, \mathbf{x}') = cov\left[\sum_{g \in G} Z_{g.\mathbf{x}}, \sum_{g' \in G} Z_{g'.\mathbf{x}'}\right] = \sum_{(g, g') \in G^2} k_Z(g.\mathbf{x}, g'.\mathbf{x}'),$$

whereof the argumentwise invariance of k_Y clearly appears.

Example 1. *Let Z be a centered Gaussian process indexed by $\mathbb{R}$, with covariance kernel $k_Z : x, x' \in \mathbb{R} \longrightarrow k_Z(x, x') = e^{-|x-x'|} \in \mathbb{R}$ (often called the Ornstein-Uhlenbeck process, Cf. [35] sec. 1.3), and $\Phi : (g, x) \in (\mathbb{Z}/2\mathbb{Z}) \times \mathbb{R} \longrightarrow \mathbb{R}$ the action of $G = \mathbb{Z}/2\mathbb{Z}$ on $\mathbb{R}$ previously considered. The process Y obtained by symmetrization of Y's restriction to $A := [0, +\infty[$, defined by $Y_x = Z_{|x|}$, has all its paths invariant under Φ. Its covariance kernel is given by $\forall x, x' \in \mathbb{R},\ k_Y(x, x') = e^{-||x|-|x'||}$. Let us notice that Y, symmetrized of the stationary process Z, is obviously not second order stationary.*

Example 2. *Let Z be a centered Gaussian field indexed by $\mathbb{R}^2$, with covariance $k_Z : \mathbf{x}, \mathbf{x}' \in \mathbb{R}^2 \longrightarrow e^{-||\mathbf{x}-\mathbf{x}'||^2} \in \mathbb{R}$, and $\Phi : (g, \mathbf{x}) \in (\mathbb{Z}/2\mathbb{Z}) \times \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ the action defined by $\phi(\overline{1}, \mathbf{x}) = s(\mathbf{x}) := (x_2, x_1)$, the symmetrized point of $\mathbf{x} = (x_1, x_2)$ with respect to the first bisector. The process Y obtained by symmetrization of Z's restriction to $A = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \leq x_2\}$ is defined by*

$$Y_{\mathbf{x}} = \begin{cases} Z_{\mathbf{x}} \text{ if } \mathbf{x} \in A \\ Z_{s(\mathbf{x})} \text{ if } \mathbf{x} \in A^c \end{cases}$$

Let us note that Y may also be defined as follows, in the spirit of Remark 4:

$$Y_{\mathbf{x}} = \frac{1}{1 + \mathbf{1}_{\{\mathbf{x} \in \mathbb{R}^2 : s(\mathbf{x}) = \mathbf{x}\}}(\mathbf{x})} Z_{\mathbf{x}} \mathbf{1}_A(\mathbf{x}) + \frac{1}{1 + \mathbf{1}_{\{\mathbf{x} \in \mathbb{R}^2 : s(\mathbf{x}) = \mathbf{x}\}}(\mathbf{x})} Z_{\mathbf{x}} \mathbf{1}_A(s(\mathbf{x}))$$

The next example illustrates that a random field with almost never Φ-invariant paths may possess a modification which paths are all Φ-invariant:

Example 3. *Let $\Omega =]0, 1[$, $\mathcal{A} = \mathcal{B}(]0, 1[)$, $\mathbb{P}$ be Lebesgue's measure on Ω , $D = \mathbb{R}$, $G = \{e, s_0\}$ (s_0 be the symmetry with respect to 0), $F : x \in \mathbb{R} \longrightarrow \int_{-\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du \in]0, 1[$, $\varepsilon : \omega \in \Omega \longrightarrow \varepsilon(\omega) = F^{-1}(\omega) \in \mathbb{R}$, and $Y : (x, \omega) \in E \times \Omega \longrightarrow Y_x(\omega) = |x|\varepsilon(\omega)\mathbf{1}_{x \neq \varepsilon(\omega)}$. The process defined by $\widetilde{Y}_x(\omega) = |x|\varepsilon(\omega)$ has clearly all its paths invariant by s_0, and $\widetilde{Y}$ is a modification of Y since $\forall x \in D,\ P(Y_x = \tilde{Y}_x) = P(\varepsilon \neq x) = 1$. However, $\left\{\omega \in \Omega \ / \ (\forall x \in D,\ Y_x(\omega) = \tilde{Y}_x(\omega))\right\} = \left\{\frac{1}{2}\right\}$ is negligible, and the two processes are hence not indistinguable.*

Example 4. *Let us come back to the notations of example 2. One can construct a process having its paths Φ-invariant based on Z by defining $\forall \mathbf{x} \in D,\ Z^{\Phi}_{\mathbf{x}} = \frac{1}{2}(Z_{\mathbf{x}} + Z_{s(\mathbf{x})}) = \frac{1}{2}(Z_{(x_1,x_2)} + Z_{(x_2,x_1)})$. The covariance kernel of this new process is given by*

$$k_{Z^{\Phi}}(\mathbf{x}, \mathbf{x}') = \frac{1}{4}[k_Z(\mathbf{x} - \mathbf{x}') + k_Z(s(\mathbf{x}) - \mathbf{x}') + k_Z(\mathbf{x} - s(\mathbf{x}')) + k_Z(s(\mathbf{x}) - s(\mathbf{x}'))]$$

One may note that here, by isotropy, $||(x_1 - x_1', x_2 - x_2')|| = ||(x_2 - x_2', x_1 - x_1')||$ and $||(x_2 - x_1', x_1 - x_2')|| = ||(x_1 - x_2', x_2 - x_1')||$, so that

$$k_{Z^{\Phi}}(\mathbf{x}, \mathbf{x}') = \frac{1}{2}\left(e^{-||(x_1 - x_1', x_2 - x_2')||^2} + e^{-||(x_2 - x_1', x_1 - x_2')||^2}\right)$$

In more general cases (e.g., when Z has a geometrical anisotropy), however, it may be necessary to calculate a sum of four different terms.

3.2. **Kriging with an argumentwise invariant kernel.** Let us now come back to our original prediction problem, and assume that we dispose of n noiseless observations $Y_{\mathbf{x}_i} = y_i$ $(1 \leq i \leq n)$ of a square-integrable centered random field $(Y_{\mathbf{x}})_{\mathbf{x} \in D}$ assumed invariant under the action Φ of a finite group G on D. As recalled in the introduction (Eq. 1), the function

$$m : \mathbf{x} \in D \longrightarrow m(\mathbf{x}) = \mathbf{k}(\mathbf{x})^T K^{-1} \mathbf{y},$$

is the Simple Kriging predictor (or "Kriging mean") of Y knowing the reponses at design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$. In addition, the Simple Kriging variance (or "Mean Squared Error") s^2 is often used as a quantifyer of m's accuracy:

$$s^2 : \mathbf{x} \in D \longrightarrow s^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}(\mathbf{x})^T K^{-1} \mathbf{k}(\mathbf{x}). \tag{16}$$

It is well known that m interpolates the observations and s^2 vanishes at the design of experiments. As we will see now, more can be said in the case where k is argumentwise invariant.

Property 3.2. *(Properties of m and s^2 when k is argumentwise invariant)*

(1) *m and s^2 are invariant*

(2) $\forall i \in \{1, \ldots, n\}$, $\forall g \in G$, $m(g.\mathbf{x}_i) = y_i$ *and* $s^2(g.\mathbf{x}_i) = 0$.

Proof. The covariance vector $\mathbf{k}(.)$ is invariant by argumentwise invariance of k. Plugging in the equality $\mathbf{k}(g.\mathbf{x}) = \mathbf{k}(\mathbf{x})$ in Eqs. 1 and 16, (1) follows. (2) basically relies on (1) in the cases where $\mathbf{x} = g.\mathbf{x}_i$ $(i \in \{1, \ldots, n\})$. □

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

$$\mathrm{cov}(Y_{\mathbf{x}}, Y_{\mathbf{x}'} | Y_{\mathbf{X}} = \mathbf{y}) = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}(\mathbf{x})^T K^{-1} \mathbf{k}(\mathbf{x}'). \tag{17}$$

In the case where Y is assumed Gaussian, the Simple Kriging mean and variance at $\mathbf{x}$ coincide respectively with the conditional expectation and variance of $Y_{\mathbf{x}}$ knowing the observations. In addition, the Gaussian assumption makes it possible to get conditional simulations of Y, relying only on the conditional mean function and covariance kernel. The following property will play a crucial role in the applications discussed in the next section.

Property 3.3. *(Properties of the conditional distribution of a Gaussian Random Field with argumentwise invariant kernel)*

(1) *The conditional random field has an argumentwise invariant kernel*

(2) *All conditional simulations are Φ-invariant*

Proof. (1) follows from the invariance of $\mathbf{k}(.)$ applied to Eq.17. 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. 3.2. Since the complement to this mean function is a centered Gaussian Field with argumentwise invariant kernel (from (1)), Prop. 3.1 implies that the conditional simulations are Φ-invariant, as sums of a Φ-invariant function plus Φ-invariant paths. $\square$

Remark 5. *In practice, the paths of Y are often simulated at a finite set of points $\mathbf{X}_{simu} = \{\mathbf{e}_1, ..., \mathbf{e}_m\} \subset D$ based on a matrix decomposition (Cholesky, Mahalanobis) of $K = (k_Y(\mathbf{e}_i, \mathbf{e}_j))_{1 \leq i,j \leq m}$. The Φ-invariance of the vectors simulated that way is thus sure (i.e. $\forall \omega \in \Omega$).*

3.3. **What about the RKHS point of view?**

As a closure to the present section on the main results of the paper, let us briefly discuss the interplay between the argumentwise invariance of a p.d. kernel and the invariance of elements from the naturally associated RKHS of real-valued functions.

Property 3.4. *The Reproducing Kernel Hilbert Space $(\mathcal{H}, \langle ., . \rangle_{\mathcal{H}})$ with reproducing kernel k has all its functions Φ-invariant if and only if k is argumentwise invariant under Φ.*

Proof. If k is argumentwise invariant and $\mathcal{H}$ is a RKHS of real-valued functions with kernel k, it is clear that any function $f \in \mathcal{H}$ is invariant under Φ. Indeed, taking arbitrarily $\mathbf{x} \in D$ and $g \in G$, we get

$$f(g.\mathbf{x}) = \langle f, k(g.\mathbf{x}, .) \rangle_{\mathcal{H}} = \langle f, k(\mathbf{x}, .) \rangle_{\mathcal{H}} = f(\mathbf{x})$$

It clearly appears from that representation that the left invariance is sufficient. This is of course related to the fact that we work here with symmetric kernels in the first place (in the sense that $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$). For the reciprocal, assuming that any $f \in \mathcal{H}$ is invariant, it is straightforward that all $k(\mathbf{x}, .)$'s ($\mathbf{x} \in D$) are invariant since they belong to $\mathcal{H}$. Hence,

$$k(g.\mathbf{x}, g'.\mathbf{x}') = \langle k(g'.\mathbf{x}', .), k(g.\mathbf{x}, .) \rangle_{\mathcal{H}} = \langle k(\mathbf{x}', .), k(\mathbf{x}, .) \rangle_{\mathcal{H}} = k(\mathbf{x}, \mathbf{x}'),$$

which proves the argumentwise invariance of k. $\square$

Remark 6. *In the case where the Mercer theorem applies, the property speaks for itself. k then possesses an orthogonal expansion of the form*

$$k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{+\infty} \lambda_i e_i(\mathbf{x}) e_i(\mathbf{x}') \tag{18}$$

where the eigenfunctions $e_i(.)$ form an orthonormal basis of $L^2(D)$. Since the $e_i(.)$'s are in the RKHS, they are invariant themselves, and it then appears directly that k is argumentwise invariant.

4. Applications

4.1. **Invariant Brownian Motion and other elementary examples.**

4.1.1. *Symmetrized BM and OU process.* Let us first consider $(B_t)_{t\in[0,+\infty[}$, a one-dimensional Brownian Motion (BM), and the symmetry with respect to the origin $s : x \in D \longrightarrow -x \in D$, where $D := \mathbb{R}$. The corresponding action of the group $G = \mathbb{Z}/2\mathbb{Z}$ on $\mathbb{R}$ is the same as in Ex. 1. In order to symmetrize B, let us first extend it to a process on the whole line by setting $\forall t < 0,\ B_t = 0$. Now, relying on the fundamental domain $A := [0, +\infty[$, a straightforward way to symmetrize B is to construct $SB^{(1)}$ as follows:

$$SB_t^{(1)} = B_{\pi_A(t)} = B_{|t|} \tag{19}$$

The resulting process is still centered and Gaussian, with covariance

$$k_{SB^{(1)}}(t, t') := \mathrm{cov}(SB_t^{(1)}, SB_{t'}^{(1)}) = \mathrm{cov}(B_{|t|}, B_{|t'|}) = \min(|t|, |t'|) \tag{20}$$

Now, as we have seen in Ex.4, another way of getting a process with symmetric paths based on B is by averaging it over the action's orbits:

$$SB_t^{(2)} := \frac{1}{2}(B_t + B_{s(t)}) = \frac{1}{2}(B_t + B_{-t}) \tag{21}$$

In that case, following the way B was extended, we thus have

$$SB_t^{(2)} = \frac{1}{2} B_{|t|}, \tag{22}$$

so that $k_{SB^{(2)}} = \frac{1}{4} k_{SB^{(1)}}$. Simulated paths of the centered Gaussian process characterized by Eq. 20 are represented on Figure 1.

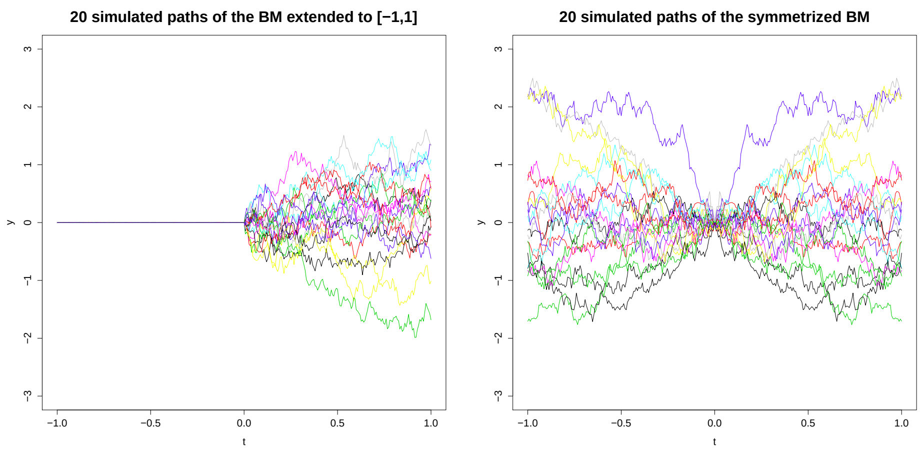

FIGURE 1. Symmetrization of the Brownian Motion relying on the symmetrized kernel (by projection on a fundamental domain) of Eq. 20.

Let us now consider an Orstein-Uhlenbeck (OU) process $(Z_t)_{t\in D}$ restricted to $D := [0, 1]$, and $s : t \in D \longrightarrow 1 - t \in D$ the symmetry with respect to $\frac{1}{2}$. This time, we choose $A := [0, \frac{1}{2}]$ as fundamental domain. A similar construction as for the first symmetrized BM leads to the process

$$Y_t^{(1)} = Z_{\pi_A(t)} = Z_{\min(t,s(t))} = Z_{\min(t,1-t)} \tag{23}$$

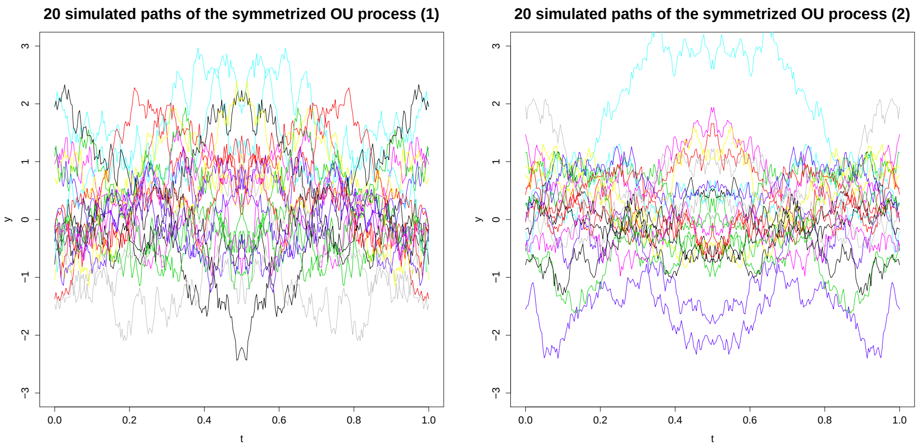

FIGURE 2. Symmetrization of the OU process relying on both kernels defined by Eqs. 24, 26. Left: by projection on a fundamental domain (Eq. 24). Right: by averaging over the orbits (Eq. 26).

This centered Gaussian process is then characterized by the kernel

$$\begin{aligned} k_{Y^{(1)}}(t,t') &:= \operatorname{cov}(Z_{\min(t,1-t)}, Z_{\min(t',1-t')}) \\ &= \exp\left(-|\min(t,1-t) - \min(t',1-t')|\right) \end{aligned} \tag{24}$$

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

$$Y_t^{(2)} = \frac{1}{2}(Z_t + Z_{s(t)}) = \frac{1}{2}(Z_t + Z_{1-t}), \tag{25}$$

and possesses the following covariance kernel:

$$\begin{aligned} k_{Y^{(2)}}(t,t') &= \frac{1}{4}\operatorname{cov}(Y_t + Y_{1-t}, Y_{t'} + Y_{1-t'}) \\ &= \frac{1}{4}\exp\left(-|t-t'|\right) + \frac{1}{4}\exp\left(-|(1-t)-t'|\right) \\ &+ \frac{1}{4}\exp\left(-|t-(1-t')|\right) + \frac{1}{4}\exp\left(-|(1-t)-(1-t')|\right) \\ &= \frac{1}{2}\exp\left(-|t-t'|\right) + \frac{1}{2}\exp\left(-|1-t-t'|\right) \end{aligned} \tag{26}$$

Simulated paths of the centered Gaussian process defined by both eq. 24 and eq. 26 are represented on figure 2.

4.1.2. *Conditional simulations of an invariant Gaussian Process.* We now assume that the invariant process $Y^{(2)}$ was observed at the 3 points $t_1 = 0.6, t_2 = 0.8, t_3 = 1$, with response values $y_1 = -0.8, y_2 = 0.5, y_3 = 0.9$. The covariance kernel of eq. 26 is used for performing simulations of $Y^{(2)}$ conditionally on the latter observations. 20 such conditional simulations are represented on Figure 3. As can be seen on Figure 3, all paths are simultaneously Φ-invariant and interpolating the conditioning data, hence

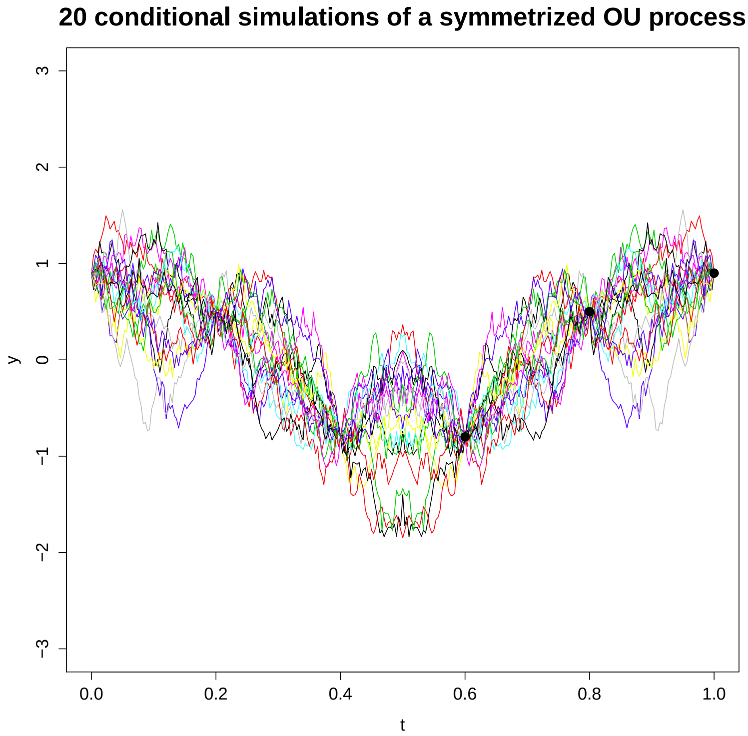

FIGURE 3. Conditional Simulations of the symmetrized OU process with the kernel of Eq. 26. The black points stand for the conditioning data.

illustrating Property 3.3 on the conditional distribution of Gaussian Random Fields with argumentwise invariant kernel.

4.2. **Kriging with an invariant kernel.** Let us now apply Kriging with an argumentwise invariant kernel to a benchmark example from the structural reliability literature exhibiting obvious symmetries.

Quoting [7] in which this test-case was recently used, "the example has been analyzed by [45] and [16] made a comparison with several meta-models proposed by [40]". The limit state function of interest reads:

$$y : (x_1, x_2) \in [-5, 5]^2 \longrightarrow \min \left\{ \begin{array}{l} 3 + 0.1(x_1 - x_2)^2 - (x_1 + x_2)/\sqrt{2} \\ 3 + 0.1(x_1 - x_2)^2 + (x_1 + x_2)/\sqrt{2} \\ (x_1 - x_2) + 6/\sqrt{2} \\ (x_2 - x_1) + 6/\sqrt{2} \end{array} \right\}$$

Figure 4 shows the contours of y, with an illustration of the three non-trivial transformations of $\mathbb{R}^2$ –denoted by s_1, s_2, s_3– leaving y invariant.
Actually, y can be shown to be left invariant by an action of the group $(\mathbb{Z}/2\mathbb{Z})^2$ on $\mathbb{R}^2$. Indeed, as illustrated on Figure 4, y is invariant under s_1, the axial symmetry with respect to the first bisector. y is also invariant

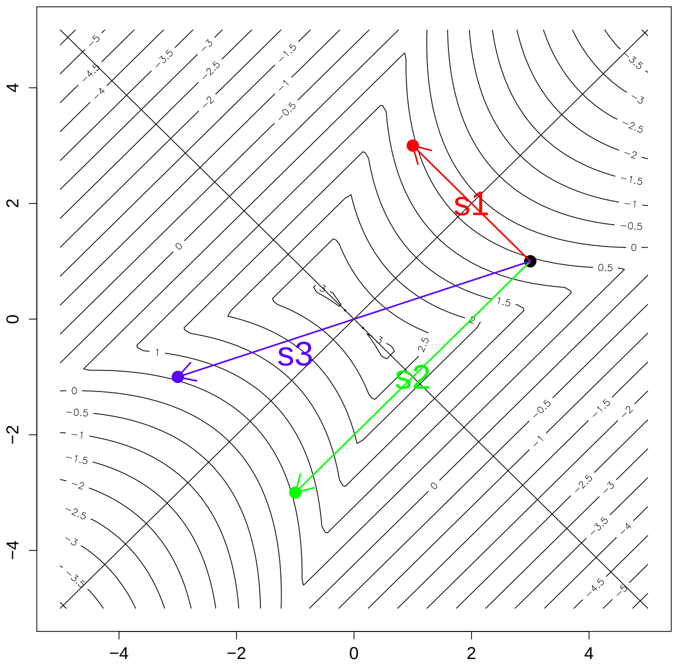

FIGURE 4. Borri and Speranzini's function, with its two axes of symmetry (black lines). The colored arrows stand for the three non-trivial transformations leaving this function unchanged.

under s_2, the axial symmetry with respect to the second bisector. Finally, y is obviously invariant under their composition, s_3, i.e. the symmetry with respect to the origin. Together with the identity of $\mathbb{R}^2$, denoted by s_0, the latter s_1, s_2, s_3 forms a group of order 4, representing $(\mathbb{Z}/2\mathbb{Z})^2$ on $\mathbb{R}^2$.

4.2.1. *Comparing three Kriging models based on different kernels.* Here we investigate using argumentwise invariant kernels for approximating this function by Simple Kriging based on 30 observations at a maximin LHS Design $\mathbf{X}$. The underlying Design of Experiments is generated using the R package *lhs*. As a preliminary step towards a comparison between different kernels, a classical Simple Kriging model with a tensor product OU kernel

$$k_Z(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{\theta}\left(|x_1 - x_1'| + |x_2 - x_2'|\right)\right) + \tau^2 \mathbf{1}_{\mathbf{x}=\mathbf{x}'} \qquad (27)$$

is fitted to the data (see Figure 5). Here the parameter are fixed to their Maximum Likelihood estimates, $\sigma^2 = 7.5$ and $\theta = 20$. In addition, a nugget effect with $\tau^2 = 0.01$ is added to k_Z for numerical purposes.

We now consider two different argumentwise invariant kernels. To start with, using similar notations as for the 1-dimensional OU example, we define a

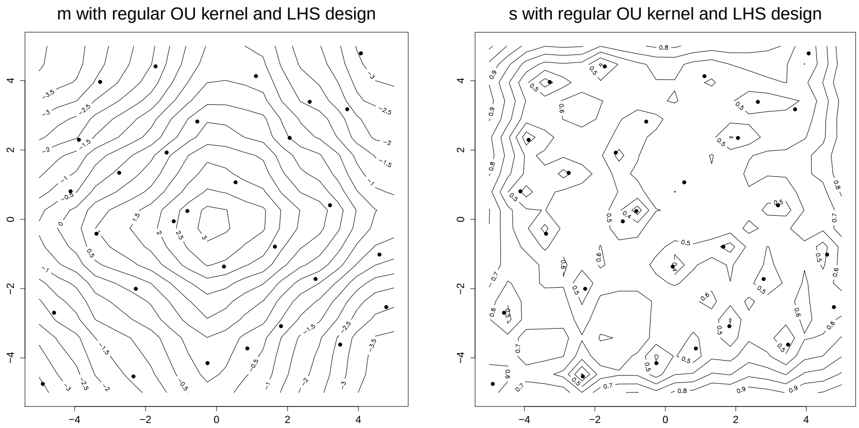

FIGURE 5. Simple Kriging mean and standard deviation with the regular tensor product kernel of Eq. 27, based on observations of y at $\mathbf{X}$.

fundamental domain for Φ (see Figure 6 for an illustration):

$$A := \{\mathbf{x} \in [-5,5]^2 : x_1 \geq 0, -x_1 < x_2 \leq x_1\} \tag{28}$$

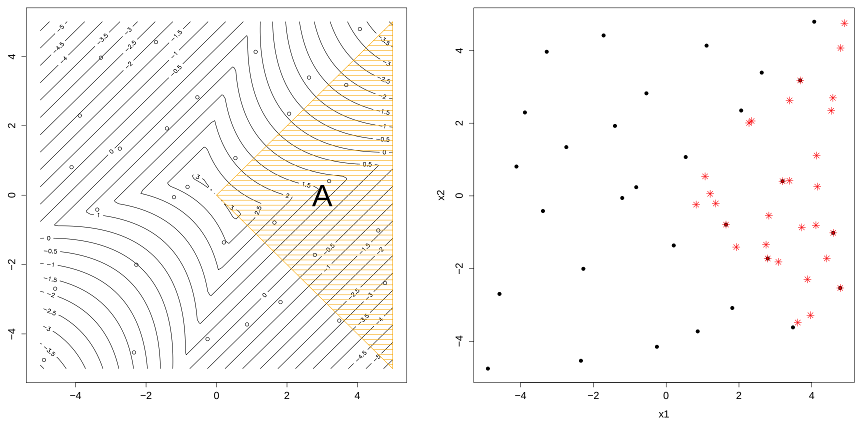

FIGURE 6. Left: The fundamental domain A for the considered action of $(\mathbb{Z}/2\mathbb{Z})^2$ on $[-5,5]^2$. The lower left boundary is excluded. Right: Projection of the design points of $\mathbf{X}$ (black points) onto A (red stars).

The first argumentwise invariant kernel considered is then constructed based on the projector $\pi_A : \mathbf{x} \in D \longrightarrow \pi_A(\mathbf{x}) = \mathcal{O}(\mathbf{x}) \cap A \in A$, as follows:

$$k_{Y^{(1)}}(\mathbf{x}, \mathbf{x}') := k_Z(\pi_A(\mathbf{x}), \pi_A(\mathbf{x}')) \tag{29}$$

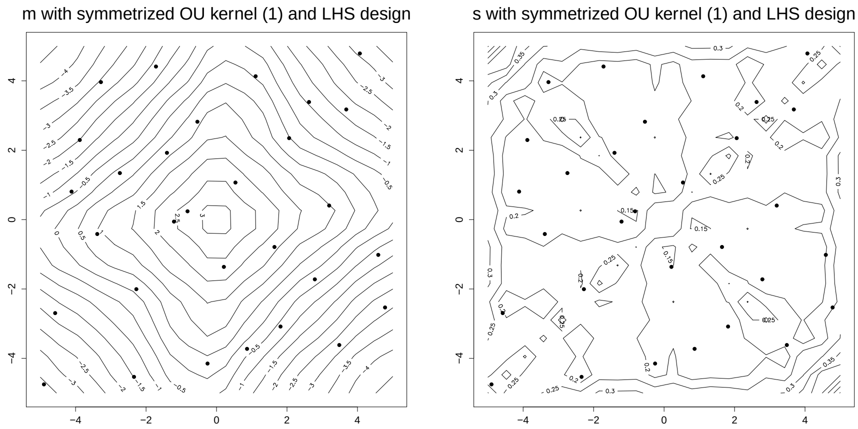

FIGURE 7. Simple Kriging mean and standard deviation with the symmetrized OU kernel of Eq. 29, based on observations of y at $\mathbf{X}$.

The second argumentwise invariant kernel considered is then constructed by averaging k_Z over the orbits of Φ:

$$k_{Y^{(2)}}(\mathbf{x}, \mathbf{x}') := \frac{1}{16} \sum_{i=0}^{3} \sum_{j=0}^{3} k_Z(s_i.\mathbf{x}, s_j.\mathbf{x}') \tag{30}$$

The results of Kriging with kernels $k_{Y^{(1)}}$ and $k_{Y^{(2)}}$ based on the observations at $\mathbf{X}$ are illustrated on Figures 7 and 8, respectively.

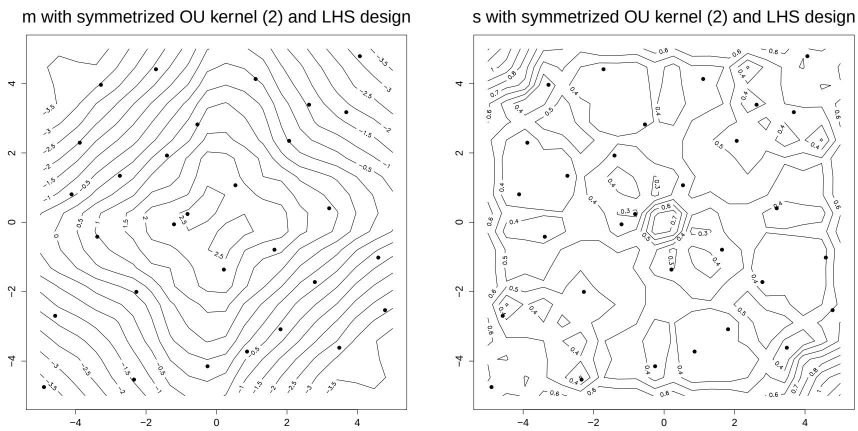

FIGURE 8. Kriging mean and standard deviation with the symmetrized OU kernel of Eq. 30, based on observations of y at $\mathbf{X}$.

Finally, for comparison, a Kriging model with regular OU kernel (the same as for the first model) but based on the design

$$\mathbf{X}_{\text{sym}} := \bigcup_{i=0}^{3} s_i.\mathbf{X} \tag{31}$$

and with the observations at $\mathbf{X}$ replicated four times is considered.

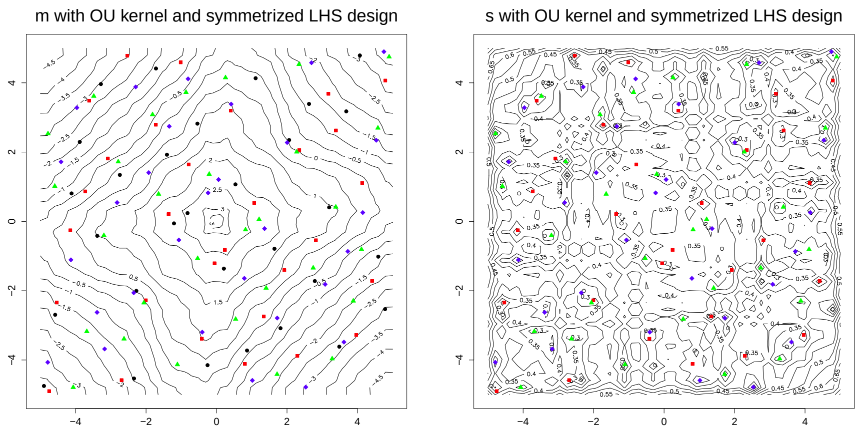

FIGURE 9. Kriging mean and standard deviation with the regular tensor product kernel of Eq. 27, based on observations of y at $\mathbf{X}_{\text{sym}}$. The solid black circles represent the LHS design $\mathbf{X}$. The red squares, blue diamonds, and green triangles represent respectively the orbits of $\mathbf{X}$ under the transformations s_1, s_2, and s_3.

4.2.2. *Discussion on the compared results.* In order to compare the prediction abilities of the considered Kriging models, we predicted y at a 50×50 out-of-sample validation design $\mathbf{X}_{\text{val}}$ using the four models, and compared the average prediction errors and the residuals. Figure 10 represents the mean predictions against reality (first line) and the standardized residuals

$$\frac{y(\mathbf{x}_{\text{val}}^{(i)}) - m(\mathbf{x}_{\text{val}}^{(i)})}{s(\mathbf{x}_{\text{val}}^{(i)})} \quad (1 \leq i \leq 2500) \tag{32}$$

Looking at the values of the Integrated Squared Error (ISE) at $\mathbf{X}_{\text{val}}$,

$$\text{ISE} = \sum_{j=1}^{2500} (y(\mathbf{x}_{\text{val}}^{(i)}) - m(\mathbf{x}_{\text{val}}^{(i)}))^2 \tag{33}$$

for the four candidate Kriging models, we first see that the first model is undoubtly dominated by the three other ones. This was to be expected since the first model is the only one which doesn't take into account the symmetry of the problem. The second model, based on a combination of the OU kernel with the projector onto the fundamental domain of Figure 6, shows

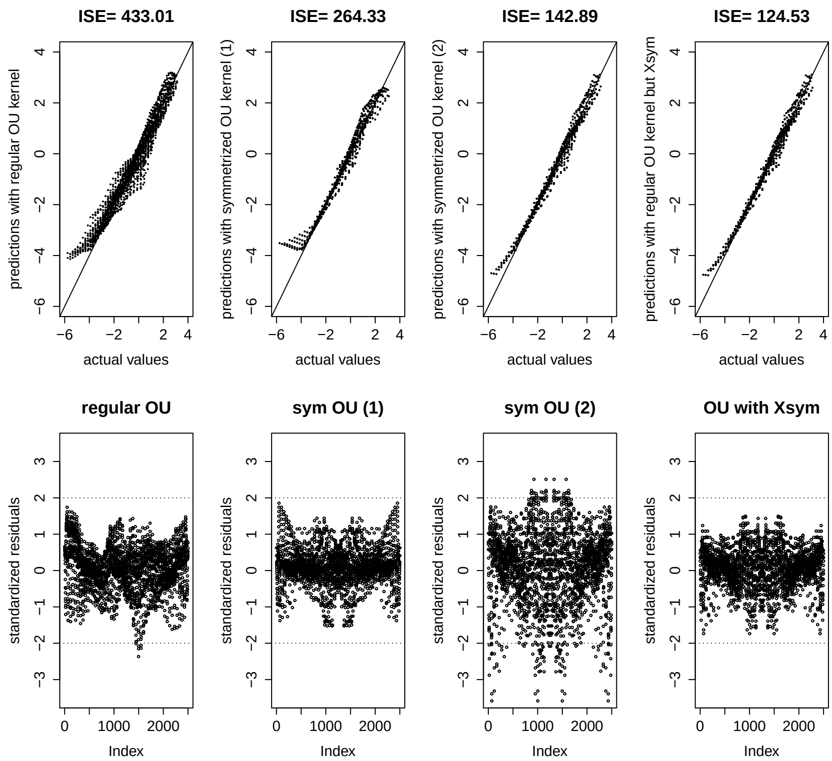

FIGURE 10. Comparison of prediction results at $\mathbf{X}_{\text{val}}$ when using the 4 Kriging models considered for Borri and Speranzini's function.

significantly better performances. Indeed, the ISE drops from 433.01 to 264.33, just by playing on the underlying kernel. However, the performances of the third model, with the kernel averaged over the action's orbits, are even better. Not only does the ISE drop to 142.89, but the order of magnitude of the standardized residuals is more in accordance with what one usually expects when Kriging under the Gaussian Process assumption (even if this is not really theoretically well-founded without further ergodicity assumptions, it is customary to expect that about 95% of the sample of standardized residuals lie between the 2.5% and 97.5% quantiles of the standard Gaussian distribution, see for instance [24]).

Perhaps surprisingly, the last model obtained by using a regular covariance kernel with a symmetrized design gave here better performances in terms of ISE (124.53) than the two previous models with argumentwise invariant kernels. This has to be tempered by the fact that doing it this way multiplies the dimension of the covariance matrix by the order of the group (i.e. 4 here), that is to say that the total number of coefficients jumps from n^2 to $n^2 \times r^2$ (i.e. from 900 to $14'400$ here). Hence, replicating the design is likely to cause problems in terms of matrix inversion, and even in terms of data storage (for the reasonable values $n = 1000$ and $r = 8$, $n^2 \times r^2 = 64'000'000$). Furthermore, the test function studied here is not very smooth (so that an OU kernel was considered instead of a Gauss or a Matérn one, more commonly used in smoother cases), which may relatively hinder the benefits of taking symmetries into account, since the latter come in more regular cases with additional smoothness properties on the axes of symmetry. Concerning the second model, let us also remark that the choice of A is arbitrary, and not always without consequences on the model obtained. In the case of an anisotropic covariance, for instance, choosing the current A or its image by a rotation of center $\mathbf{0}$ and angle $\frac{\pi}{2}$ may lead to substantially different predictions. This has to be studied in more detail in further works.

To finish with this application, let us point out the fact that among the considered models, only the ones based on an argumentwise invariant covariance kernel enables conditional simulations with invariant paths. 4 such simulated paths with the kernel of Eq. 29 conditional on the observations at $\mathbf{X}$ are represented on Figure 11.

4.3. **Kriging with a wrong kind of invariant kernel.** We now simulate realizations of a two-dimensional centered Gaussian Random Field with argumentwise invariant kernel, and compare the predictive performances obtained by using Kriging models with different configurations:

- Model A: Symmetrized kernel obtained by projection over a fundamental domain (analogue of the kernel defined in Eq. 29)
- Model B: Symmetrized kernel obtained by double sum over the orbits (analogue of the kernel defined in Eq. 30)
- Model C: Stationary kernel with symmetrized design (Same design as defined in Eq. 31)

However, contrarily to the previous example and in order to investigate a different class of fields for which invariances and the different kinds of symmetrization may have a more crucial impact, we chose here for the stationary kernel k_Z underlying the three models above an anisotropic Gaussian kernel:

$$k_Z(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\left(\frac{(x_1 - x_1')^2}{20} + \frac{(x_2 - x_2')^2}{10}\right)\right) + \tau^2 \mathbf{1}_{\mathbf{x}=\mathbf{x}'}, \qquad (34)$$

still with $\sigma^2 = 7.5$ and $\tau^2 = 0.01$. The same design of experiments $\mathbf{X}$ (up to a symmetrization in the case of model C) was used for the three models. By the way, we took the same 30-points LH Design as for the previous example.

Before analyzing statistical performance results, let us focus on the particular example addressed in Figure 12. Here, as in the other results presented in this section, the underlying model of the simulated Gaussian Field is Model B. On the right hand side of Figure 12, in the first line, one can observe that the Kriging mean surfaces obtained with the three different kernels look rather similar. Looking at the Integrated Square Error values, however, one

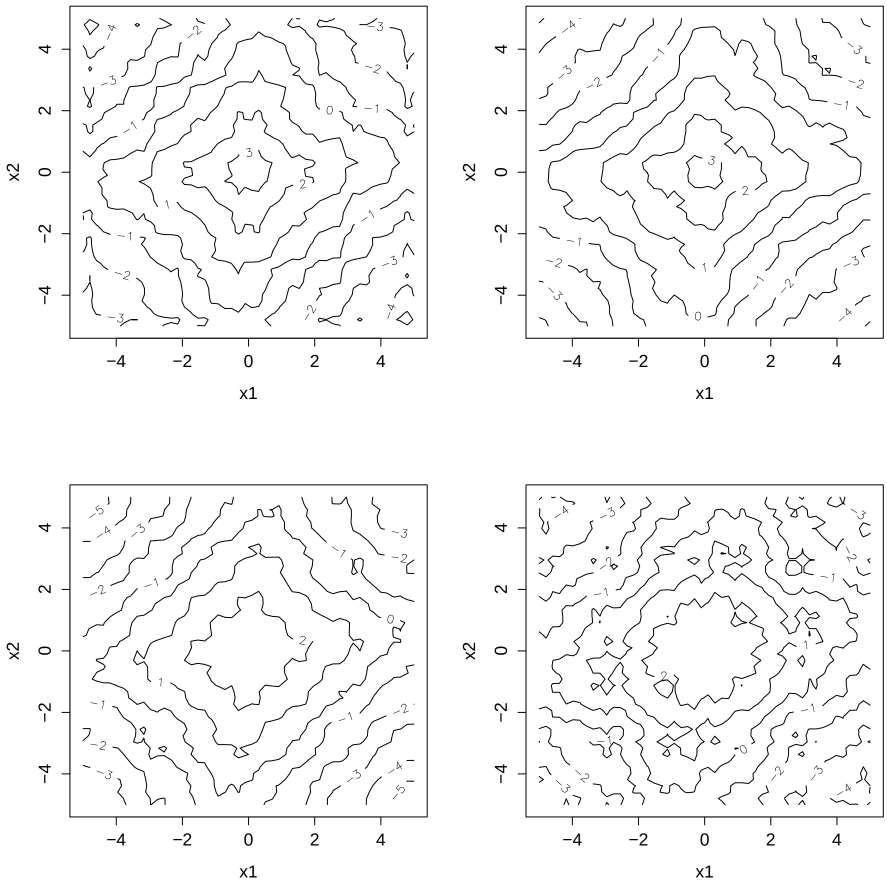

FIGURE 11. Four conditional simulations of Borri and Speranzini's function with symmetrized OU kernel (1), based on observations at $\mathbf{X}$.

discovers a clear distinction in favour of Kriging with the kernel of Model B (ISE $= 1.93$), compared with the ones of Model C (ISE $= 3.16$) and Model A (ISE $= 3.44$). Furthermore, a visual inspection of the standardized residuals leads to a similar distinction: While the standardized residuals obtained with the kernel of Model B look in accordance with a Gaussian assumption, with a majority of points between -2 and 2 and a few values outside this interval, the standardized residuals obtained with the two other kernels are rather confined to the interval $[-1, 1]$, with a few outsiders. This is quantitatively confirmed by a Kolmogorov-Smirnov test, which statistic values (denoted D_n) are respectively given by 0.078 (B), 0.171 (C), and 0.18 (A).

Repeating the experiments with 100 simulated Gaussian Field realizations (corresponding to Model B), we obtain boxplots for the (logarithm of the) ISE values (Figure 13, left) and for the D_n values (Figure 13, right). The

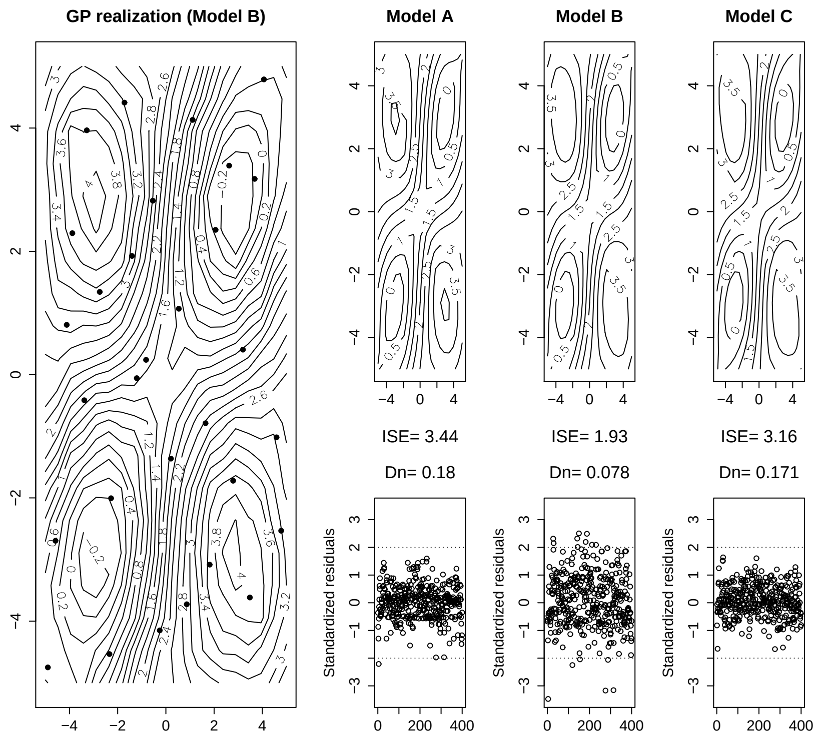

FIGURE 12. Left: Simulation of a Gaussian Field with argumentwise invariant kernel (Model B), and learning set $\mathbf{X}$ (black points). Right, for each of the three columns: Kriging mean (above) and standardized residuals when predicting at $\mathbf{X}_{\text{val}}$ with Kriging models A, B, and C. The ISE values quantify the departure of the Kriging mean from the actual function. The D_n values (Kolmogorov's statistic) quantify the departure from normality of the standardized residuals.

superiority of Model B in coverage (D_n statistic) clearly appears, the departure from normality of the standardized residuals being much larger for Model A and Model C. However, note that, even for Model B, the empirical distribution of standardized residuals slightly differs from a standard Gaussian one: The D_n values are often higher than expected at confidence levels 0.05 or 0.01 (Figure 13, right hand side, horizontal dotted lines). The superiority of Model B is also confirmed in prediction (ISE criterion). Visible on the boxplots by checking that the notches are not overlapping (Figure 13, left), it can be checked in a quantitative way by means of a two-sample Wilcoxon test: For instance, the null hypothesis "ISE (Model B) = ISE (Model C)" with the one-sided alternative "ISE (Model B) $\leq$ ISE (Model C)" is strongly rejected with a p-value of $4.344e-06$.

These results confirm the first impressions left by the the particular case commented above and illustrated in Figure 12: The performances of Kriging

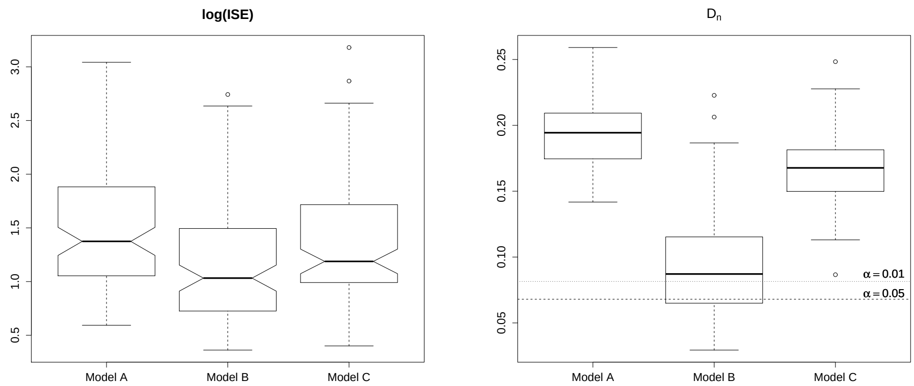

FIGURE 13. Left: log-ISE values (at $\mathbf{X}_{\text{val}}$) when using the three Kriging models A, B, C for predicting 100 Gaussian Field realizations generated with Model B. Right: Kolmogorov-Smirnov test statistic values for comparing the standardized residuals to a standard Gaussian distribution.

with an argumentwise invariant kernel are sensitive to the adequacy of the kernel chosen to the kind of invariant process underlying the data, at the same time in terms of prediction accuracy and of coverage. In particular, approximating an invariant Gaussian Field with a Kriging model relying on a symmetrized kernel obtained by projection over a fundamental domain may give very different results than using a "double sum over the orbits" kernel, even if the underlying stationary kernel k_Z is the same in both cases.

5. Conclusion and perspectives

We proposed a class of covariance kernels, called *argumentwise invariant kernels*, characterizing (up to a modification) squared integrable random fields with invariant paths under an arbitrary action of a finite group on the index set, as well as Reproducing Kernel Hilbert Spaces of invariant functions (still with a finite group acting on the source space).

These kernels can be used for different purposes. We focused here on modeling invariant functions by Kriging. As discussed along the paper, Kriging models with an argumentwise invariant kernel have interesting properties, including the invariance of both Kriging mean and variance functions, but also the invariance of paths emanating from conditional simulations.

Among the two variants for making up invariant kernels based on arbitrary kernels proposed in the last section, summing a kernel over the orbits of the considered group action gave more convincing results than composing the basis kernel with a projection onto a fundamental domain. However, this may not hold in the general case, and further works may focus on identifying and unlocking the potential weak points of both considered approaches.

Acknowledgements: The authors would like to thank Anestis Antoniadis for a decisive question on the non necessary symmetric paths of a version of a process with symmetric paths, and Yann Richet (IRSN) for having provided them with a physical application rising the question of embedding symmetry properties within Kriging. Many thanks as well to Alain Valette for his precious advise in algebra, and to Cédric Boutillier who helped improving a former version of the present paper. Last but not least, this paper wouldn't exist in its present without the contribution of Yves Deville, who helped creating a replicate of the DiceKriging package allowing to extend the use of the prediction and simulation methods to arbitrary classes of kernels. Finally, the authors would like the anonymous referee and associate editor for their constructive suggestions and comments.

References

[1] P. Abrahamsen. A review of Gaussian random fields and correlation functions, second edition. Technical report, Norwegian Computing Center, 1997.

[2] R.J. Adler and J.E. Taylor. *Random Fields and Geometry*. Springer, Boston, 2007.

[3] S. Amari and H. Nagaoka. *Transactions of mathematical monographs: methods of information geometry*, volume 191. Oxford University Press, 2000.

[4] E. B. Anderes. *Estimating deformations of isotropic Gaussian random fields*. PhD thesis, Univ. Chicago., 2005.

[5] E.B. Anderes and S. Chatterjee. Consistent estimates of deformed isotropic Gaussian random fields on the plane. *The Annals of Statistics*, 37, No. 5A:2324–2350, 2009.

[6] E.B. Anderes and M.L. Stein. Estimating deformations of isotropic Gaussian random fields on the plane. *The Annals of Statistics*, Vol. 36, No. 2:719–741, 2008.

[7] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 2011.

[8] B. Bekka, P. de la Harpe, and A. Valette. *Kazhdan's property (T)*. Cambridge University Press, 2008.

[9] C. Berg, J.P.R. Christensen, and P. Ressel. *Harmonic Analysis on Semigroups*. Springer, 1984.

[10] A. Berlinet and C. Thomas Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, 2003.

[11] H. Cartan and R. Godement. Théorie de la dualité et analyse harmonique dans les groupes abéliens localement compacts. *Annales Scientifiques de l'Ecole Normale Supérieure*, 64:79–99, 1947.

[12] Harald Cramér and M. R. Leadbetter. *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*. Dover, 1967.

[13] N.A.C. Cressie. *Statistics for spatial data*. Wiley series in probability and mathematical statistics, 1993.

[14] F.C. Curriero. On the use of non-euclidean isotropy in geostatistics. Technical report, Johns Hopkins University, 2005.

[15] M.L. Eaton. *Group invariance applications in statistics*. Regional Conference Series in Probability and Statistics Volume 1, 1989.

[16] B. Echard, N. Gayton, and M. Lemaire. 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*, 2010.

[17] C. Gaetan and X. Guyon. *Modélisation et Statistique Spatiales*. Springer, 2008.

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

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

[20] P. Guttorp and P.D. Sampson. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87, No. 417:108–119, 1992.

[21] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.

[22] D.T. Hristopoulos. Spatial random field models inspired from statistical physics with applications in the geosciences. *Physica A*, (365):211–216, June 2006.

[23] J. Istas. Manifold indexed fractional fields. *ESAIM P&S*, 2011.

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

[25] D.G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *J. of the Chem., Metal. and Mining Soc. of South Africa*, 52 (6):119139, 1951.

[26] S. Lang. *Algebra*. Addison-Wesley, Reading, Mass., 1965.

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

[28] P. Lévy. *Processus stochastiques et mouvement Brownien*. 1965.

[29] G. Matheron. La théorie des variables régionalisées et ses applications. Technical report, Centre de Morphologie Mathématique de Fontainebleau, Ecole Nationale Supérieure des Mines de Paris, 1970.

[30] A. O'Hagan. Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering and System Safety*, 91(91):1290–1300, 2006.

[31] C. J. Paciorek and M. J. Schervish. Nonstationary covariance functions for Gaussian process regression. In *Advances in Neural Information Processing Systems 16 (NIPS)*, 2004.

[32] C.J. Paciorek. *Nonstationary Gaussian Processes for Regression and Spatial Modelling*. PhD thesis, Carnegie Mellon University, 2003.

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

[34] C.E. Rasmussen and K.I. Williams. *Gaussian Processes for Machine Learning*. M.I.T. Press, 2006.

[35] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer-Verlag, 1991.

[36] T.J. Santner, B.J. Williams, and W.J. Notz. *The Design and Analysis of Computer Experiments*. Springer, 2003.

[37] A. M. Schmidt and A. O'Hagan. Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society B*, 65(3):745–758, 2003.

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

[39] B. Schölkopf. The kernel trick for distances. In *Neural Information Processing Systems*, 2000.

[40] L. Schueremans and D. Van Gemert. Benefit of splines and neural networks in simulation based structural reliability analysis. *Structural safety*, 27(3):246–261, 2005.

[41] M.L. Stein. *Interpolation of Spatial Data, Some Theory for Kriging*. Springer, 1999.

[42] V.N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.

[43] E. Vazquez. *Modélisation Comportementale de Systèmes Non-linéaires Multivariables par Méthodes à Noyaux et Applications*. PhD thesis, Université Paris XI Orsay, 2005.

[44] J.-P. Vert. *Kernel Methods*. Centre for Computational Biology, Ecole des Mines de Paris, 2007.

[45] P. H. Waarts. *Structural reliability using finite element methods: an appraisal of DARS*. PhD thesis, Delft University of Technology, 2000.

[46] G. Wahba. *Spline Models for Observational Data*. Siam, 1990.

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

David Ginsbourger, University of Bern, Institute of Mathematical Statistics and Actuarial Science, Alpeneggstrasse 22, CH-3012 Bern, Switzerland, david.ginsbourger@stat.unibe.ch

Xavier Bay and Olivier Roustant, Ecole Nationale Supérieure des Mines, Fayol-EMSE, LSTI, 158 cours Fauriel, F-42023 Saint-Etienne, France, roustant@emse.fr; bay@emse.fr

Laurent Carraro, Télécom Saint-Etienne, 25 rue du Docteur Rémy Annino, F-42000 Saint-Etienne, France, laurent.carraro@telecom-st-etienne.fr