On ANOVA decompositions of kernels and Gaussian random field paths

David Ginsbourger, Olivier Roustant, Dominic Schuhmacher, Nicolas Durrande, Nicolas Lenz

On ANOVA decompositions of kernels and Gaussian random field paths

D. Ginsbourger, O. Roustant, D. Schuhmacher, N. Durrande, and N. Lenz

Abstract The FANOVA (or "Sobol'-Hoeffding") decomposition of multivariate functions has been used for high-dimensional model representation and global sensitivity analysis. When the objective function f has no simple analytic form and is costly to evaluate, a practical limitation is that computing FANOVA terms may be unaffordable due to numerical integration costs. Several approximate approaches relying on random field models have been proposed to alleviate these costs, where f is substituted by a (kriging) predictor or by conditional simulations. In the present work, we focus on FANOVA decompositions of Gaussian random field sample paths, and we notably introduce an associated kernel decomposition (into 2^{2d} terms) called KANOVA. An interpretation in terms of tensor product projections is obtained, and it is shown that projected kernels control both the sparsity of Gaussian random field sample paths and the dependence structure between FANOVA effects. Applications on simulated data show the relevance of the approach for designing new classes of covariance kernels dedicated to high-dimensional kriging.

1 Introduction: Metamodel-based Global Sensitivity Analysis

Global Sensitivity Analysis (GSA) is a topic of importance for the study of complex systems as it aims at uncovering among many candidates which variables and inter-

David Ginsbourger · Nicolas Lenz
Department of Mathematics and Statistics, University of Bern, Alpeneggstrasse 22, CH-3012 Bern, Switzerland. e-mail: ginsbourger@stat.unibe.ch e-mail: nicolas.lenz@gmail.com

Olivier Roustant · Nicolas Durrande
École Nationale Supérieure des Mines, FAYOL-EMSE, LSTI, F-42023 Saint-Etienne, France. e-mail: roustant@emse.fr e-mail: durrande@emse.fr

Dominic Schuhmacher
Institut für Mathematische Stochastik, Georg-August-Universität Göttingen, Goldschmidtstraße 7, D-37077 Göttingen, Germany. e-mail: dominic.schuhmacher@mathematik.uni-goettingen.de

actions thereof are influential with respect to some response of interest. FANOVA (Functional ANalysis Of VAriance) [12, 30, 9, 2] has become commonplace for decomposing a real-valued function f of d-variables into a sum of 2^d functions (a.k.a. *effects*) of increasing dimensionality, and quantifying the influence of each variable or group of variables through non-negative indices summing up to one, the celebrated *Sobol' indices* [31, 25]. In practice f is rarely known analytically and a number of statistical procedures have been proposed for estimating Sobol' indices based on a finite sample of evaluations of f; see [13] and the references therein. Alternatively, a pragmatic approach to GSA, when the evaluation budget is drastically limited by computational cost or time, is to first approximate f by employing some class of surrogate models (e.g., regression, neural nets, splines, wavelets, kriging; see [35] for an overview) and then to perform the analysis on the obtained cheap-to-evaluate surrogate model. Here we focus essentially on kriging and Gaussian random field (GRF) models, with an emphasis on the interplay between covariance kernels and FANOVA decompositions of corresponding centred GRF sample paths.

While screening and GSA relying on kriging have been used for at least two decades [38], probabilistic GSA in the Bayesian set-up seems to originate in [22], where posterior effects and related quantities were derived under a GRF prior. Later on, posterior distributions of Sobol' indices were investigated in [20] relying on conditional simulations, an approach revisited and extended to multi-fidelity computer codes in [18]. From a different perspective, FANOVA-graphs were used in [21] to incorporate GSA information into a kriging model, and a special class of kernels was introduced in [6] for which Sobol' indices of the kriging predictor are analytically tractable. Moreover, a class of kernels leading to centred GRFs with additive paths has been discussed in [5], and FANOVA decompositions of GRFs and their covariance were touched upon in [19] where GRFs with ortho-additive paths were introduced. More recently, a variant of the kernel investigated in [6] was revisited in [4] with a focus on GSA with dependent inputs, and a class of kernels related to ANOVA decompositions was studied in [8, 7]. In a different setting, GRF priors have been used for Bayesian FANOVA when the responses are curves [14].

In the present paper we investigate ANOVA decompositions both for (symmetric positive definite) kernels and for associated centred GRFs. We show that under standard integrability conditions, s.p.d. kernels can be decomposed into 4^d terms that govern the joint distribution of the 2^d terms of the associated GRF FANOVA decomposition. This has some serious consequences in kriging-based GSA, as for instance the choice of a sparse kernel induces almost sure sparsity of the associated GRF paths, and that such phenomenon cannot be compensated by data acquisition.

2 Preliminaries and notation

FANOVA. We focus on measurable $f : D \subseteq \mathbb{R}^d \longrightarrow \mathbb{R}$ ($d \in \mathbb{N}\backslash\{0\}$). In FANOVA with independent inputs, D is typically assumed to be of the form $D = \prod_{i=1}^{d} D_i$ for some measurable subsets $D_i \in \mathscr{B}(\mathbb{R})$, where each D_i is endowed with a probability

measure ν_i and D is equipped with the product measure $\nu = \bigotimes_{i=1}^d \nu_i$. Assuming further that f is square-integrable w.r.t. ν, f can be expanded into as sum of 2^d terms indexed by the subsets $\mathfrak{u} \subseteq I = \{1,\ldots,d\}$ of the d variables,

$$f = \sum_{\mathfrak{u} \subseteq I} f_{\mathfrak{u}}, \tag{1}$$

where $f_{\mathfrak{u}} \in \mathscr{F} = \mathrm{L}^2(\nu)$ depend only on the variables x_j with $j \in \mathfrak{u}$ (up to an a.e. equality, as all statements involving L^2 from Equation (1) on). Uniqueness of this decomposition is classically guaranteed by imposing that $\int f_{\mathfrak{u}}\, \nu_j(\mathrm{d}x_j) = \mathbf{0}$ for every $j \in \mathfrak{u}$, in which case the FANOVA *effects* $f_{\mathfrak{u}}$ can be expressed in closed form as

$$f_{\mathfrak{u}} : \mathbf{x} \in D \longrightarrow f_{\mathfrak{u}}(x_1,\ldots,x_d) = \sum_{\mathfrak{u}' \subseteq \mathfrak{u}} (-1)^{|\mathfrak{u}|-|\mathfrak{u}'|} \int f(x_1,\ldots,x_d)\, \nu_{-\mathfrak{u}'}(\mathrm{d}\mathbf{x}_{-\mathfrak{u}'}), \tag{2}$$

where $\nu_{-\mathfrak{u}'} = \bigotimes_{j \in I \setminus \mathfrak{u}'} \nu_j$ and $\mathbf{x}_{-\mathfrak{u}'} = (x_i)_{i \in I \setminus \mathfrak{u}'}$. As developed in [17], Equation (2) is a special case of a decomposition relying on commuting projections. Denoting by $P_j : f \in \mathscr{F} \longrightarrow \int f \mathrm{d}\nu_j$ the orthogonal projector onto the subspace $\mathscr{F}_j$ of $f \in \mathscr{F}$ not depending on x_j, the identity on $\mathscr{F}$ can be expanded as

$$I_{\mathscr{F}} = \prod_{j=1}^d \left[(I_{\mathscr{F}} - P_j) + P_j\right] = \sum_{\mathfrak{u} \subseteq I} \left(\prod_{j \in \mathfrak{u}} (I_{\mathscr{F}} - P_j)\right)\left(\prod_{j \in I \setminus \mathfrak{u}} P_j\right). \tag{3}$$

FANOVA effects appear then as images of f under the orthogonal projection operators onto the associated subspaces $\mathscr{F}_{\mathfrak{u}} = \left(\bigcap_{j \notin \mathfrak{u}} \mathscr{F}_j\right) \cap \left(\bigcap_{j \in \mathfrak{u}} \mathscr{F}_j^{\perp}\right)$, i.e. we have that $f_{\mathfrak{u}} = T_{\mathfrak{u}}(f)$, where $T_{\mathfrak{u}} = \left(\prod_{j \in \mathfrak{u}} (I_{\mathscr{F}} - P_j)\right)\left(\prod_{j \notin \mathfrak{u}} P_j\right)$. Finally, the squared norm of f decomposes by orthogonality as $\|f\|^2 = \sum_{\mathfrak{u} \subseteq I} \|T_{\mathfrak{u}}(f)\|^2$ and the influence of each (group of) variable(s) on f can be quantified via the Sobol' indices

$$S_{\mathfrak{u}}(f) = \frac{\|T_{\mathfrak{u}}(f - T_{\emptyset}(f))\|^2}{\|f - T_{\emptyset}(f)\|^2} = \frac{\|T_{\mathfrak{u}}(f)\|^2}{\|f - T_{\emptyset}(f)\|^2}, \quad \mathfrak{u} \neq \emptyset. \tag{4}$$

Gaussian random fields (GRFs). A random field indexed by D is a collection of random variables $Z = (Z_{\mathbf{x}})_{\mathbf{x} \in D}$ defined on a common probability space $(\Omega, \mathscr{A}, \mathbb{P})$. The random field is called a Gaussian random field (GRF) if $(Z_{\mathbf{x}^{(1)}}, \ldots, Z_{\mathbf{x}^{(n)}})$ is n-variate normally distributed for any $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)} \in D$ ($n \geq 1$). The distribution of Z is then characterized by its mean function $m(\mathbf{x}) = \mathbb{E}[Z_{\mathbf{x}}]$, $\mathbf{x} \in D$, and covariance function $k(\mathbf{x},\mathbf{y}) = \mathrm{Cov}(Z_{\mathbf{x}}, Z_{\mathbf{y}})$, $\mathbf{x},\mathbf{y} \in D$. It is well-known that admissible covariance functions coincide with symmetric positive definite (s.p.d.) kernels on $D \times D$ [3].

A *multivariate* GRF taking values in $\mathbb{R}^p$ is a collection of $\mathbb{R}^p$-valued random vectors $Z = (Z_{\mathbf{x}})_{\mathbf{x} \in D}$ such that $Z_{\mathbf{x}_{(i)}}^{(j)}$, $1 \leq i \leq n$, $1 \leq j \leq p$, are jointly np-variate normally distributed for any $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)} \in D$. The distribution of Z is characterized by its $\mathbb{R}^p$-valued mean function and a matrix-valued covariance function $(k_{ij})_{i,j \in \{1,\ldots,p\}}$.

In both real- and vector-valued cases (assuming additional technical conditions where necessary) k governs a number of *pathwise* properties ranging from square-

integrability to continuity, differentiability and more; see e.g. Section 1.4 of [1] or Chapter 5 of [28] for details. As we will see in Section 4, k actually also governs the FANOVA decomposition of GRF paths $\omega \in \Omega \longrightarrow Z_\bullet(\omega) \in \mathbb{R}^D$. Before establishing this result, let us first introduce a functional ANOVA decomposition for kernels.

3 KANOVA: A kernel ANOVA decomposition

Essentially we apply the $2d$-dimensional version of the decomposition introduced in Section 2 to $\nu \otimes \nu$-square integrable kernels k (s.p.d. or not). From a formal point of view it is more elegant and leads to more efficient notation if we work with the tensor products $T_\mathfrak{u} \otimes T_\mathfrak{v} \colon \mathscr{F} \otimes \mathscr{F} \longrightarrow \mathscr{F} \otimes \mathscr{F}$. It is well known that $L^2(\nu \otimes \nu)$ and $\mathscr{F} \otimes \mathscr{F}$ are isometrically isomorphic (see [15] for details on tensor products of Hilbert spaces), and we silently identify them here for simplicity. Then $T_\mathfrak{u} \otimes T_\mathfrak{v} = T_\mathfrak{u}^{(1)} T_\mathfrak{v}^{(2)} = T_\mathfrak{v}^{(2)} T_\mathfrak{u}^{(1)}$, where $T_\mathfrak{u}^{(1)}, T_\mathfrak{v}^{(2)} \colon L^2(\nu \otimes \nu) \longrightarrow L^2(\nu \otimes \nu)$ are given by $(T_\mathfrak{u}^{(1)} k)(\mathbf{x}, \mathbf{y}) = (T_\mathfrak{u}(k(\bullet, \mathbf{y}))(\mathbf{x})$ and $(T_\mathfrak{v}^{(2)} k)(\mathbf{x}, \mathbf{y}) = (T_\mathfrak{v}(k(\mathbf{x}, \bullet))(\mathbf{y})$.

Theorem 1. *Let k be $\nu \otimes \nu$-square integrable.*

a) There exist $k_{\mathfrak{u},\mathfrak{v}} \in \mathrm{L}^2(\nu \otimes \nu)$ depending solely on $(\mathbf{x}_\mathfrak{u}, \mathbf{y}_\mathfrak{v})$ such that k can be decomposed in a unique way as $k = \sum_{\mathfrak{u},\mathfrak{v} \subseteq I} k_{\mathfrak{u},\mathfrak{v}}$ under the conditions

$$\forall \mathfrak{u}, \mathfrak{v} \subseteq I \ \forall i \in \mathfrak{u} \ \forall j \in \mathfrak{v} \ \int k_{\mathfrak{u},\mathfrak{v}} \ \nu_i(\mathrm{d}x_i) = \mathbf{0} \text{ and } \int k_{\mathfrak{u},\mathfrak{v}} \ \nu_j(\mathrm{d}y_j) = \mathbf{0}. \quad (5)$$

We have

$$k_{\mathfrak{u},\mathfrak{v}}(\mathbf{x}, \mathbf{y}) = \sum_{\mathfrak{u}' \subseteq \mathfrak{u}} \sum_{\mathfrak{v}' \subseteq \mathfrak{v}} (-1)^{|\mathfrak{u}|+|\mathfrak{v}|-|\mathfrak{u}'|-|\mathfrak{v}'|} \int k(\mathbf{x}, \mathbf{y}) \ \nu_{-\mathfrak{u}'}(\mathrm{d}\mathbf{x}_{-\mathfrak{u}'}) \ \nu_{-\mathfrak{v}'}(\mathrm{d}\mathbf{y}_{-\mathfrak{v}'}). \quad (6)$$

Moreover, $k_{\mathfrak{u},\mathfrak{v}}$ may be written concisely as $k_{\mathfrak{u},\mathfrak{v}} = [T_\mathfrak{u} \otimes T_\mathfrak{v}]k$.

b) Suppose that D is compact and k is a continuous s.p.d. kernel. Then, for any $(\alpha_\mathfrak{u})_{\mathfrak{u} \subseteq I} \in \mathbb{R}^{2^d}$, the following function is also a s.p.d. kernel:

$$(\mathbf{x}, \mathbf{y}) \in D \times D \longrightarrow \sum_{\mathfrak{u} \subseteq I} \sum_{\mathfrak{v} \subseteq I} \alpha_\mathfrak{u} \alpha_\mathfrak{v} k_{\mathfrak{u},\mathfrak{v}}(\mathbf{x}, \mathbf{y}) \in \mathbb{R}. \quad (7)$$

Proof. The proofs are in the appendix (Section 8) to facilitate the reading.

Example 1 (The Brownian kernel). Consider the covariance kernel $k(x,y) = \min(x,y)$ of the Brownian motion on $D = [0,1]$, and suppose that ν is the Lebesgue measure. The $k_{\mathfrak{u},\mathfrak{v}}$'s can then easily be obtained by direct calculation: $k_{\emptyset,\emptyset} = \frac{1}{3}$, $k_{\emptyset,\{1\}}(y) = y - \frac{y^2}{2} - \frac{1}{3}$, $k_{\{1\},\emptyset}(x) = x - \frac{x^2}{2} - \frac{1}{3}$, and $k_{\{1\},\{1\}}(x,y) = \min(x,y) - x + \frac{x^2}{2} - y + \frac{y^2}{2} + \frac{1}{3}$.

Example 2. Consider the very common class of tensor product kernels: $k(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^d k_i(x_i, y_i)$ where the k_i's are 1-dimensional symmetric kernels. It turns out that Equation (6) boils down to a sum depending on 1- and 2-dimensional integrals, since

$$\int k(\mathbf{x},\mathbf{y})d\nu_{-\mathfrak{u}}(\mathbf{x}_{-\mathfrak{u}})d\nu_{-\mathfrak{v}}(\mathbf{y}_{-\mathfrak{v}}) =$$
$$\prod_{i\in\mathfrak{u}\cap\mathfrak{v}} k_i(x_i,y_i)\cdot\prod_{i\in\mathfrak{u}\setminus\mathfrak{v}}\int k_i(x_i,\cdot)d\nu_i\cdot\prod_{i\in\mathfrak{v}\setminus\mathfrak{u}}\int k_i(\cdot,y_i)d\nu_i\cdot\prod_{i\notin\mathfrak{u}\cup\mathfrak{v}}\iint k_i d(\nu_i\otimes\nu_i). \quad (8)$$

By symmetry of k, Equation (8) solely depends on the integrals $\iint k_i d(\nu_i\otimes\nu_i)$ and integral functions $t\mapsto\int k_i(\cdot,t)d\nu_i$, $i=1,\dots,d$. We refer to Section 9 for explicit calculations using typical k_i's. A particularly convenient case is considered next.

Corollary 1. *Let $k_i^{(0)}\colon D_i\times D_i\longrightarrow\mathbb{R}$ $(1\le i\le d)$ be argumentwise centred, i.e. such that $\int k_i^{(0)}(\cdot,t)\mathrm{d}\nu_i=\int k_i^{(0)}(s,\cdot)\mathrm{d}\nu_i=0$ for all $i\in I$ and $s,t\in D_i$, and consider $k(\mathbf{x},\mathbf{y})=\prod_{i=1}^d(1+k_i^{(0)}(x_i,y_i))$. Then the KANOVA decomposition of k consists of the terms $[T_\mathfrak{u}\otimes T_\mathfrak{u}]k(\mathbf{x},\mathbf{y})=\prod_{i\in\mathfrak{u}}k_i^{(0)}(x_i,y_i)$ and $[T_\mathfrak{u}\otimes T_\mathfrak{v}]k=\mathbf{0}$ if $\mathfrak{u}\neq\mathfrak{v}$.*

Remark 1. By taking $k(\mathbf{x},\mathbf{y})=\prod_{i=1}^d(1+k_i^{(0)}(x_i,y_i))$, where k_i^0 are s.p.d., we recover the so-called ANOVA kernels [37, 36, 6]. Corollary 1 guarantees for argumentwise centred $k_i^{(0)}$ that the associated k has a simple KANOVA decomposition, with analytically tractable $k_{\mathfrak{u},\mathfrak{u}}$ terms and vanishing $k_{\mathfrak{u},\mathfrak{v}}$ terms (for $\mathfrak{u}\neq\mathfrak{v}$).

4 FANOVA decomposition of Gaussian random field paths

Let $Z=(Z_\mathbf{x})_{\mathbf{x}\in D}$ be a centred GRF with covariance function k. To simplify the arguments we make an assumption (for the rest of the article) that is often satisfied in practice: *let D_i be compact subsets of $\mathbb{R}$ and assume that Z has continuous sample paths.* The latter can be guaranteed by a weak condition on the covariance kernel; see [1], Theorem 1.4.1. For $r\in\mathbb{N}\setminus\{0\}$ write $C_b(D,\mathbb{R}^r)$ for the space of (bounded) continuous functions $D\to\mathbb{R}^r$ equipped with the supremum norm, and set in particular $C_b(D)=C_b(D,\mathbb{R})$. We reinterpret $T_\mathfrak{u}$ as maps $C_b(D)\to C_b(D)$, which are still bounded linear operators.

Theorem 2. *The 2^d-dimensional vector-valued random field $(Z_\mathbf{x}^{(\mathfrak{u})},\mathfrak{u}\subseteq I)_{\mathbf{x}\in D}$ is Gaussian, centred, and has continuous sample paths again. Its matrix-valued covariance function is given by*

$$\mathrm{Cov}(Z_\mathbf{x}^{(\mathfrak{u})},Z_\mathbf{y}^{(\mathfrak{v})})=[T_\mathfrak{u}\otimes T_\mathfrak{v}]k(\mathbf{x},\mathbf{y}). \quad (9)$$

Example 3. Continuing from Example 1, let $B=(B_x)_{x\in[0,1]}$ be Brownian motion on $D=[0,1]$, which is a centred Gaussian random field with continuous paths. Theorem 2 yields that $(T_\emptyset B,T_{\{1\}}B)=(\int_0^1 B_u\mathrm{d}u,B_x-\int_0^1 B_u\mathrm{d}u)_{x\in D}$ is a bivariate random field on D, where $T_\emptyset B$ is a $\mathcal{N}(0,1/3)$-distributed random variable, while $(T_{\{1\}}B_x)$ is a centred Gaussian process with covariance kernel $k_{\{1\},\{1\}}(x,y)=\min(x,y)-x+\frac{x^2}{2}-y+\frac{y^2}{2}+\frac{1}{3}$. The cross-covariance function of the components is given by $\mathrm{Cov}(T_\emptyset B,T_{\{1\}}B_x)=x-\frac{x^2}{2}-\frac{1}{3}$.

Remark 2. Under our conditions on Z and using the notation from the proof of Theorem 1, we have a Karhunen–Loève expansion $Z_{\mathbf{x}} = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\varepsilon_i\phi_i(\mathbf{x})$, where $\varepsilon = (\varepsilon_i)_{i\in\mathbb{N}\setminus\{0\}}$ is a standard Gaussian white noise sequence and the series converges uniformly (i.e. in $C_b(D)$) with probability 1 (and in $L^2(\Omega)$); for $d = 1$ see [16, 1]. Thus by the continuity of $T_{\mathfrak{u}}$, we can expand the projected random field as

$$Z_{\mathbf{x}}^{(\mathfrak{u})} = T_{\mathfrak{u}}\left(\sum_{i=1}^{\infty}\sqrt{\lambda_i}\varepsilon_i\phi_i(\mathbf{x})\right) = \sum_{i=1}^{\infty}\sqrt{\lambda_i}\varepsilon_i T_{\mathfrak{u}}\left(\phi_i\right)(\mathbf{x}), \tag{10}$$

where the series converges uniformly in $\mathbf{x}$ with probability 1 (and in $L^2(\Omega)$). This is the basis for an alternative proof of Theorem 2. We can also verify Equation (9) under these conditions. Using the left/right-continuity of cov in $L^2(\Omega)$, we obtain indeed $\operatorname{cov}\left(Z_{\mathbf{x}}^{(\mathfrak{u})}, Z_{\mathbf{y}}^{(\mathfrak{v})}\right) = \sum_{i=1}^{\infty}\lambda_i\, T_{\mathfrak{u}}(\phi_i)(\mathbf{x})\, T_{\mathfrak{v}}(\phi_i)(\mathbf{y}) = k_{\mathfrak{u},\mathfrak{v}}(\mathbf{x},\mathbf{y})$.

Corollary 2. *(a) For any $\mathfrak{u} \subseteq I$ the following statements are equivalent:*

(i) $T_{\mathfrak{u}}(k(\bullet,\mathbf{y})) = \mathbf{0}$ *for every* $\mathbf{y} \in D$
(ii) $[T_{\mathfrak{u}} \otimes T_{\mathfrak{u}}]k = \mathbf{0}$
(iii) $[T_{\mathfrak{u}} \otimes T_{\mathfrak{u}}]k(\mathbf{x},\mathbf{x}) = 0$ *for every* $\mathbf{x} \in D$
(iv) $\mathbb{P}(Z^{(\mathfrak{u})} = \mathbf{0}) = 1$

(b) For any $\mathfrak{u},\mathfrak{v} \subseteq I$ with $\mathfrak{u} \neq \mathfrak{v}$ the following statements are equivalent:

(i) $[T_{\mathfrak{u}} \otimes T_{\mathfrak{v}}]k = \mathbf{0}$
(ii) $Z^{(\mathfrak{u})}$ *and* $Z^{(\mathfrak{v})}$ *are two independent GRFs*

Remark 3. A consequence of Corollary 2 is that choosing a kernel without $\mathfrak{u}$ component in GRF-based GSA will lead to a posterior distribution without $\mathfrak{u}$ component whatever the assimilated data, i.e. $\mathbb{P}(Z^{(\mathfrak{u})} = \mathbf{0} \,|\, Z_{\mathbf{x}_1}, \ldots, Z_{\mathbf{x}_n}) = 1$ (a.s.). Indeed, an a.s. constant random element remains a.s. constant under any conditioning.

However, the analogous result does not hold for cross-covariances between $Z^{(\mathfrak{u})}$ and $Z^{(\mathfrak{v})}$ for $\mathfrak{u} \neq \mathfrak{v}$. Let us take for instance $D = [0,1]$, ν arbitrary, and $Z_t = U + Y_t$, where $U \sim \mathcal{N}(0,\sigma^2)$ ($\sigma > 0$) and (Y_t) is a centred Gaussian process with argumentwise centred covariance kernel $k^{(0)}$. Assuming that U and Y are independent, it is clear that $(T_\emptyset Z)_s = 0$ and $(T_{\{1\}}Z)_t = Y_t$, so $\operatorname{cov}((T_\emptyset Z)_s, (T_{\{1\}}Z)_t) = 0$. If in addition Z was observed at a point $r \in D$, Equation (9) yields $\operatorname{cov}((T_\emptyset Z)_s, (T_{\{1\}}Z)_t | Z_r) = (T_\emptyset \otimes T_{\{1\}})(k(\bullet,\star) - k(\bullet,r)k(r,\star)/k(r,r))(s,t)$, where $k(s,t) = \sigma^2 + k^{(0)}(s,t)$ is the covariance kernel of Z. By Equation (6) we obtain $\operatorname{cov}((T_\emptyset Z)_s, (T_{\{1\}}Z)_t | Z_r) = -\sigma^2 k^{(0)}(t,r)/(\sigma^2 + k^{(0)}(r,r))$, which in general is nonzero.

Remark 4. Coming back to the ANOVA kernels discussed in Remark 1, Corollary 2(b) implies that for a centred GRF with continuous sample paths and covariance kernel of the form $k(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{d}(1 + k_i^{(0)}(x_i,y_i))$, where $k_i^{(0)}$ is argumentwise centred, the FANOVA effects $Z^{(\mathfrak{u})}$, $\mathfrak{u} \subseteq I$, are actually independent.

To close this section, let us finally touch upon the distribution of Sobol' indices of GRF sample paths, relying on Theorem 2 and Remark 2.

Corollary 3. *For $\mathfrak{u} \subseteq I$, $\mathfrak{u} \neq \emptyset$, we can represent the Sobol' indices of Z as*

$$S_{\mathfrak{u}}(Z) = \frac{Q_{\mathfrak{u}}(\varepsilon,\varepsilon)}{\sum_{\mathfrak{v}\neq\emptyset} Q_{\mathfrak{v}}(\varepsilon,\varepsilon)},$$

where the $Q_{\mathfrak{u}}$'s are quadratic forms in a standard Gaussian white noise sequence. In the notation of Remark 2, $Q_{\mathfrak{u}}(\varepsilon,\varepsilon) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sqrt{\lambda_i\lambda_j}\langle T_{\mathfrak{u}}\phi_i, T_{\mathfrak{u}}\phi_j\rangle\varepsilon_i\varepsilon_j$, where the convergence is uniform with probability 1.

Remark 5. Consider the GRF $Z' = Z - T_{\emptyset}Z$ with Karhunen–Loève expansion $Z'_{\mathbf{x}} = \sum_{i=1}^{\infty}\sqrt{\lambda'_i}\phi'_i(\mathbf{x})\varepsilon_i$. From Equation (4) and (the proof of) Corollary 3 we can see that $S_u(Z) = S_u(Z') = \sum_{i,j=1}^{\infty} g'_{ij}\varepsilon_i\varepsilon_j \big/ \sum_{i=1}^{\infty}\lambda'_i\varepsilon_i^2$, where $g'_{ij} = \sqrt{\lambda'_i\lambda'_j}\langle T_{\mathfrak{u}}\phi'_i, T_{\mathfrak{u}}\phi'_j\rangle$. Truncating both series above at $K \in \mathbb{N}$, applying the theorem in Section 2 of [27] and then Lebesgue's theorem for $K \to \infty$, we obtain

$$\mathbb{E}S_{\mathfrak{u}}(Z) = \sum_{i=1}^{\infty} g'_{ii}\int_0^{\infty}\left((1+2\lambda'_i t)^{3/2}\prod_{l\neq i}(1+2\lambda'_l t)^{1/2}\right)^{-1} dt,$$

$$\mathbb{E}S_{\mathfrak{u}}(Z)^2 = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(g'_{ii}g'_{jj} + 2{g'_{ij}}^2)\int_0^{\infty} t\left((1+2\lambda'_i t)^{3/2}\prod_{l\notin\{i,j\}}(1+2\lambda'_l t)^{1/2}\right)^{-1} dt.$$

5 Making new kernels from old with KANOVA

While kernel methods and Gaussian process modelling have proven efficient in a number of classification and prediction problems, finding a suitable kernel for a given application is often judged difficult. It should simultaneously express the desired features of the problem at hand while respecting positive definiteness, a mathematical constraint that is not straightforward to check in practice. In typical implementations of kernel methods, a few classes of standard stationary kernels are available for which positive definiteness was established analytically based on the Bochner theorem. On the other hand, some operations on kernels are known to preserve positive-definiteness, which enables enriching the available dictionary of kernels notably by multiplication by a positive constant, convex combinations, products and convolutions of kernels, or deformations of the input space. The section *Making new kernels from old* of [24] (Section 4.2.4) covers a number of such operations. We now consider some new ways of creating admissible kernels in the context of the KANOVA decomposition of Section 3. Let us first consider as before some square-integrable symmetric positive definite kernel k_{old} and take $\mathfrak{u} \subseteq I$.

One straightforward approach to create a kernel whose associated Gaussian random field has paths in $\mathscr{F}_{\mathfrak{u}}$ is then to plainly take the "simple" projected kernel

$$k_{\text{new}} = \pi_{\mathfrak{u}}k_{\text{old}} \text{ with } \pi_{\mathfrak{u}} = T_{\mathfrak{u}} \otimes T_{\mathfrak{u}}. \tag{11}$$

From Theorem 1(b) it is clear that such kernels are s.p.d.; however, they will generally not be strictly positive definite.

Going one step further, one obtains a richer class of 2^{2^d} positive definite kernels by considering parts of $\mathscr{P}(I)$, and designing kernels accordingly. Taking $U \subset \mathscr{P}(I)$, we obtain a further class of projected kernels as follows:

$$k_{\text{new}} = \pi_U k_{\text{old}} \text{ with } \pi_U = T_U \otimes T_U = \sum_{\mathfrak{u}\in U} \sum_{\mathfrak{v}\in U} T_{\mathfrak{u}} \otimes T_{\mathfrak{v}}, \text{ where } T_U = \sum_{\mathfrak{u}\in U} T_{\mathfrak{u}}. \quad (12)$$

The resulting kernel is again s.p.d., which follows from Theorem 1(b) by choosing $\alpha_{\mathfrak{u}} = 1$ if $\mathfrak{u} \in U$ and $\alpha_{\mathfrak{u}} = 0$ otherwise. Such a kernel contains not only the covariances induced by the effects associated with the different subsets of U, but also cross-covariances between these effects. Finally, another relevant class of positive definite projected kernels can be designed by taking

$$k_{\text{new}} = \pi^\star_U k_{\text{old}} \text{ with } \pi^\star_U = \sum_{\mathfrak{u}\in U} T_{\mathfrak{u}} \otimes T_{\mathfrak{u}}. \quad (13)$$

This kernel corresponds to the one of a sum of independent random fields with same individual distributions as the $Z^{(\mathfrak{u})}$ $(\mathfrak{u} \in U)$. In addition, projectors of the form $\pi_{U_1}, \pi^\star_{U_2}$ $(U_1, U_2 \subset \mathscr{P}(I))$ can be combined (e.g. by sums or convex combinations) in order to generate a large class of s.p.d. kernels, as illustrated here and in Section 6.

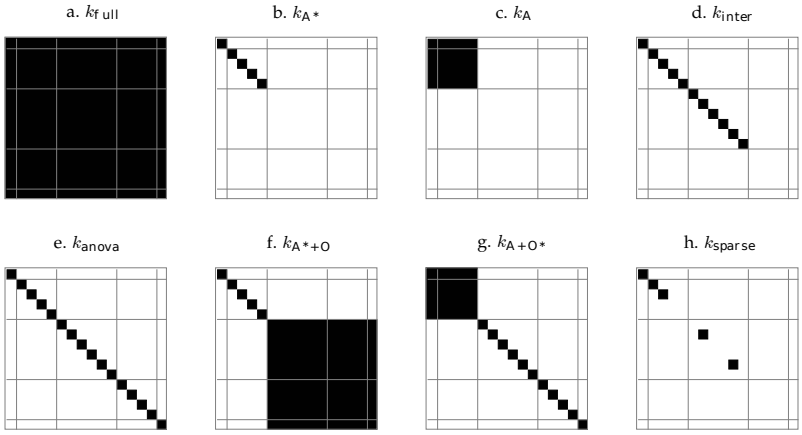

Fig. 1 Schematic representations of a reference kernel k_{full} and various projections or sums of projections. The expressions of these kernels are detailed in Section 6.

Example 4. Let us consider $A = \{\emptyset, \{1\}, \{2\}, \ldots, \{d\}\}$ and O, the complement of A in $\mathscr{P}(I)$. While A corresponds to the constant and main effects forming the additive component in the FANOVA decomposition, O corresponds to all higher-order terms, referred to as *ortho-additive* component in [19]. Taking $\pi_A k = (T_A \otimes T_A)k$ amounts to

extracting the additive component of k with cross-covariances between the various main effects (including the constant); see Figure 1(c). On the other hand, $\pi_A^\star k = \sum_{\mathfrak{u}\in A}\pi_{\mathfrak{u}}k$ retains these main effects without their possible cross-covariances; see Figure 1(b). In the next theorem (proven in [19]), analytical formulae are given for $\pi_A k$ and related terms for the class of tensor product kernels.

Theorem 3. *Let $D_i = [a_i, b_i]$ ($a_i < b_i$) and $k = \prod_{i=1}^d k_i$, where the k_i are s.p.d. kernels on D_i such that $k_i(x_i, y_i) > 0$ for all $x_i, y_i \in D_i$. Then, the additive and ortho-additive components of k with their cross-covariances are given by*

$$(\pi_A k)(\mathbf{x},\mathbf{y}) = \frac{a(\mathbf{x})a(\mathbf{y})}{\mathscr{E}} + \mathscr{E}\cdot\sum_{i=1}^d\left(\frac{k_i(x_i,y_i)}{\mathscr{E}_i} - \frac{E_i(x_i)E_i(y_i)}{\mathscr{E}_i^2}\right)$$

$$(T_O\otimes T_A k)(\mathbf{x},\mathbf{y}) = (T_A\otimes T_O k)(\mathbf{y},\mathbf{x}) = E(\boldsymbol{x})\cdot\left(1-d+\sum_{j=1}^d\frac{k_j(x_j,y_j)}{E_j(x_j)}\right) - (\pi_A k)(\mathbf{x},\mathbf{y})$$

$$(\pi_O k)(\mathbf{x},\mathbf{y}) = k(\mathbf{x},\mathbf{y}) - (T_A\otimes T_O k)(\mathbf{x},\mathbf{y}) - (T_O\otimes T_A k)(\mathbf{x},\mathbf{y}) - (\pi_A k)(\mathbf{x},\mathbf{y})$$

where $E_i(x_i) = \int_{a_i}^{b_i} k_i(x_i,y_i)\,dy_i$, $E(\mathbf{x}) = \prod_{i=1}^d E_i(x_i)$, $\mathscr{E}_i = \int_{a_i}^{b_i} E_i(x_i)\nu_i(\mathrm{d}x_i)$, $\mathscr{E} = \prod_{i=1}^d \mathscr{E}_i$, and $a(\mathbf{x}) = \mathscr{E}\left(1-d+\sum_{i=1}^d\frac{E_i(x_i)}{\mathscr{E}_i}\right)$.

6 Numerical experiments

We now consider a 30 dimensional numerical experiment where we compare the prediction abilities of various sparse kernels obtained from the KANOVA decomposition of a squared-exponential kernel

$$k_{\text{full}}(\mathbf{x},\mathbf{y}) = \exp(-||\mathbf{x}-\mathbf{y}||^2), \qquad \mathbf{x},\mathbf{y}\in(0,1)^{30}. \tag{14}$$

As detailed in the previous sections, k_{full} can be expanded as a sum of 4^{30} terms, and sparsified versions of k_{full} can be obtained by projections such as in Example 4. We will focus hereafter on seven sub-kernels (all summations are over $\mathfrak{u},\ \mathfrak{v}\subseteq I$):

$$\begin{array}{ll}
k_{\text{anova}} = \sum\pi_{\mathfrak{u}}k & k_{A^\star} = \sum_{|\mathfrak{u}|\le 1}\pi_{\mathfrak{u}}k \\
k_A = \sum_{|\mathfrak{u}|\le 1}\sum_{|\mathfrak{v}|\le 1}(T_{\mathfrak{u}}\otimes T_{\mathfrak{v}})k & k_{A^\star+O} = \pi_O k + k_{A^\star} \\
k_{A+O^\star} = k_{\text{anova}} - k_{A^\star} + k_A & k_{\text{inter}} = \sum_{|\mathfrak{u}|\le 2}\pi_{\mathfrak{u}}k \\
k_{\text{sparse}} = (\pi_\emptyset + \pi_{\{1\}} + \pi_{\{2\}} + \pi_{\{2,3\}} + \pi_{\{4,5\}})k. &
\end{array} \tag{15}$$

A schematic representation of these kernels can be found in Figure 1. Note that the tensor product structure of k_{full} allows to use Theorem 3 in order to get more tractable expressions for all kernels above. Furthermore, the integrals appearing in the E_i and $\mathscr{E}_i$ terms can be calculated analytically as detailed in Section 9.

We now compare the predictions obtained by GRF modelling with these kernels on a benchmark of test functions given by sample paths from centred GRFs pos-

sessing the same set of kernels (200 paths per kernel). Whenever the kernel used for prediction is not the same as the one used for simulation, a Gaussian observation noise with variance τ^2 is assumed in the models used in prediction, where τ^2 is chosen so as to reflect the part of variance that cannot be approximated by the model.

We consider a training set X_{train} of 500 points and a test set X_{test} of 200 points given by a Latin Hypercube design with optimized maximin criterion [26, 10]. The accuracy of the fit is measured using the following criterion :

$$C = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum y_i^2} \tag{16}$$

where y is the vector of the test function values at the test points and $\hat{y}$ is the vector of predicted values. This criterion is equal to one when the prediction error is null and it is equal to zero when the model predicts as bad as the null function.

The values of the criterion for all couples of test functions and models are summarized in Table 1. Let us stress three important points from these results.

First, this example illustrates that, unless the correlation range is increased, predicting a GRF based on 500 points in dimension 30 is hopeless when the covariance structure is full or close to full (first four rows of Table 1) no matter what sub-kernel is chosen for prediction. However, for GRFs with sparser covariance, prediction performances are strongly increased (last four rows of Table 1).

Second, still focusing on the four last lines of Table 1, k_{inter} seems to offer a nice compromise as it works much better than other sub-kernels on Z_{inter} and achieves very good performances on the sparser GRF sample paths. Besides this, it is not doing much worse than the best sub-kernels on lines 1 to 4.

Third, we observe that neglecting cross-correlations between blocks has very little or no influence on the results, so that the Gaussian kernel appears to have a structure relatively close to a diagonal one. This point remains to be studied analytically.

	k_{full}	k_{anova}	$k_{A^\star+O}$	$k_{A+O^\star}$	k_{inter}	$k_{A^\star}$	k_A	k_{sparse}
Z_{full}	0.06	0.05	0.06	0.05	0.05	0.03	0.04	0.01
Z_{anova}	0.05	0.05	0.05	0.05	0.04	0.03	0.03	0.01
$Z_{A^\star+O}$	0.05	0.04	0.05	0.04	0.04	0.03	0.03	0.01
$Z_{A+O^\star}$	0.06	0.06	0.06	0.06	0.05	0.04	0.04	0.01
Z_{inter}	0.33	0.37	0.34	0.37	0.7	0.28	0.28	0.07
$Z_{A^\star}$	0.67	0.76	0.71	0.75	0.96	1	1	0.2
Z_A	0.69	0.77	0.71	0.77	0.96	1	1	0.18
Z_{sparse}	0.75	0.83	0.8	0.78	0.95	0.9	0.9	1
mean	0.33	0.37	0.35	0.36	0.47	0.41	0.42	0.19

Table 1 Average value of C over the 200 replications of the experiment. Lines correspond to classes of test functions (GRF models used for simulation) while columns correspond to the kernels used for prediction. The four last lines of the k_{inter} column are coloured to highlight the superior performances of that kernel when the class of test functions is as sparse or sparser than Z_{inter}.

7 Conclusion and perspectives

We have proposed an ANOVA decomposition of kernels (KANOVA), and shown how KANOVA governs the probability distribution of FANOVA effects of Gaussian random field paths. This has enabled us in turn to establish that ANOVA kernels correspond to centred Gaussian random fields with independent FANOVA effects, to make progress towards the distribution of Sobol' indices of Gaussian random fields, and also to suggest a number of operations for making new symmetric positive definite kernels from existing ones. Particular cases include the derivation of additive and ortho-additive kernels extracted from tensor product kernels, for which a closed form formula was given. Besides this, a 30-dimensional numerical experiment supports the hypothesis that KANOVA may be a useful approach to designing kernels for high-dimensional kriging, as the performances of the interaction kernel suggest. Perspectives include analytically calculating the norm of terms appearing in the KANOVA decomposition to better understand the structure of common GRF models. From a practical point of view, a next challenge will be to parametrize decomposed kernels adequately so as to recover from data which terms of the FANOVA decomposition are dominating and to automatically design adapted kernels from this.

Acknowledgements The authors would like to thank Dario Azzimonti for proofreading.

References

1. Adler, R., Taylor, J.: Random Fields and Geometry. Springer, Boston (2007)
2. Antoniadis, A.: Analysis of variance on function spaces. Statistics **15**, 59–71 (1984)
3. Berlinet, A., Thomas-Agnan, C.: Reproducing kernel Hilbert spaces in probability and statistics. Kluwer Academic Publishers (2004)
4. Chastaing, G., Le Gratiet, L.: ANOVA decomposition of conditional Gaussian processes for sensitivity analysis with dependent inputs. Journal of Statistical Computation and Simulation (2014). DOI 10.1080/00949655.2014.925111
5. Durrande, N., Ginsbourger, D., Roustant, O.: Additive covariance kernels for high-dimensional Gaussian process modeling. Annales de la faculté des Sciences de Toulouse **21**, 481 – 499 (2012)
6. Durrande, N., Ginsbourger, D., Roustant, O., Carraro, L.: ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis. Journal of Multivariate Analysis **115**, 57–67 (2013)
7. Duvenaud, D.: Automatic model construction with Gaussian processes. Ph.D. thesis, Computational and Biological Learning Laboratory, University of Cambridge (2014)
8. Duvenaud, D., Nickisch, H., Rasmussen, C.: Additive Gaussian Processes (2011). NIPS conference
9. Efron, B., Stein, C.: The jackknife estimate of variance. The Annals of Statistics **9**, 586–596 (1981)
10. Franco, J., Dupuy, D., Roustant, O., Damblin, G., Iooss, B.: DiceDesign: Designs of Computer Experiments (2014). R package version 1.6
11. Gikhman, I.I., Skorokhod, A.V.: The theory of stochastic processes. I. Springer, Berlin (2004). Translated from the Russian by S. Kotz, Reprint of the 1974 edition

12. Hoeffding, W.: A class of statistics with asymptotically normal distributions. Annals of Mathematical Statistics **19**, 293–325 (1948)
13. Janon, A., Klien, T., Lagnoux, A., Nodet, M., Prieur, C.: Asymptotic normality and efficiency of two sobol index estimators. ESAIM: Probability and Statistics (2013)
14. Kaufman, C., Sain, S.: Bayesian functional ANOVA modeling using Gaussian process prior distributions. Bayesian Analysis **5**, 123–150 (2010)
15. Krée, P.: Produits tensoriels complétés d'espaces de Hilbert. Séminaire Paul Krée **Vol 1**, No. 7 (1974–1975)
16. Kuelbs, J.: Expansions of vectors in a Banach space related to Gaussian measures. Proceedings of the American Mathematical Society **27 (2)**, 364–370 (1971)
17. Kuo, F.Y., Sloan, I.H., Wasilkowski, G.W., Wozniakowski, H.: On decompositions of multivariate functions. Mathematics of Computation **79**, 953–966 (2010)
18. Le Gratiet, L., Cannamela, C., Iooss, B.: A Bayesian approach for global sensitivity analysis of (multi-fidelity) computer codes. SIAM/ASA J. Uncertainty Quantification **2(1)**, 336–363 (2014)
19. Lenz, N.: Additivity and ortho-additivity in Gaussian random fields. Master's thesis, Departement of Mathematics and Statistics, University of Bern (2013). URL http://hal.archives-ouvertes.fr/hal-01063741
20. Marrel, A., Iooss, B., Laurent, B., Roustant, O.: Calculations of Sobol indices for the Gaussian process metamodel. Reliability Engineering & System Safety **94**, 742–751 (2009)
21. Muehlenstaedt, T., Roustant, O., Carraro, L., Kuhnt, S.: Data-driven Kriging models based on FANOVA-decomposition. Statistics and Computing **22**(3), 723–738 (2012)
22. Oakley, J., O'Hagan, A.: Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society **66**, 751–769 (2004)
23. Rajput, B.S., Cambanis, S.: Gaussian processes and Gaussian measures. Ann. Math. Statist. **43**, 1944–1952 (1972)
24. Rasmussen, C.R., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press (2006)
25. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S.: Global sensitivity analysis: the primer. Wiley Online Library (2008)
26. Santner, T., Williams, B., Notz, W.: The design and analysis of computer experiments. Springer Verlag (2003)
27. Sawa, T.: The exact moments of the least squares estimator for the autoregressive model. J. Econometrics **8**(2), 159–172 (1978)
28. Scheuerer, M.: A comparison of models and methods for spatial interpolation in statistics and numerical analysis. Ph.D. thesis, Georg-August-Universität Göttingen (2009)
29. Schuhmacher, D.: Distance estimates for Poisson process approximations of dependent thinnings. Electron. J. Probab. **10**, no. 5, 165–201 (electronic) (2005)
30. Sobol', I.: Multidimensional Quadrature Formulas and Haar Functions. Nauka, Moscow (1969). (In Russian)
31. Sobol, I.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation **55**(1-3), 271–280 (2001)
32. Steinwart, I., Scovel, C.: Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs. Constructive Approximation **35 (3)**, 363–417 (2012)
33. Talagrand, M.: Regularity of Gaussian processes. Acta Math. **159**(1-2), 99–149 (1987)
34. Tarieladze, V., Vakhania, N.: Disintegration of Gaussian measures and average-case optimal algorithms. J. Complexity **23**(4-6), 851–866 (2007)
35. Touzani, S.: Response surface methods based on analysis of variance expansion for sensitivity analysis. Ph.D. thesis, Université de Grenoble (2011)
36. Vapnik, V.: Statistical Learning Theory. Wiley, New York (1998)
37. Wahba, G.: Spline Models for Observational Data. Siam (1990)
38. Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., Morris, M.D.: Screening, predicting, and computer experiments. Technometrics **34**, 15–25 (1992)

8 Proofs

Proof (Theorem 1). a) The first part and the concrete solution (6) follow directly from the corresponding statements in Section 2. Having established (6), it is easily seen that $[T_\mathfrak{u} \otimes T_\mathfrak{v}]k = T_\mathfrak{u}^{(1)} T_\mathfrak{v}^{(2)} k$ coincides with $k_{\mathfrak{u},\mathfrak{v}}$.

b) Under these conditions Mercer's theorem applies (see [32] for an overview and recent extensions). So there exist a non-negative sequence $(\lambda_i)_{i\in\mathbb{N}\setminus\{0\}}$, and continuous representatives $(\phi_i)_{i\in\mathbb{N}\setminus\{0\}}$ of an orthonormal basis of $\mathrm{L}^2(\nu)$ such that $k(\mathbf{x},\mathbf{y}) = \sum_{i=1}^\infty \lambda_i \phi_i(\mathbf{x})\phi_i(\mathbf{y})$, $\mathbf{x},\mathbf{y} \in D$, where the convergence is absolute and uniform. Noting that $T_\mathfrak{u}, T_\mathfrak{v}$ are also bounded as operators on continuous functions, applying $T_\mathfrak{u}^{(1)} T_\mathfrak{v}^{(2)}$ from above yields that

$$\sum_{\mathfrak{u}\subseteq I} \sum_{\mathfrak{v}\subseteq I} \alpha_\mathfrak{u}\alpha_\mathfrak{v} k_{\mathfrak{u},\mathfrak{v}}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^\infty \lambda_i \psi_i(\mathbf{x})\psi_i(\mathbf{y}), \tag{17}$$

where $\psi_i = \sum_{\mathfrak{u}\subseteq I} \alpha_\mathfrak{u}(T_\mathfrak{u}\phi_i)$. Thus the considered function is indeed s.p.d.

Proof (Corollary 1). Expand the product $\prod_{l=1}^d (1 + k_l^{(0)}(x_l,y_l))$ and conclude by uniqueness of the KANOVA decomposition, noting that $\int \prod_{l\in\mathfrak{u}} k_l^{(0)}(x_l,y_l)\nu_i(\mathrm{d}x_i) = \int \prod_{l\in\mathfrak{u}} k_l^{(0)}(x_l,y_l)\nu_j(\mathrm{d}y_j) = 0$ for any $\mathfrak{u}\subseteq I$ and any $i,j\in\mathfrak{u}$.

Proof (Theorem 2). Sample path continuity implies product-measurability of Z and $Z^{(\mathfrak{u})}$, respectively, as can be shown by an approximation argument; see e.g. Prop. A.D. in [29]. Due to Theorem 3 in [33], the covariance kernel k is continuous, hence $\int_D \mathbb{E}|Z_\mathbf{x}|\,\nu_{-\mathfrak{u}}(\mathrm{d}\mathbf{x}_{-\mathfrak{u}}) \le (\int_D k(\mathbf{x},\mathbf{x})\,\nu_{-\mathfrak{u}}(\mathrm{d}\mathbf{x}_{-\mathfrak{u}}))^{1/2} < \infty$ for any $\mathfrak{u}\subseteq I$ and by Cauchy–Schwarz $\int_D\int_D \mathbb{E}|Z_\mathbf{x}Z_\mathbf{y}|\,\nu_{-\mathfrak{u}}(\mathrm{d}\mathbf{x}_{-\mathfrak{u}})\nu_{-\mathfrak{v}}(\mathrm{d}\mathbf{y}_{-\mathfrak{v}}) < \infty$ for any $\mathfrak{u},\mathfrak{v}\subseteq I$. Replacing f by Z in Formula (2), taking expectations and using Fubini's theorem yields that $Z^{(\mathfrak{u})}$ is centred again. Combining (2), Fubini's theorem, and (6) yields

$$\begin{aligned}
&\mathrm{Cov}(Z_\mathbf{x}^{(\mathfrak{u})}, Z_\mathbf{y}^{(\mathfrak{v})}) \\
&\quad= \sum_{\mathfrak{u}'\subseteq\mathfrak{u}} \sum_{\mathfrak{v}'\subseteq\mathfrak{v}} (-1)^{|\mathfrak{u}|+|\mathfrak{v}|-|\mathfrak{u}'|-|\mathfrak{v}'|} \mathrm{Cov}\left(\int Z_\mathbf{x}\, \nu_{-\mathfrak{u}'}(\mathrm{d}\mathbf{x}_{-\mathfrak{u}'}), \int Z_\mathbf{y}\, \nu_{-\mathfrak{v}'}(\mathrm{d}\mathbf{y}_{-\mathfrak{v}'}) \right) \\
&\quad= \sum_{\mathfrak{u}'\subseteq\mathfrak{u}} \sum_{\mathfrak{v}'\subseteq\mathfrak{v}} (-1)^{|\mathfrak{u}|+|\mathfrak{v}|-|\mathfrak{u}'|-|\mathfrak{v}'|} \int \mathrm{Cov}(Z_\mathbf{x}, Z_\mathbf{y})\, \nu_{-\mathfrak{u}'}(\mathrm{d}\mathbf{x}_{-\mathfrak{u}'})\, \nu_{-\mathfrak{v}'}(\mathrm{d}\mathbf{y}_{-\mathfrak{v}'}) \\
&\quad= [T_\mathfrak{u}\otimes T_\mathfrak{v}]k\,(\mathbf{x},\mathbf{y}).
\end{aligned} \tag{18}$$

It remains to show the joint Gaussianity of the $Z^{(\mathfrak{u})}$. First note that $C_b(D,\mathbb{R}^r)$ is a separable Banach space for $r\in\mathbb{N}\setminus\{0\}$. We may and do interprete Z as a random element of $C_b(D)$, equipped with the σ-algebra $\mathscr{B}^D$ generated by the evaluation maps $[C_b(D) \ni f \mapsto f(\mathbf{x}) \in \mathbb{R}]$. By Theorem 2 in [23] the distribution $\mathbb{P}Z^{-1}$ of Z is a Gaussian measure on $\big(C_b(D), \mathscr{B}(C_b(D))\big)$. Since $T_\mathfrak{u}$ is a bounded linear operator $C_b(D)\to C_b(D)$, we obtain immediately that the "combined operator" $\mathfrak{T}\colon C_b(D)\to C_b(D,\mathbb{R}^{2^d})$, defined by $(\mathfrak{T}(f))(\mathbf{x}) = (T_\mathfrak{u} f(\mathbf{x}))_{\mathfrak{u}\subseteq I}$, is also bounded and

linear. Corollary 3.7 of [34] yields that the image measure $(\mathbb{P}Z^{-1})\mathfrak{T}^{-1}$ is a Gaussian measure on $C_b(D,\mathbb{R}^{2^d})$. This means that for every bounded linear operator $\Lambda\colon C_b(D,\mathbb{R}^{2^d}) \to \mathbb{R}$ the image measure $((\mathbb{P}Z^{-1})\mathfrak{T}^{-1})\Lambda^{-1}$ is a univariate normal distribution, i.e. $\Lambda(\mathfrak{T}Z)$ is a Gaussian random variable. Thus, for all $n \in \mathbb{N}$, $\mathbf{x}^{(i)} \in D$ and $a_i^{(\mathfrak{u})} \in \mathbb{R}$, where $1 \le i \le n$, $u \subseteq I$, we obtain that $\sum_{i=1}^n \sum_{\mathfrak{u}\subseteq I} a_i^{(\mathfrak{u})}(T_{\mathfrak{u}}Z)_{\mathbf{x}^{(i)}}$ is Gaussian by the fact that $[C_b(D) \ni f \mapsto f(\mathbf{x}) \in \mathbb{R}]$ is continuous (and linear) for every $\mathbf{x} \in D$. We conclude that $\mathfrak{T}Z = (Z_{\mathbf{x}}^{(\mathfrak{u})}, \mathfrak{u} \subseteq I)_{\mathbf{x}\in D}$ is a vector-valued Gaussian random field.

Proof (Corollary 2). (a) If (i) holds, $[T_{\mathfrak{u}} \otimes T_{\mathfrak{u}}]k = T_{\mathfrak{u}}^{(2)}(T_{\mathfrak{u}}^{(1)}k) = \mathbf{0}$ by $(T_{\mathfrak{u}}^{(1)}k)(\bullet,\mathbf{y}) = T_{\mathfrak{u}}(k(\bullet,\mathbf{y}))$; thus (ii) holds. (ii) trivially implies (iii). Statement (iii) means that $\mathrm{Var}(Z_{\mathbf{x}}^{(\mathfrak{u})}) = 0$, which implies that $Z_{\mathbf{x}}^{(\mathfrak{u})} = 0$ a.s., since $Z^{(\mathfrak{u})}$ is centred. (iv) follows by noting that $\mathbb{P}(Z_{\mathbf{x}}^{(\mathfrak{u})} = 0) = 1$ for all $\mathbf{x} \in D$ implies $\mathbb{P}(Z^{(\mathfrak{u})} = \mathbf{0}) = 1$ by the fact that $Z^{(\mathfrak{u})}$ has continuous sample paths and is therefore separable. Finally, (iv) implies (i) because $T_{\mathfrak{u}}(k(\bullet,\mathbf{y})) = \mathrm{Cov}(Z_{\bullet}^{(\mathfrak{u})}, Z_{\mathbf{y}}) = \mathbf{0}$; compare (18) for the first equality.

(b) For any $m,n \in \mathbb{N}$ and $\mathbf{x}_1,\dots,\mathbf{x}_m,\mathbf{y}_1,\dots,\mathbf{y}_n \in D$ we obtain by Theorem 2 that $Z_{\mathbf{x}_1}^{(\mathfrak{u})},\dots,Z_{\mathbf{x}_m}^{(\mathfrak{u})},Z_{\mathbf{y}_1}^{(\mathfrak{v})},\dots,Z_{\mathbf{y}_n}^{(\mathfrak{v})}$ are jointly normally distributed. Statement (i) is equivalent to saying that $\mathrm{Cov}(Z_{\mathbf{x}}^{(\mathfrak{u})}, Z_{\mathbf{y}}^{(\mathfrak{v})}) = 0$ for all $\mathbf{x},\mathbf{y} \in D$. Thus $(Z_{\mathbf{x}_1}^{(\mathfrak{u})},\dots,Z_{\mathbf{x}_m}^{(\mathfrak{u})})$ and $(Z_{\mathbf{y}_1}^{(\mathfrak{v})},\dots,Z_{\mathbf{y}_n}^{(\mathfrak{v})})$ are independent. Since the sets

$$\{(f,g) \in \mathbb{R}^D \times \mathbb{R}^D\colon (f(\mathbf{x}_1),\dots,f(\mathbf{x}_m)) \in A, (g(\mathbf{y}_1),\dots,g(\mathbf{y}_n)) \in B\} \tag{19}$$

with $m,n \in \mathbb{N}$, $\mathbf{x}_1,\dots,\mathbf{x}_m,\mathbf{y}_1,\dots,\mathbf{y}_n \in D$, $A \in \mathscr{B}(\mathbb{R}^m)$, $B \in \mathscr{B}(\mathbb{R}^n)$ generate $\mathscr{B}^D \otimes \mathscr{B}^D$ (and the system of such sets is stable under intersections), statement (ii) follows. The converse direction is straightforward.

Proof (Corollary 3). By Remark 2, there is a Gaussian white noise sequence $\varepsilon = (\varepsilon_i)_{i\in\mathbb{N}\setminus\{0\}}$ such that $Z_{\mathbf{x}} = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\varepsilon_i\phi_i(\mathbf{x})$ uniformly with probability 1. From $Z_{\mathbf{x}}^{(\mathfrak{u})} = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\varepsilon_i T_{\mathfrak{u}}\phi_i(\mathbf{x})$, we obtain $\|Z^{(\mathfrak{u})}\|^2 = Q_{\mathfrak{u}}(\varepsilon,\varepsilon)$ with $Q_{\mathfrak{u}}$ as defined in the statement. A similar calculation for the denominator of $S_{\mathfrak{u}}(Z)$ leads to $\sum_{\mathfrak{v}\neq\emptyset} Q_{\mathfrak{v}}(\varepsilon,\varepsilon)$, which concludes the proof.

9 Additional examples

Here we give useful expressions to compute the KANOVA decomposition of some famous tensor product kernels with respect to the uniform measure on $[0,1]^d$. For the sake of simplicity we denote the 1-dimensional kernels that they are based on by k (corresponding to the notation k_i in Example 2). The uniform measure on $[0,1]$ is denoted by λ.

Example 5 (Exponential kernel). If $k(x,y) = \exp\left(-\frac{|x-y|}{\theta}\right)$, then:

- $\int_0^1 k(.,y)d\lambda = \theta \times [2-k(0,y)-k(1,y)]$
- $\iint_{[0,1]^2} k(.,.)d(\lambda\otimes\lambda) = 2\theta(1-\theta+\theta e^{-1/\theta})$

Example 6 (Matérn kernel, $\nu = p+\frac{1}{2}$). Define for $\nu = p+\frac{1}{2}$ ($p\in\mathbb{N}$):

$$k(x,y) = \frac{p!}{(2p)!}\sum_{i=0}^{p}\frac{(p+i)!}{i!(p-i)!}\left(\frac{|x-y|}{\theta/\sqrt{8\nu}}\right)^{p-i}\times\exp\left(-\frac{|x-y|}{\theta/\sqrt{2\nu}}\right).$$

Then, denoting $\zeta_p = \frac{\theta}{\sqrt{2\nu}}$, we have:

$$\int_0^1 k(.,y)d\lambda = \zeta_p\frac{p!}{(2p)!}\times\left[2c_{p,0}-A_p\left(\frac{y}{\zeta_p}\right)-A_p\left(\frac{1-y}{\zeta_p}\right)\right],$$

where

$$A_p(u) = \left(\sum_{\ell=0}^{p}c_{p,\ell}u^\ell\right)e^{-u}\quad\text{with}\quad c_{p,\ell} = \frac{1}{\ell!}\sum_{i=0}^{p-\ell}\frac{(p+i)!}{i!}2^{p-i}.$$

This generalizes Example 5, corresponding to $\nu = 1/2$. Also, this result can be written more explicitly for the commonly selected value $\nu = 3/2$ ($p=1,\zeta_1=\theta/\sqrt{3}$):

- $k(x,y) = \left(1+\frac{|x-y|}{\zeta_1}\right)\exp\left(-\frac{|x-y|}{\zeta_1}\right)$
- $\int_0^1 k(.,y)d\lambda = \zeta_1\times\left[4-A_1\left(\frac{y}{\zeta_1}\right)-A_1\left(\frac{1-y}{\zeta_1}\right)\right]$ with $A_1(u) = (2+u)e^{-u}$
- $\iint_{[0,1]^2} k(.,.)d(\lambda\otimes\lambda) = 2\zeta_1\left[2-3\zeta_1+(1+3\zeta_1)e^{-1/\zeta_1}\right]$

Similarly, for $\nu = 5/2$ ($p=2,\zeta_2=\theta/\sqrt{5}$):

- $k(x,y) = \left(1+\frac{|x-y|}{\zeta_2}+\frac{1}{3}\frac{(x-y)^2}{(\zeta_2)^2}\right)\exp\left(-\frac{|x-y|}{\zeta_2}\right)$
- $\int_0^1 k(.,y)d\lambda = \frac{1}{3}\zeta_2\times\left[16-A_2\left(\frac{y}{\zeta_2}\right)-A_2\left(\frac{1-y}{\zeta_2}\right)\right]$ with $A_2(u) = (8+5u+u^2)e^{-u}$
- $\iint_{[0,1]^2} k(.,.)d(\lambda\otimes\lambda) = \frac{1}{3}\zeta_2(16-30\,\zeta_2)+\frac{2}{3}(1+7\,\zeta_2+15\,(\zeta_2)^2)e^{-1/\zeta_2}$

Example 7 (Gaussian kernel). If $k(x,y) = \exp\left(-\frac{1}{2}\frac{(x-y)^2}{\theta^2}\right)$, then

- $\int_0^1 k(.,y)d\lambda = \theta\sqrt{2\pi}\times\left[\Phi\left(\frac{1-y}{\theta}\right)+\Phi\left(\frac{y}{\theta}\right)-1\right]$
- $\iint_{[0,1]^2} k(.,.)d(\lambda\otimes\lambda) = 2(e^{-1/(2\theta^2)}-1)+\theta\sqrt{2\pi}\times\left(2\Phi\left(\frac{1}{\theta}\right)-1\right)$

where Φ denotes the cdf of the standard normal distribution.