Incorporating structural priors in Gaussian Random Field models

David Ginsbourger [1]

Joint work with Nicolas Durrande, Olivier Roustant, Nicolas Lenz, and Dominic Schuhmacher

[1] Institute of Mathematical Statistics and Actuarial Science
University of Bern, Switzerland

MascotNum Workshop on Computer Experiments and Meta-models for Uncertainty Quantification

ETH Zürich, April 24. 2014

Outline

Outline

Black box functions

Here we mainly focus on cases where a system of interest can be modelled as (or involves) a costly-to-evaluate deterministic function:

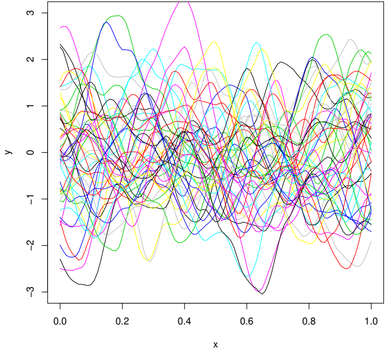
$$f : \mathbf{x} \in D \subset E \longrightarrow f(\mathbf{x}) \in F$$

for some given *input space* E and *output space* F –often $E \subset \mathbb{R}^d$ and $F \subset \mathbb{R}$.

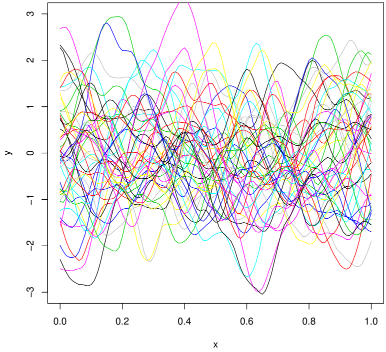
This talk is essentially about (Gaussian) Random Field models. . .

This talk is essentially about (Gaussian) Random Field models. . . *what do such models have to do with black box functions?*

This talk is essentially about (Gaussian) Random Field models. . . *what do such models have to do with black box functions?*

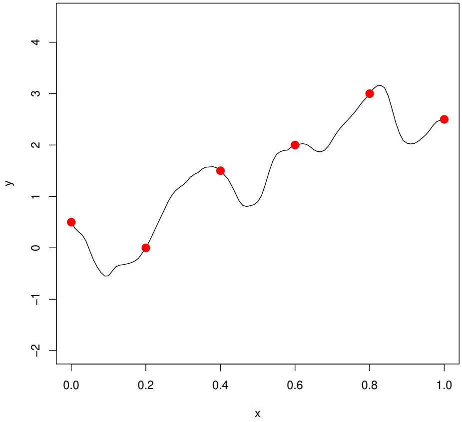


This talk is essentially about (Gaussian) Random Field models. . . *what do such models have to do with black box functions?*

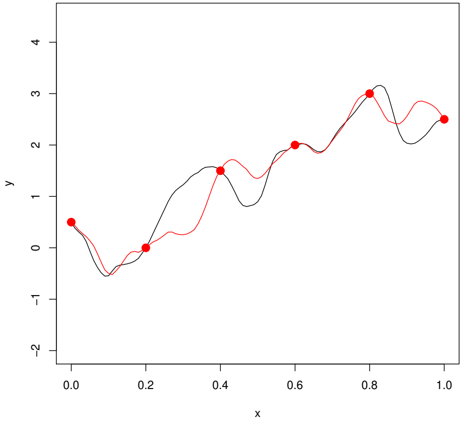


Actually, they can serve as prior distribution on function spaces!

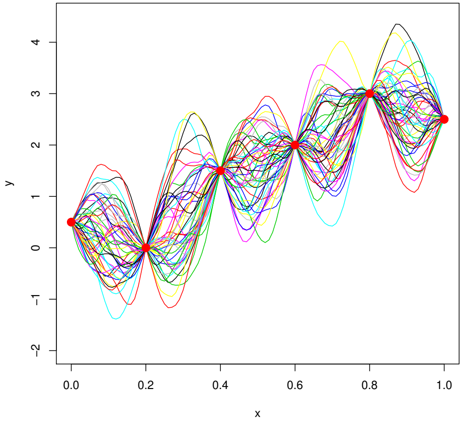
Conditional simulations (1D)



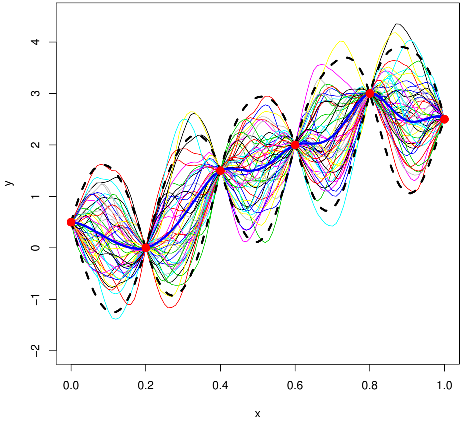
Conditional simulations (1D)



Conditional simulations (1D)



Conditional simulations and Kriging (1D)



Kriging at a glance: from geostats to machine learning

Originally, Kriging refers to ”optimal” linear prediction of a random field $(Z(\mathbf{x}))_{\mathbf{x}\in D}$ ($D \subset \mathbb{R}^2$ or $\mathbb{R}^3$) based on observations at $\mathbf{X}_n := \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, i.e.

$$A_n := \{(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n)) = \mathbf{z}_n\}$$

where $\mathbf{z}_n = (z(\mathbf{x}_1), \dots, z(\mathbf{x}_n))$ with $z(.) = Z(.;\omega)$ for some $\omega \in \Omega$.

Kriging at a glance: from geostats to machine learning

Originally, Kriging refers to "optimal" linear prediction of a random field $(Z(\mathbf{x}))_{\mathbf{x}\in D}$ ($D \subset \mathbb{R}^2$ or $\mathbb{R}^3$) based on observations at $\mathbf{X}_n := \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, i.e.

$$A_n := \{(Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)) = \mathbf{z}_n\}$$

where $\mathbf{z}_n = (z(\mathbf{x}_1), \ldots, z(\mathbf{x}_n))$ with $z(.) = Z(.;\omega)$ for some $\omega \in \Omega$.

Kriging may be cast as an ancester/a particular or more general case of various contemporary methods from different fields, including

- Gaussian Process Regression
- Interpolation Splines
- Kernel methods and regularization in RKHS

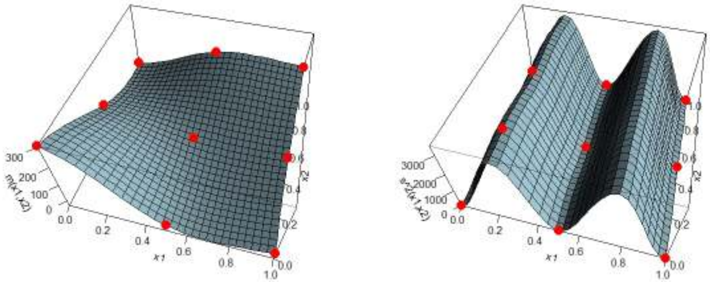
A few references about those 3 facets

C. E. Rasmussen and C. K. I. Williams (2006)
Gaussian Processes for Machine Learning
The MIT Press

G. Wahba (1990)
Spline Models for Observational Data
CBMS-NSF Regional Conference Series in Applied Mathematics

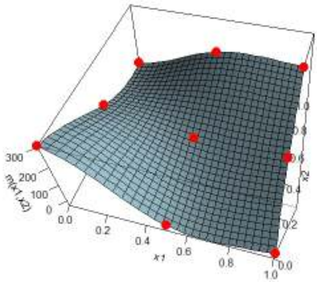
A. Berlinet, C. Thomas-Agnan (2004)
Reproducing Kernel Hilbert Spaces in Probability and Statistics
Kluwer Academic Publishers

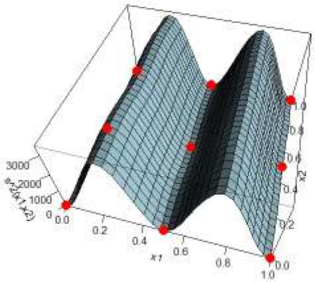
Interpolating deterministic functions by Kriging



Prediction by Kriging (based on 9 points) of the Branin-Hoo function.

Interpolating deterministic functions by Kriging





Prediction by Kriging (based on 9 points) of the Branin-Hoo function.

The covariance is here a **stationary** anisotropic Matérn kernel ($\nu = 5/2$) with scale and range parameters estimated by Maximum Likelihood.

Ordinary Kriging Equations –for completeness!–

Assume Z has a covariance kernel k, and constant mean $\mu \in \mathbb{R}$

Ordinary Kriging Equations –for completeness!–

Assume Z has a covariance kernel k, and constant mean $\mu \in \mathbb{R}$

$$\begin{cases} m_n(\mathbf{x}) = \mathbf{k}_n(\mathbf{x})^T \mathbf{K}_n^{-1} \mathbf{z}_n + \widehat{\mu}_n (1 - \mathbf{k}_n(\mathbf{x})^T \mathbf{K}_n^{-1} \mathbb{1}_n) \\ k_n(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}_n(\mathbf{x})^T \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}') + \frac{\left(1 - \mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x})\right)\left(1 - \mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}')\right)}{\left(\mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbb{1}_n\right)} \end{cases}$$

Ordinary Kriging Equations –for completeness!–

Assume Z has a covariance kernel k, and constant mean $\mu \in \mathbb{R}$

$$\begin{cases} m_n(\mathbf{x}) = \mathbf{k}_n(\mathbf{x})^T \mathbf{K}_n^{-1} \mathbf{z}_n + \widehat{\mu}_n (1 - \mathbf{k}_n(\mathbf{x})^T \mathbf{K}_n^{-1} \mathbb{1}_n) \\ k_n(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}_n(\mathbf{x})^T \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}') + \frac{(1 - \mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}))(1 - \mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}'))}{(\mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbb{1}_n)} \end{cases}$$

$$\mathbf{K}_n = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \dots & k(\mathbf{x}_1, \mathbf{x}_n) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \dots & k(\mathbf{x}_2, \mathbf{x}_n) \\ \dots & \dots & \dots. & \dots. \\ k(\mathbf{x}_n, \mathbf{x}_1) & \dots & \dots. & k(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix}, \mathbf{k}_n(\mathbf{x}) = \begin{pmatrix} k(\mathbf{x}, \mathbf{x}_1) \\ k(\mathbf{x}, \mathbf{x}_2) \\ \dots \\ k(\mathbf{x}, \mathbf{x}_n) \end{pmatrix}, \widehat{\mu}_n = \frac{\mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbf{z}_n}{(\mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbb{1}_n)}.$$

Ordinary Kriging Equations –for completeness!–

Assume Z has a covariance kernel k, and constant mean $\mu \in \mathbb{R}$

$$\begin{cases} m_n(\mathbf{x}) = \mathbf{k}_n(\mathbf{x})^T \mathbf{K}_n^{-1} \mathbf{z}_n + \widehat{\mu}_n (1 - \mathbf{k}_n(\mathbf{x})^T \mathbf{K}_n^{-1} \mathbb{1}_n) \\ k_n(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}, \mathbf{x}') - \mathbf{k}_n(\mathbf{x})^T \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}') + \frac{(1 - \mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}))(1 - \mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}'))}{(\mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbb{1}_n)} \end{cases}$$

$$\mathbf{K}_n = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \dots & k(\mathbf{x}_1, \mathbf{x}_n) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \dots & k(\mathbf{x}_2, \mathbf{x}_n) \\ \dots & \dots & \dots. & \dots. \\ k(\mathbf{x}_n, \mathbf{x}_1) & \dots & \dots. & k(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix}, \mathbf{k}_n(\mathbf{x}) = \begin{pmatrix} k(\mathbf{x}, \mathbf{x}_1) \\ k(\mathbf{x}, \mathbf{x}_2) \\ \dots \\ k(\mathbf{x}, \mathbf{x}_n) \end{pmatrix}, \widehat{\mu}_n = \frac{\mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbf{z}_n}{(\mathbb{1}_n^T \mathbf{K}_n^{-1} \mathbb{1}_n)}.$$

If μ is known (or with improper uniform prior) and $Z - \mu$ is assumed Gaussian, then m_n **and** k_n **are** Z**'s conditional mean and covariance** and

$$\boxed{\mathcal{L}(Z|A_n) = \mathcal{GRF}\left(m_n(\cdot), k_n(\cdot, \cdot')\right)}$$

More on the Bayesian approach: selected references

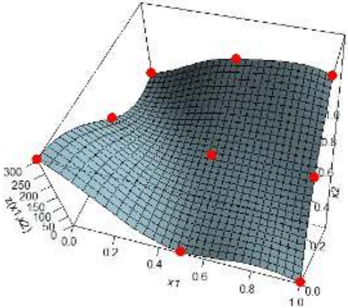
H. Omre and K. Halvorsen (1989).
The bayesian bridge between simple and universal kriging.
Mathematical Geology, 22 (7):767-786.

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

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

A.W. Van der Vaart and J. H. Van Zanten (2008)
Rates of contraction of posterior distributions based on Gaussian process priors.
Annals of Statistics, 36:1435-1463.

Conditional simulations of the Branin-Hoo function



In second-order random field models with constant mean, prior assumptions on f are implicitly accounted for through the choice of the covariance

$$k : (\mathbf{x}, \mathbf{x}') \in D \times D \longrightarrow k(\mathbf{x}, \mathbf{x}') = \text{cov}(Z_{\mathbf{x}}, Z_{\mathbf{x}'}) \in \mathbb{R}$$

In second-order random field models with constant mean, prior assumptions on f are implicitly accounted for through the choice of the covariance

$$k : (\mathbf{x}, \mathbf{x}') \in D \times D \longrightarrow k(\mathbf{x}, \mathbf{x}') = \text{cov}(Z_{\mathbf{x}}, Z_{\mathbf{x}'}) \in \mathbb{R}$$

Classical invariance notions for k

- Second-order stationarity (*k invariant under simultaneous translations of $\mathbf{x}$ and $\mathbf{x}'$*)
- Isotropy (*k invariant under simultaneous rigid motions of $\mathbf{x}$ and $\mathbf{x}'$*).

These properties are rather to be understood in a "mean square" sense.

In second-order random field models with constant mean, prior assumptions on f are implicitly accounted for through the choice of the covariance

$$k : (\mathbf{x}, \mathbf{x}') \in D \times D \longrightarrow k(\mathbf{x}, \mathbf{x}') = \mathrm{cov}(Z_{\mathbf{x}}, Z_{\mathbf{x}'}) \in \mathbb{R}$$

Classical invariance notions for k

- Second-order stationarity (*k invariant under simultaneous translations of $\mathbf{x}$ and $\mathbf{x}'$*)
- Isotropy (*k invariant under simultaneous rigid motions of $\mathbf{x}$ and $\mathbf{x}'$*).

These properties are rather to be understood in a "mean square" sense.

The main focus here is on functional properties of random field paths driven by k, both in Gaussian and in more general second-order settings.

Outline

Invariance under the action of a finite group

Let us assume that f is known *a priori* to be left unchanged by a set of symmetries (e.g., by physical arguments).

Is it possible to incorporate such "structural prior" into a random field model?

Invariance under the action of a finite group

Let us assume that f is known *a priori* to be left unchanged by a set of symmetries (e.g., by physical arguments).

Is it possible to incorporate such "structural prior" into a random field model?

Property

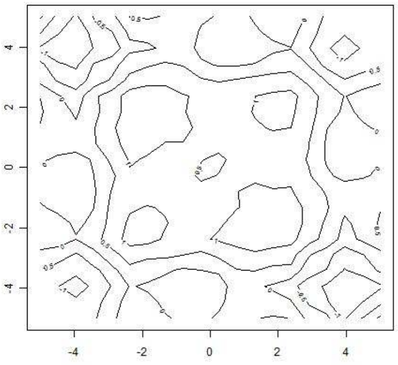
Let G be a finite groupe acting measurably on D via

$$\Phi : (\mathbf{x}, g) \in D \times G \longrightarrow \Phi(\mathbf{x}, g) = g.\mathbf{x} \in D$$

and Z be a second-order random field indexed by D with constant mean.

$$(\forall \mathbf{x} \in D,\ \mathbb{P}(\forall g \in G,\ Z_{\mathbf{x}} = Z_{g.\mathbf{x}}) = 1) \Leftrightarrow (\forall \mathbf{x} \in D,\ \forall g \in G,\ k(g.\mathbf{x}, \cdot) = k(\mathbf{x}, \cdot))$$

Invariant kernels enable invariant simulations



Another invariance: random fields with additive paths

Let $D = \prod_i^d D_i$ where $D_i \subset \mathbb{R}$. $f \in \mathbb{R}^D$ is called additive when there exists $f_i \in \mathbb{R}^{D_i}$ $(1 \leq i \leq d)$ such that $f(\mathbf{x}) = \sum_{i=1}^d f_i(x_i)$ $(\mathbf{x} = (x_1, \ldots, x_d) \in D)$.

Another invariance: random fields with additive paths

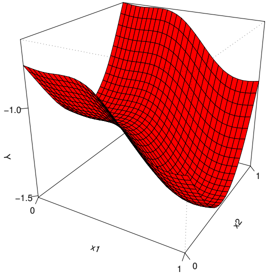
Let $D = \prod_i^d D_i$ where $D_i \subset \mathbb{R}$. $f \in \mathbb{R}^D$ is called additive when there exists $f_i \in \mathbb{R}^{D_i}$ $(1 \leq i \leq d)$ such that $f(\mathbf{x}) = \sum_{i=1}^d f_i(x_i)$ $(\mathbf{x} = (x_1, \ldots, x_d) \in D)$.

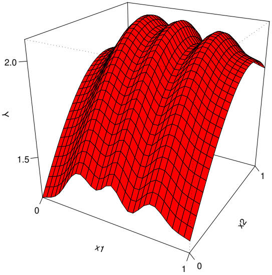
GRF models possessing additive paths (with $k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^d k_i(x_i, x_i'))$ have been considered in Nicolas Durrande's Ph.D. thesis (2011):

Another invariance: random fields with additive paths

Let $D = \prod_i^d D_i$ where $D_i \subset \mathbb{R}$. $f \in \mathbb{R}^D$ is called additive when there exists $f_i \in \mathbb{R}^{D_i}$ $(1 \leq i \leq d)$ such that $f(\mathbf{x}) = \sum_{i=1}^d f_i(x_i)$ $(\mathbf{x} = (x_1, \ldots, x_d) \in D)$.

GRF models possessing additive paths (with $k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^d k_i(x_i, x_i'))$ have been considered in Nicolas Durrande's Ph.D. thesis (2011):





Pathwise properties of fields with invariant kernels

Definition: Composition operator

Let us consider a (non-necessarily bi/in/sur-jective) function $v : \mathbf{x} \in D \longrightarrow v(\mathbf{x}) \in D$.

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

defines the *composition operator* associated with v.

Pathwise properties of fields with invariant kernels

Definition: Composition operator

Let us consider a (non-necessarily bi/in/sur-jective) function $v : \mathbf{x} \in D \longrightarrow v(\mathbf{x}) \in D$.

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

defines the *composition operator* associated with v.

Property

Let Z be a centred second-order RF with covariance kernel k and T be a finite linear combination of composition operators. Then k is T-invariant, i.e.

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

If and only if $\mathbb{P}\left(Z_{\mathbf{x}} = T(Z)_{\mathbf{x}}\right) = 1 \;\; (\mathbf{x} \in D)$.

Particular case of additivity

One can show that f is additive $\iff$ f is invariant under

$$T(f)(\mathbf{x}) = \sum_{i=1}^{d} f(\mathbf{v}_i(\mathbf{x})) - (d-1)f(\mathbf{a})$$

where $\mathbf{a} \in D$ is arbitrary and $\mathbf{v}_i(\mathbf{x}) = (a_1, \ldots, a_{i-1}, \underbrace{x_i}_{\text{ith coordinate}}, a_{i+1}, \ldots, a_d)$

This leads to Z additive **if and only** if k (is positive definite and) writes

$$\boxed{k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{d} \sum_{j=1}^{d} k_{ij}(x_i, x_j')}$$

Particular case of additivity

One can show that f is additive $\Longleftrightarrow$ f is invariant under

$$T(f)(\mathbf{x}) = \sum_{i=1}^{d} f(\mathbf{v}_i(\mathbf{x})) - (d-1)f(\mathbf{a})$$

where $\mathbf{a} \in D$ is arbitrary and $\mathbf{v}_i(\mathbf{x}) = (a_1, \dots, a_{i-1}, \underbrace{x_i}_{\text{ith coordinate}}, a_{i+1}, \dots, a_d)$

This leads to Z additive **if and only** if k (is positive definite and) writes

$$\boxed{k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{d} \sum_{j=1}^{d} k_{ij}(x_i, x_j')}$$

Particular case of group invariance

$T(f)(\mathbf{x}) = \sum_{i=1}^{\#G} \frac{1}{\#G} f(\mathbf{v}_i(\mathbf{x}))$ with $\mathbf{v}_i(\mathbf{x}) := g_i.\mathbf{x}$
leads to Z **Φ-invariant if and only if k is argumentwise invariant**.

Extension to further operators in the Gaussian case

In the Gaussian case, the last results can be extended to a wider class of operators using the Loève isometry Ψ between $\mathcal{L}(Z)$ (The Hilbert space generated by Z) and the RKHS associated with k, $\mathcal{H}(k)$.

Extension to further operators in the Gaussian case

In the Gaussian case, the last results can be extended to a wider class of operators using the Loève isometry Ψ between $\mathcal{L}(Z)$ (The Hilbert space generated by Z) and the RKHS associated with k, $\mathcal{H}(k)$.

Let T be an operator defined on the paths of Z such that $T(Z)_{\mathbf{x}} \in \mathcal{L}(Z)$ $(\mathbf{x} \in D)$. T induces an operator $\mathcal{T}$ from $\mathcal{H}(k)$ to $\mathbb{R}^D$, defined by

$$\boxed{\mathcal{T}(h)(\mathbf{x}) = \mathrm{cov}(T(Z)_{\mathbf{x}}, \Psi(h))}$$

Extension to further operators in the Gaussian case

In the Gaussian case, the last results can be extended to a wider class of operators using the Loève isometry Ψ between $\mathcal{L}(Z)$ (The Hilbert space generated by Z) and the RKHS associated with k, $\mathcal{H}(k)$.

Let T be an operator defined on the paths of Z such that $T(Z)_{\mathbf{x}} \in \mathcal{L}(Z)$ $(\mathbf{x} \in D)$. T induces an operator $\mathcal{T}$ from $\mathcal{H}(k)$ to $\mathbb{R}^D$, defined by

$$\boxed{\mathcal{T}(h)(\mathbf{x}) = \mathrm{cov}(T(Z)_{\mathbf{x}}, \Psi(h))}$$

Theorem

$$(\forall x \in D,\ \mathbb{P}(Z_{\mathbf{x}} = T(Z)_{\mathbf{x}}) = 1) \Leftrightarrow (\mathcal{T} = \mathrm{Id}_{\mathcal{H}})$$

Examples (Gaussian case)

a) Let ν be a measure on D s.t. $\int_D \sqrt{k(\mathbf{u},\mathbf{u})}d\nu(\mathbf{u}) < +\infty$.
Then Z has centred paths iff $\int_D k(\mathbf{x},\mathbf{u})\mathrm{d}\nu(\mathbf{u}) = 0, \forall \mathbf{x} \in D$.

For instance, given any p.d. kernel k, k_0 defined by

$$k_0(\mathbf{x},\mathbf{y}) = k(\mathbf{x},\mathbf{y}) - \int k(\mathbf{x},\mathbf{u})\mathrm{d}\nu(\mathbf{u}) - \int k(\mathbf{y},\mathbf{u})\mathrm{d}\nu(\mathbf{u}) + \int k(\mathbf{u},\mathbf{v})\mathrm{d}\nu(\mathbf{u})\mathrm{d}\nu(\mathbf{v})$$

satisfies the above condition.

Examples (Gaussian case)

a) Let ν be a measure on D s.t. $\int_D \sqrt{k(\mathbf{u},\mathbf{u})} d\nu(\mathbf{u}) < +\infty$.
Then Z has centred paths iff $\int_D k(\mathbf{x},\mathbf{u}) \mathrm{d}\nu(\mathbf{u}) = 0, \forall \mathbf{x} \in D$.

For instance, given any p.d. kernel k, k_0 defined by

$$k_0(\mathbf{x},\mathbf{y}) = k(\mathbf{x},\mathbf{y}) - \int k(\mathbf{x},\mathbf{u})\mathrm{d}\nu(\mathbf{u}) - \int k(\mathbf{y},\mathbf{u})\mathrm{d}\nu(\mathbf{u}) + \int k(\mathbf{u},\mathbf{v})\mathrm{d}\nu(\mathbf{u})\mathrm{d}\nu(\mathbf{v})$$

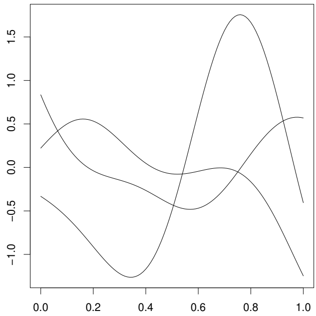
satisfies the above condition.

b) Solutions to the *Laplace equation* are called harmonic functions. Let us call harmonic any p.d. kernel solving the Laplace equation argumentwise: $(\Delta k(\cdot,\mathbf{x}')) = 0$ $(\mathbf{x}' \in D)$.

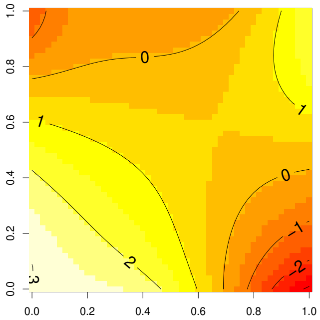
An example of such harmonic kernel over $\mathbb{R}^2 \times \mathbb{R}^2$ can be found in the recent literature (Schaback et al. 2009):

$$k_{harm}(\mathbf{x},\mathbf{y}) = \exp\left(\frac{x_1 y_1 + x_2 y_2}{\theta^2}\right) \cos\left(\frac{x_2 y_1 - x_1 y_2}{\theta^2}\right).$$

Example sample paths invariant under various T's

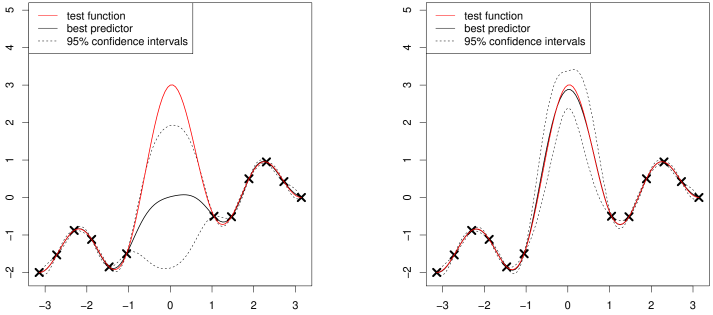


(a) Zero-mean paths of the centred GP with kernel k_0.



(b) Harmonic path of a GRF with kernel k_{harm}.

Kriging with invariant kernels: example a)

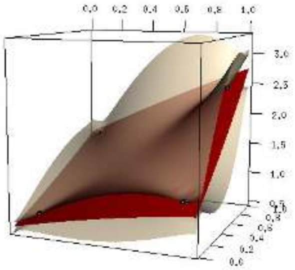


(c) GPR with kernel k

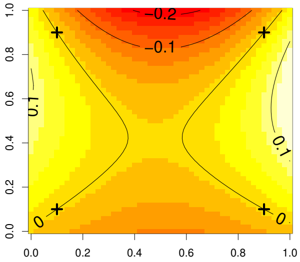
(d) GPR with kernel k_0

Figure: Comparison of two kriging models. The left one is based on a Gaussian kernel. The right one incorporates the zero-mean property.

Kriging with invariant kernels: example b)



(a) Mean predictor and 95% confidence intervals



(b) prediction error

Figure: Example of kriging model based on a harmonic kernel.

Outline

Set up: Functional ANOVA decomposition

Specific assumptions on D and f: $D = D_1 \times \cdots \times D_d$, where each $D_i \subset \mathbb{R}$ is endowed with a probability measure μ_i, D is endowed with the product measure $\mu := \mu_1 \times \cdots \times \mu_d$, and $f \in L^2(\mu)$.

The Functional ANOVA (or *Sobol'-Hoeffding*) decomposition consists in expanding f into a sum of orthogonal terms of increasing complexity:

$$\boxed{f = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}} f_{\mathbf{u}}}$$

Set up: Functional ANOVA decomposition

Specific assumptions on D and f: $D = D_1 \times \cdots \times D_d$, where each $D_i \subset \mathbb{R}$ is endowed with a probability measure μ_i, D is endowed with the product measure $\mu := \mu_1 \times \cdots \times \mu_d$, and $f \in L^2(\mu)$.

The Functional ANOVA (or *Sobol'-Hoeffding*) decomposition consists in expanding f into a sum of orthogonal terms of increasing complexity:

$$\boxed{f = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}} f_{\mathbf{u}}}$$

By orthogonality, $||f||^2_{L^2(\mu)} = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}} \sigma^2_{\mathbf{u}}$, where $\sigma^2_{\mathbf{u}} := ||f_{\mathbf{u}}||^2_{L^2(\mu)}$.

The ratios $S_{\mathbf{u}} := \sigma^2_{\mathbf{u}} / (\sum_{\mathbf{v} \neq \emptyset} \sigma^2_{\mathbf{v}})$ are referred to as Sobol' indices ($\mathbf{u} \neq \emptyset$).

Before going further: A few fundamental references

W. Hoeffding (1948)
A class of statistics with asymptotically normal distribution
Annals of Mathematical Statistics, 19, 293-325

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

A. Antoniadis (1984)
Analysis of variance on function spaces
Math. Oper. Forsch. und Statist., series Statistics, 15(1):59-71

I.M. Sobol' (1993)
Sensitivity estimates for nonlinear mathematical models
Mathematical Modelling and Computational Experiments, 1:407-414.

Revisiting FANOVA: an operator approach

Let $\mathcal{F}$ be a subspace of $\mathbb{R}^D$, P_j ($1 \leq j \leq d$) be set of commuting projections on $\mathcal{F}$ s.t. $P_j(f) = f$ if f does not depend on x_j, and $P_j(f)$ does not depend on x_j.

Revisiting FANOVA: an operator approach

Let $\mathcal{F}$ be a subspace of $\mathbb{R}^D$, P_j ($1 \leq j \leq d$) be set of commuting projections on $\mathcal{F}$ s.t. $P_j(f) = f$ if f does not depend on x_j, and $P_j(f)$ does not depend on x_j.

The identity operator $I_{\mathcal{F}} : \mathcal{F} \longrightarrow \mathcal{F}$ can be decomposed as

$$I_{\mathcal{F}} = \prod_{j=1}^{d} [(I - P_j) + P_j] = \sum_{\mathbf{u} \subseteq \{1,\dots,d\}} \overbrace{\left(\prod_{j \in \mathbf{u}} (I - P_j) \right) \left(\prod_{j \in \{1,\dots,d\} \setminus \mathbf{u}} P_j \right)}^{T_{\mathbf{u}}},$$

Revisiting FANOVA: an operator approach

Let $\mathcal{F}$ be a subspace of $\mathbb{R}^D$, P_j ($1 \leq j \leq d$) be set of commuting projections on $\mathcal{F}$ s.t. $P_j(f) = f$ if f does not depend on x_j, and $P_j(f)$ does not depend on x_j.

The identity operator $I_{\mathcal{F}} : \mathcal{F} \longrightarrow \mathcal{F}$ can be decomposed as

$$I_{\mathcal{F}} = \prod_{j=1}^{d} [(I - P_j) + P_j] = \sum_{\mathbf{u} \subseteq \{1,\dots,d\}} \overbrace{\left(\prod_{j \in \mathbf{u}} (I - P_j)\right) \left(\prod_{j \in \{1,\dots,d\} \setminus \mathbf{u}} P_j\right)}^{T_{\mathbf{u}}},$$

whereof $\forall f \in \mathcal{F}, f = \sum_{\mathbf{u} \subseteq \{1,\dots,d\}} f_{\mathbf{u}}$ with $f_{\mathbf{u}} := T_{\mathbf{u}}(f)$.

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

The standard FANOVA is obtained when $\mathcal{F}$ is the set of square integrable functions on $D = [0,1]^d$, and the P_j's are partial integration operators:

$$P_j(f)(\mathbf{x}) = \int_0^1 f(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_d) \mathrm{d}\mu_j(t),$$

leading to the following operators:

$$T_{\mathbf{u}}(f) = \sum_{\mathbf{v} \subseteq \mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|} \int_{[0,1]^{d-|\mathbf{v}|}} f(\mathbf{x}) \mathrm{d}\mu_{-\mathbf{v}}(\mathbf{x}_{-\mathbf{v}})$$

where $\mu_{-\mathbf{v}}$ denotes $\prod_{j \in [1 \ldots d] \setminus \mathbf{v}} \mu_j$ and $\mathbf{x}_{-\mathbf{v}}$ is defined accordingly.

The standard FANOVA is obtained when $\mathcal{F}$ is the set of square integrable functions on $D = [0,1]^d$, and the P_j's are partial integration operators:

$$P_j(f)(\mathbf{x}) = \int_0^1 f(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_d) \mathrm{d}\mu_j(t),$$

leading to the following operators:

$$T_{\mathbf{u}}(f) = \sum_{\mathbf{v} \subseteq \mathbf{u}} (-1)^{|\mathbf{u}| - |\mathbf{v}|} \int_{[0,1]^{d-|\mathbf{v}|}} f(\mathbf{x}) \mathrm{d}\mu_{-\mathbf{v}}(\mathbf{x}_{-\mathbf{v}})$$

where $\mu_{-\mathbf{v}}$ denotes $\prod_{j \in [1\dots d] \setminus \mathbf{v}} \mu_j$ and $\mathbf{x}_{-\mathbf{v}}$ is defined accordingly.

Example: Low order terms of the decomposition ($i, j \in \{1, \dots, d\}$)

$$T_{\emptyset}(f) = \int_{[0,1]^d} f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})$$

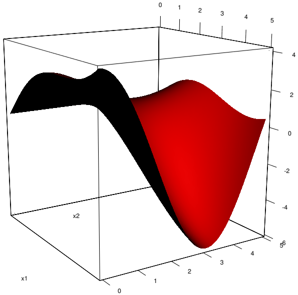
$$T_{\{i\}}(f)(x_i) = \int_{[0,1]^{d-1}} f(\mathbf{x}) \mathrm{d}\mu_{-\{i\}}(\mathbf{x}_{-\{i\}}) - \int_{[0,1]^d} f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})$$

$$T_{\{i,j\}}(f)(x_i, x_j) = \int_{[0,1]^{d-2}} f(\mathbf{x}) \mathrm{d}\mu_{-\{i,j\}}(\mathbf{x}_{-\{i,j\}}) - \int_{[0,1]^{d-1}} f(\mathbf{x}) \mathrm{d}\mu_{-\{i\}}(\mathbf{x}_{-\{i\}}) - \int_{[0,1]^{d-1}} f(\mathbf{x}) \mathrm{d}\mu_{-\{j\}}(\mathbf{x}_{-\{j\}}) + \int_{[0,1]^d} f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})$$

Example on a simple test function

We consider the following test function over $[0,5]^2$:

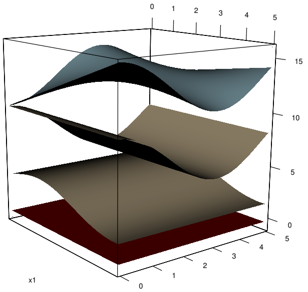
$$f(\mathbf{x}) = \sin(x_1) + x_1 \cos(x_2)$$



Example on a simple test function

The FANOVA decomposition of f can be obtained analytically:

$$f_{\emptyset}(\mathbf{x}) = 0.2(1 - \cos(5)) + 0.5\sin(5)$$
$$f_{\{1\}}(\mathbf{x}) = \sin(x_1) + 0.2x_1\sin(5) - f_{\emptyset}$$
$$f_{\{2\}}(\mathbf{x}) = 0.2(1 - \cos(5)) + 0.1\cos(x_2) - f_{\emptyset}$$
$$f_{\{1,2\}}(\mathbf{x}) = f(\mathbf{x}) - f_{\{1\}}(\mathbf{x}) - f_{\{2\}}(\mathbf{x}) - f_{\emptyset}$$



Back to the set up: how do deal with a costly f?

Assuming that the value of f at points $\mathbf{X}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset D$ is known, how to estimate ANOVA decompositions terms and Sobol' indices?

Back to the set up: how do deal with a costly f?

Assuming that the value of f at points $\mathbf{X}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset D$ is known, how to estimate ANOVA decompositions terms and Sobol' indices?

Popular workflow: replace f in ANOVA decompositions and global SA by a (cheaper) approximation $\tilde{f}$ based on $\{(\mathbf{x}_i, f(\mathbf{x}_i)), 1 \leq i \leq n\}$, e.g.,

- Standard linear models
- Polynomial chaos models (Sudret et al.)
- Smoothing spline models (Wahba et al.)

Back to the set up: how do deal with a costly f?

Assuming that the value of f at points $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset D$ is known, how to estimate ANOVA decompositions terms and Sobol' indices?

Popular workflow: replace f in ANOVA decompositions and global SA by a (cheaper) approximation $\tilde{f}$ based on $\{(\mathbf{x}_i, f(\mathbf{x}_i)), 1 \leq i \leq n\}$, e.g.,

- Standard linear models
- Polynomial chaos models (Sudret et al.)
- Smoothing spline models (Wahba et al.)
- Kriging and Gaussian random field (GRF) models

About Oakley and O'Hagan's contributions

O&O'H have suggested to estimate ANOVA terms in the Bayesian framework, where a GRF model, say $(Z_{\mathbf{x}})_{\mathbf{x} \in D}$, is assumed for f. Analytical expressions of the posterior means of ANOVA terms were derived in

J.E. Oakley and A. O'Hagan (2004)
Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach
Journal of the Royal Statistical Society (Series B), 66(3):751-769

About Oakley and O'Hagan's contributions

O&O'H have suggested to estimate ANOVA terms in the Bayesian framework, where a GRF model, say $(Z_{\mathbf{x}})_{\mathbf{x} \in D}$, is assumed for f. Analytical expressions of the posterior means of ANOVA terms were derived in

J.E. Oakley and A. O'Hagan (2004)
Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach
Journal of the Royal Statistical Society (Series B), 66(3):751-769

- Posterior means computed by multi-dimensional numerical integration.
- k is by default a stationary Gaussian kernel
- $S_{\mathbf{u}}$'s are estimated through a ratio of posterior means.

About Marrel et al.'s contribution

Marrel et al. have proposed to investigate (approximate) posterior distributions of the $S_{\mathbf{u}}$'s by appealing to conditional simulations.

A. Marrel, B. Iooss, B. Laurent, and O. Roustant (2009)
Calculations of Sobol indices for the Gaussian process metamodel
Reliability Engineering and System Safety 94:742-751

About Marrel et al.'s contribution

Marrel et al. have proposed to investigate (approximate) posterior distributions of the $S_{\mathbf{u}}$'s by appealing to conditional simulations.

A. Marrel, B. Iooss, B. Laurent, and O. Roustant (2009)
Calculations of Sobol indices for the Gaussian process metamodel
Reliability Engineering and System Safety 94:742-751

$\sigma_{\mathbf{u}}^2$'s are simulated by combining the known distributions of the $T_{\mathbf{u}}(Z)$'s and numerical approximation schemes for the integrals.

About Marrel et al.'s contribution

Marrel et al. have proposed to investigate (approximate) posterior distributions of the $S_{\mathbf{u}}$'s by appealing to conditional simulations.

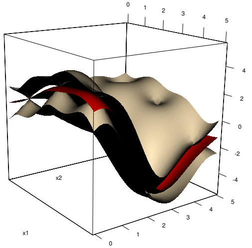
A. Marrel, B. Iooss, B. Laurent, and O. Roustant (2009)
Calculations of Sobol indices for the Gaussian process metamodel
Reliability Engineering and System Safety 94:742-751

$\sigma_{\mathbf{u}}^2$'s are simulated by combining the known distributions of the $T_{\mathbf{u}}(Z)$'s and numerical approximation schemes for the integrals.

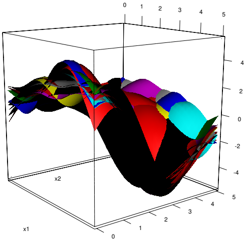
- Again, k is chosen among standard stationary covariance kernels
- The approach proves useful on a 20-dimensional test case
- Links between the chosen k and Sobol' indices are not discussed

Back to our simple test function

Here the test function $f(\mathbf{x}) = \sin(x_1) + x_1 \cos(x_2)$ is evaluated at a 9-point grid design, and a GRF model with Gaussian kernel is fitted to the data.



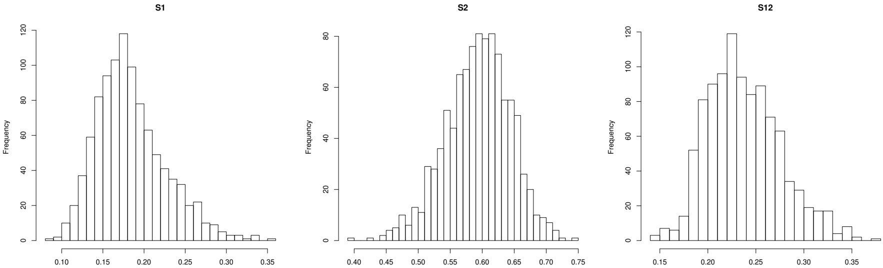
GRF model



GRF conditional simulations

Back to our simple test function

Following Marrel et al., posterior distributions of Sobol' indices are approximated relying on numerical integration and Monte Carlo :



About Durrande et al.'s contribution

Durrande et al. have focused on the choice of k, and showed that for a particular class of so-called ANOVA kernels

$$k(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{d} (1 + k_0(x_i, x_i')),$$

the FANOVA decomposition of the kriging mean predictor m can be calculated without numerical integration.

About Durrande et al.'s contribution

Durrande et al. have focused on the choice of k, and showed that for a particular class of so-called ANOVA kernels

$$k(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{d} (1 + k_0(x_i, x_i')),$$

the FANOVA decomposition of the kriging mean predictor m can be calculated without numerical integration. Further, m's Sobol' indices write:

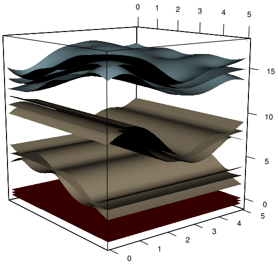
$$S_{\mathbf{u}}(m) = \frac{Z_{\mathbf{X}_n}^t \mathrm{K}^{-1} \left(\bigodot_{i \in \mathbf{u}} \Gamma_i \right) \mathrm{K}^{-1} Z_{\mathbf{X}_n}}{Z_{\mathbf{X}_n}^t \mathrm{K}^{-1} \left(\bigodot_{i=1}^{d} (1_{n \times n} + \Gamma_i) - 1_{n \times n} \right) \mathrm{K}^{-1} Z_{\mathbf{X}_n}}$$

where Γ_i is the $n \times n$ matrix $\Gamma_i = \int_{D_i} \mathbf{k}_0^i(x_i) \mathbf{k}_0^i(x_i)^t \mathrm{d}\mu_i(x_i)$.

N. Durrande, D. Ginsbourger, O. Roustant, and L. Carraro (2013)
ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis
Journal of Multivariate Analysis, 155:57-67

Back to our simple example

Using an ad hoc ANOVA kernel, the FANOVA decomposition of the GRF posterior mean is obtained analytically:



Focus and starting research questions

Claim: Assuming that f is some realization of a centred GRF Z with kernel k goes with implicit assumptions concerning f's FANOVA decomposition. . .

Focus and starting research questions

Claim: Assuming that f is some realization of a centred GRF Z with kernel k goes with implicit assumptions concerning f's FANOVA decomposition...

- How are the different FANOVA terms (jointly) distributed?
- What is the interplay between k and this distribution?
- How does conditioning on data affect those results?
- What consequences for Sobol' indices estimation under a GRF prior?

A fundamental decomposition result for GRFs

Let $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ be a centred GRF with squared-integrable paths (a.s.) and denote by $k : D \times D \to \mathbb{R}$ its covariance kernel.

Then Z admits the following *pathwise ANOVA decomposition* almost surely:

$$\boxed{Z_{\cdot} = \sum_{\mathbf{u}\subseteq\{1,\dots,d\}} T_{\mathbf{u}}(Z)_{\cdot}}$$

where the $(T_{\mathbf{u}}(Z)_{\mathbf{x}})_{\mathbf{x}\in D}$ are centred GRFs with respective covariance kernels $T_{\mathbf{u}} \otimes T_{\mathbf{u}}(k)$ ($\mathbf{u} \subseteq \{1,\dots,d\}$).

Moreover, $(T_{\mathbf{u}}(Z)_{\mathbf{x}}; \mathbf{u} \subseteq \{1,\dots,d\})_{\mathbf{x}\in D}$ defines a 2^d-dimensional vector-valued GRF with cross-covariances $T_{\mathbf{u}} \otimes T_{\mathbf{v}}(k)$ $((\mathbf{u},\mathbf{v}) \subseteq \{1,\dots,d\}^2)$.

A fundamental decomposition result for GRFs

Let $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ be a centred GRF with squared-integrable paths (a.s.) and denote by $k : D \times D \to \mathbb{R}$ its covariance kernel.

Then Z admits the following *pathwise ANOVA decomposition* almost surely:

$$\boxed{Z_{\cdot} = \sum_{\mathbf{u}\subseteq\{1,\dots,d\}} T_{\mathbf{u}}(Z)_{\cdot}}$$

where the $(T_{\mathbf{u}}(Z)_{\mathbf{x}})_{\mathbf{x}\in D}$ are centred GRFs with respective covariance kernels $T_{\mathbf{u}} \otimes T_{\mathbf{u}}(k)$ ($\mathbf{u} \subseteq \{1,\dots,d\}$).

Moreover, $(T_{\mathbf{u}}(Z)_{\mathbf{x}}; \mathbf{u} \subseteq \{1,\dots,d\})_{\mathbf{x}\in D}$ defines a 2^d-dimensional vector-valued GRF with cross-covariances $T_{\mathbf{u}} \otimes T_{\mathbf{v}}(k)$ $((\mathbf{u},\mathbf{v}) \subseteq \{1,\dots,d\}^2)$.

Remark: The same would hold for any finite additive decomposition of the identity operator on a space containing the paths with probability 1.

Example: The 1-dimensional Brownian Motion

Let $(B_t)_{t\in[0,1]}$ be a Brownian Motion on $[0,1]$, i.e. a centred Gaussian field ($d = 1 \to$ "*process*") with kernel $(s,t) \in [0,1]^2 \to k(s,t) = \min(s,t)$.

The 1-dimensional pathwise ANOVA decomposition of B writes

$$B_t = \underbrace{\int_0^1 B_v dv}_{\left(T_\emptyset B\right)_t} + \underbrace{\left(B_t - \int_0^1 B_v \mathrm{d}v\right)}_{\left(T_{\{1\}} B\right)_t}$$

Example: The 1-dimensional Brownian Motion

Let $(B_t)_{t\in[0,1]}$ be a Brownian Motion on $[0,1]$, i.e. a centred Gaussian field ($d = 1 \to$ "*process*") with kernel $(s,t) \in [0,1]^2 \to k(s,t) = \min(s,t)$.

The 1-dimensional pathwise ANOVA decomposition of B writes

$$B_t = \underbrace{\int_0^1 B_v dv}_{\color{blue}{(T_\emptyset B)_t}} + \underbrace{\left(B_t - \int_0^1 B_v \mathrm{d}v \right)}_{\color{red}{(T_{\{1\}} B)_t}}$$

$T_\emptyset B$ is a centred Gaussian process with covariance kernel

$$(T_\emptyset \otimes T_\emptyset k)(s,t) = \left(\int_0^1 \int_0^1 \min(s,t) \mathrm{d}s\mathrm{d}t \right) \mathbf{1}_{[0,1]^2}(s,t) = \color{blue}{\frac{1}{3} \mathbf{1}_{[0,1]^2}(s,t)}$$

Example: The 1-dimensional Brownian Motion

Let $(B_t)_{t\in[0,1]}$ be a Brownian Motion on $[0,1]$, i.e. a centred Gaussian field ($d=1 \to$ "*process*") with kernel $(s,t)\in[0,1]^2 \to k(s,t)=\min(s,t)$.

The 1-dimensional pathwise ANOVA decomposition of B writes

$$B_t = \underbrace{\int_0^1 B_v dv}_{\color{blue}{(T_\emptyset B)_t}} + \underbrace{\left(B_t - \int_0^1 B_v \mathrm{d}v\right)}_{\color{red}{(T_{\{1\}} B)_t}}$$

$T_\emptyset B$ is a centred Gaussian process with covariance kernel

$$(T_\emptyset \otimes T_\emptyset k)(s,t) = \left(\int_0^1\int_0^1 \min(s,t)\mathrm{d}s\mathrm{d}t\right)\mathbf{1}_{[0,1]^2}(s,t) = \color{blue}{\frac{1}{3}\mathbf{1}_{[0,1]^2}(s,t)}$$

$T_{\{1\}} B$ is a centred Gaussian process with covariance kernel

$$(T_{\{1\}} \otimes T_{\{1\}} k)(s,t) = \color{red}{\left(\min(s,t) - (t-\frac{t^2}{2}) - (s-\frac{s^2}{2}) + \frac{1}{3}\right)\mathbf{1}_{[0,1]^2}(s,t)}$$

Associated ANOVA decomposition of p.d. kernels

Likewise, starting from a squared-integrable p.d. kernel $k : D \times D \to \mathbb{R}$, we obtain the following "double" (*tensor product*) decomposition :

$$\boxed{k = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}} \sum_{\mathbf{v} \subseteq \{1,\ldots,d\}} T_{\mathbf{u}} \otimes T_{\mathbf{v}}(k).}$$

Associated ANOVA decomposition of p.d. kernels

Likewise, starting from a squared-integrable p.d. kernel $k : D \times D \to \mathbb{R}$, we obtain the following "double" (*tensor product*) decomposition :

$$k = \sum_{\mathbf{u} \subseteq \{1,\dots,d\}} \sum_{\mathbf{v} \subseteq \{1,\dots,d\}} T_{\mathbf{u}} \otimes T_{\mathbf{v}}(k).$$

Remark: This decomposition may be seen as consequence of the previous result, or directly obtained by identifying $I_{L^2(D \times D)}$ *with* $I_{L^2(D)} \times I_{L^2(D)}$.

Comments

- This decomposition consists of 2^{2d} terms! (e.g., 16 terms for $d = 2$)
- $T_{\mathbf{u}} \otimes T_{\mathbf{v}}(k)$ can actually be interpreted as the orthogonal projection of k onto $\mathrm{Ran}(T_{\mathbf{u}} \otimes T_{\mathbf{v}})$, in the sense of the tensor product structure.

Example: Back to the 1-dimensional Brownian Motion

The two projected processes $(T_\emptyset B)_t$ and $\left(T_{\{1\}} B\right)_t$ are correlated, with cross-covariance kernels

$$(T_\emptyset \otimes T_{\{1\}} k)(s,t) = \left(t - \frac{t^2}{2} - \frac{1}{3}\right) \mathbf{1}_{[0,1]^2}(s,t)$$

$$(T_{\{1\}} \otimes T_\emptyset k)(s,t) = \left(s - \frac{s^2}{2} - \frac{1}{3}\right) \mathbf{1}_{[0,1]^2}(s,t)$$

Example: Back to the 1-dimensional Brownian Motion

The two projected processes $(T_\emptyset B)_t$ and $\left(T_{\{1\}}B\right)_t$ are correlated, with cross-covariance kernels

$$(T_\emptyset \otimes T_{\{1\}}k)(s,t) = \left(t - \frac{t^2}{2} - \frac{1}{3}\right) \mathbf{1}_{[0,1]^2}(s,t)$$

$$(T_{\{1\}} \otimes T_\emptyset k)(s,t) = \left(s - \frac{s^2}{2} - \frac{1}{3}\right) \mathbf{1}_{[0,1]^2}(s,t)$$

To sum up, the double FANOVA decomposition of $k(s,t) = \min(s,t)$ writes

$$k(s,t) = \frac{1}{3}\mathbf{1}_{[0,1]^2}(s,t) + \left(s - \frac{s^2}{2} - \frac{1}{3}\right) \mathbf{1}_{[0,1]^2}(s,t) + \left(t - \frac{t^2}{2} - \frac{1}{3}\right) \mathbf{1}_{[0,1]^2}(s,t) + \left(\min(s,t) - (t - \frac{t^2}{2}) - (s - \frac{s^2}{2}) + \frac{1}{3}\right) \mathbf{1}_{[0,1]^2}(s,t)$$

Sparsity and independence properties

Let $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ be a centred GRF as before. Then, for any given $\mathbf{u} \subseteq \{1, \ldots, d\}$ the two following assertions are equivalent:

- $T_{\mathbf{u}} \otimes T_{\mathbf{u}}(k) = \mathbf{0}$ $(\mu \otimes \mu - \text{a.e.})$
- $\mathbb{P}(T_{\mathbf{u}} Z = \mathbf{0}) = 1$

Furthermore, for any $\mathbf{v} \subseteq \{1, \ldots, d\}$, we have the second equivalence

- $T_{\mathbf{u}} \otimes T_{\mathbf{v}}(k) = \mathbf{0}$ $(\mu \otimes \mu - \text{a.e.})$
- $T_{\mathbf{u}} Z$ and $T_{\mathbf{v}} Z$ are two independent GRFs

Sparsity and independence properties

Let $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ be a centred GRF as before. Then, for any given $\mathbf{u} \subseteq \{1, \ldots, d\}$ the two following assertions are equivalent:

- $T_{\mathbf{u}} \otimes T_{\mathbf{u}}(k) = \mathbf{0}$ ($\mu \otimes \mu -$ a.e.)
- $\mathbb{P}(T_{\mathbf{u}} Z = \mathbf{0}) = 1$

Furthermore, for any $\mathbf{v} \subseteq \{1, \ldots, d\}$, we have the second equivalence

- $T_{\mathbf{u}} \otimes T_{\mathbf{v}}(k) = \mathbf{0}$ ($\mu \otimes \mu -$ a.e.)
- $T_{\mathbf{u}} Z$ and $T_{\mathbf{v}} Z$ are two independent GRFs

Fundamental example: ANOVA kernels based on a zero-mean 1d kernel

Let Z be a centred GRF with kernel of the form $k(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{d} (1 + k_0(x_i, x_i'))$.

Property: If k_0 is argumentwise zero-mean, then the "non-diagonal" $T_{\mathbf{u}} \otimes T_{\mathbf{v}} k$'s are null, and so Z decomposes as a sum of independent $T_{\mathbf{u}} Z's$.

Spectral interpretation

Let $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ be decomposable as $Z_{\mathbf{x}} = \sum_{i=1}^{+\infty} \xi_i \phi_i(\mathbf{x})$, where $\xi_i \sim \mathcal{N}(0, \lambda_i)$ independently, and the ϕ_i's form an orthonormal basis of $L^2(\mu)$. Then,

$$\boxed{(T_{\mathbf{u}}Z)_{\mathbf{x}} = \sum_{i=1}^{+\infty} \xi_i (T_{\mathbf{u}}\phi_i)(\mathbf{x}) \quad (\mathbf{u} \subset \{1, \ldots, n\})}$$

Spectral interpretation

Let $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ be decomposable as $Z_{\mathbf{x}} = \sum_{i=1}^{+\infty} \xi_i \phi_i(\mathbf{x})$, where $\xi_i \sim \mathcal{N}(0, \lambda_i)$ independently, and the ϕ_i's form an orthonormal basis of $L^2(\mu)$. Then,

$$\boxed{(T_{\mathbf{u}} Z)_{\mathbf{x}} = \sum_{i=1}^{+\infty} \xi_i (T_{\mathbf{u}} \phi_i)(\mathbf{x}) \quad (\mathbf{u} \subset \{1, \dots, n\})}$$

From $\sigma^2_{\mathbf{u}}(Z) := ||T_{\mathbf{u}} Z||^2_{L^2(\mu)} = \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} \xi_i \xi_j \langle T_{\mathbf{u}} \phi_i, T_{\mathbf{u}} \phi_j \rangle_{L^2(\mu)}$, we then get

$$\mathbb{E}[\sigma^2_{\mathbf{u}}(Z)] = \sum_{i=1}^{+\infty} \lambda_i ||T_{\mathbf{u}} \phi_i||^2_{L^2(\mu)}$$

Spectral interpretation

Let $(Z_{\mathbf{x}})_{\mathbf{x}\in D}$ be decomposable as $Z_{\mathbf{x}} = \sum_{i=1}^{+\infty} \xi_i \phi_i(\mathbf{x})$, where $\xi_i \sim \mathcal{N}(0, \lambda_i)$ independently, and the ϕ_i's form an orthonormal basis of $L^2(\mu)$. Then,

$$(T_{\mathbf{u}}Z)_{\mathbf{x}} = \sum_{i=1}^{+\infty} \xi_i (T_{\mathbf{u}}\phi_i)(\mathbf{x}) \quad (\mathbf{u} \subset \{1, \dots, n\})$$

From $\sigma_{\mathbf{u}}^2(Z) := ||T_{\mathbf{u}}Z||_{L^2(\mu)}^2 = \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} \xi_i \xi_j \langle T_{\mathbf{u}}\phi_i, T_{\mathbf{u}}\phi_j \rangle_{L^2(\mu)}$, we then get

$$\mathbb{E}[\sigma_{\mathbf{u}}^2(Z)] = \sum_{i=1}^{+\infty} \lambda_i ||T_{\mathbf{u}}\phi_i||_{L^2(\mu)}^2$$

Using that $(\mathbb{P}(T_{\mathbf{u}}Z = \mathbf{0}) = 1) \Leftrightarrow (\mathbb{E}[\sigma_{\mathbf{u}}^2(Z)] = 0)$, both are then equivalent to

$$(\forall i : \lambda_i \neq 0, \quad T_{\mathbf{u}}\phi_i = \mathbf{0} \;\; \mu - \text{a.e.})$$

which is in turn also equivalent to $T_{\mathbf{u}} \otimes T_{\mathbf{u}}(k) = \mathbf{0}$ ($\mu \otimes \mu$ – a.e.).

About the distribution of Sobol' indices

Property

For a Gaussian random field Z possessing a Karhunen-Loève expansion $\sum_{i=1}^{+\infty} \sqrt{\lambda_i}\varepsilon_i\phi_i(\cdot)$, the Sobol' indices form a 2^d-dimensional random vector which probability distribution is characterized by

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

where the $Q_{\mathbf{u}}$'s are the quadratic forms on $\ell^2(\mathbb{R})$ defined by

$$Q_{\mathbf{u}}(\mathbf{e}_i, \mathbf{e}_j) = \sqrt{\lambda_i \lambda_j} \langle T_{\mathbf{u}}\phi_i, T_{\mathbf{u}}\phi_j \rangle_{L^2(\mu)},$$

and $\{\mathbf{e}_k, k \in \mathbb{N}\}$ is the canonical basis of $\ell^2(\mathbb{R})$.

Sparsity is stable under conditioning

Property

Let Z be a centred GRF with kernel k, $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset D$, $Z_{\mathbf{X}_n}$ be the vector of evaluations of Z at $\mathbf{X}_n$, and $\mathbf{u} \subset \{1, \ldots, n\}$. Then, the sparsity of Z with respect to $T_\mathbf{u}$ is preserved by conditioning, i.e.

$$\text{If } \mathbb{P}(T_\mathbf{u} Z = 0) = 1, \text{ then } \mathbb{P}(T_\mathbf{u} Z = 0 | Z_{\mathbf{X}_n} = \mathbf{z}_n) = 1 \quad (\forall \mathbf{z}_n \in \mathbb{R}^n)$$

Sparsity is stable under conditioning

Property

Let Z be a centred GRF with kernel k, $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset D$, $Z_{\mathbf{X}_n}$ be the vector of evaluations of Z at $\mathbf{X}_n$, and $\mathbf{u} \subset \{1, \ldots, n\}$. Then, the sparsity of Z with respect to $T_{\mathbf{u}}$ is preserved by conditioning, i.e.

$$\text{If } \mathbb{P}(T_{\mathbf{u}} Z = 0) = 1, \text{ then } \mathbb{P}(T_{\mathbf{u}} Z = 0 | Z_{\mathbf{X}_n} = \mathbf{z}_n) = 1 \quad (\forall \mathbf{z}_n \in \mathbb{R}^n)$$

Practical consequence: if one takes a “sparse” kernel without **u** component from the beginning, the corresponding ANOVA terms and Sobol’ indices will remain null **forever**, whatever the strength of the **u** component in the data.

Extraction of subkernels

In the following applications, we will extract subkernels corresponding to prescribed sparse structures by appealing to operators of the form

a) $(I_{\mathcal{F}} - \pi_{\mathcal{Q}}) \otimes (I_{\mathcal{F}} - \pi_{\mathcal{Q}})$ where $\pi_{\mathcal{Q}}$ is the orthogonal projector onto $\mathcal{Q} \subset \mathcal{F}$ for removing the $\mathcal{Q}$-component of Z's trajectories

Extraction of subkernels

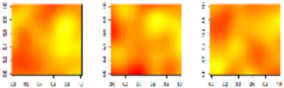
In the following applications, we will extract subkernels corresponding to prescribed sparse structures by appealing to operators of the form

a) $(I_{\mathcal{F}} - \pi_{\mathcal{Q}}) \otimes (I_{\mathcal{F}} - \pi_{\mathcal{Q}})$ where $\pi_{\mathcal{Q}}$ is the orthogonal projector onto $\mathcal{Q} \subset \mathcal{F}$ for removing the $\mathcal{Q}$-component of Z's trajectories

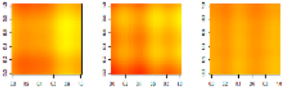
b) $I_{\mathcal{F} \times \mathcal{F}} - \sum_{\mathbf{u},\mathbf{v} \in \mathbf{W} : \mathbf{u} \neq \mathbf{v}} T_{\mathbf{u}} \otimes T_{\mathbf{v}}$ where $\mathbf{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_q\} \subset \mathcal{P}(\{1, \ldots, n\})$, for removing all cross-covariances between contributions among $\mathbf{W}$.

Additive and ortho-additive projections of a GRF

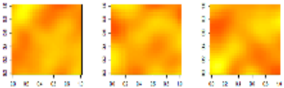
Realizations of a centred GRF Z with isotropic kernel $k(\mathbf{x}, \mathbf{x}') = \sigma^2 \cdot e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{\theta^2}}$



Corresponding realizations of $\pi_{\mathcal{A}} Z$, where $\pi_{\mathcal{A}} := T_{\{\emptyset\}} + \sum_{i=1}^{d} T_{\{i\}}$



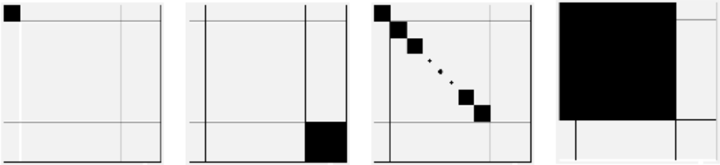
Corresponding realizations of $\pi_{\mathcal{O}} Z$, where $\pi_{\mathcal{O}} := I - \pi_{\mathcal{A}}$



Schematic representation of (sub)kernels

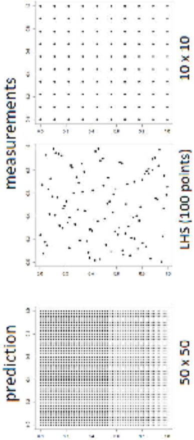
Applying $\mathcal{P} = \{T_{\{\emptyset\}}, T_{\{1\}}, \dots, T_{\{d\}}, \pi_{\mathcal{O}}\}$ to a kernel gives us a decomposition into $(d+2)^2$ parts.

We identify a projected kernel figuratively by a $(d+2) \times (d+2)$ matrix:



constant | ortho-add. | additive | (full) add.

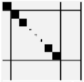
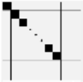
Numerical Experiment



- Define learning and test set on a domain $D = [0, 1]^2$
- Generate $Z := Z(\omega)$ using the sub(kernels):

- Calculate the predictor $\hat{Z} := \hat{Z}(\omega)$ for every trajectory with all (sub)kernels
- Estimate $\int_D \left(\hat{Z}(x) - Z(x)\right)^2 d\mu$
- Replicate the experiment 200 times and rank (sub)kernels according to average prediction error

Results (kernels in line, trajectories in columns)

				
	(0.086) 1	(0.033) 3	(0.055) 3	(0.088) 2
	(0.456) 3	(0.032) 1	(0.429) 4	(0.457) 3
	(6.472) 4	(5.927) 4	(0.043) 1	(5.579) 4
	(0.087) 2	(0.032) 2	(0.055) 2	(0.087) 1

Let $k_{\mathcal{G}}(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{||\mathbf{x}-\mathbf{y}||^2_{\mathbb{R}^d}}{\theta^2}\right)$ with $\theta = 0.5$ and let $k_{anova}(\mathbf{x}, \mathbf{y}) = \prod_{j=1}^{d}(k_1 + k_0(x_j, y_j))$ be the associated ANOVA kernel.

We consider hereafter the following sparse kernels built upon k_{anova}:

$$k_{spa}(\mathbf{x}, \mathbf{y}) := \sigma^2_{spa}\Big(k_1^d + k_1^{d-1}(k_0(x_1, y_1) + k_0(x_2, y_2)) + k_1^{d-2}(k_0(x_2, y_2)k_0(x_3, y_3) + k_0(x_4, y_4)k_0(x_5, y_5))\Big)$$

$$k_{add}(\mathbf{x}, \mathbf{y}) := \sigma^2_{add}\left(k_1^d + \sum_{j=1}^{d} k_1^{d-1}k_0(x_j, y_j)\right)$$

$$k_{inter}(\mathbf{x}, \mathbf{y}) = \sigma^2_{inter}\left(k_1^d + k_1^{d-1}\sum_{j=1}^{d} k_0(x_j, y_j) + \sum_{i<j} k_1^{d-2}k_0(x_i, y_i)k_0(x_j, y_j)\right).$$

Note: The σ^2 factors are chosen so as to ensure that $\int_D k_{\cdot}(\mathbf{x}, \mathbf{x})d\mathbf{x} = 1$.

Selected numerical experiment ($d = 50$)

Experimental design

Training set: A $10d$-point Latin hypercube design

Test set: 200 points uniformly distributed over $[0, 1]^d$.

Fitness: **Q2** $\left(1 - \frac{\sum_{i=1}^{200}(z_{\text{test},i} - \hat{z}(\mathbf{x}_{\text{test},i}))^2}{\sum_{i=1}^{200}(z_{\text{test},i} - \bar{z}_{\text{test}})^2}\right)$, averaged over 100 replications.

Results

	k_{spa}	k_{add}	k_{inter}	k_{anova}	$k_{\mathcal{G}}$
Z_{spa}	1 (0)	0.77 (0.31)	0.8 (0.19)	0.59 (0.19)	0.53 (0.17)
Z_{add}	-0.06 (0.13)	1 (0)	0.93 (0.01)	0.7 (0.04)	0.63 (0.05)
Z_{inter}	-0.03 (0.13)	0.19 (0.1)	0.47 (0.07)	0.27 (0.09)	0.25 (0.1)
Z_{anova}	0.29 (0.27)	0.25 (0.23)	0.31 (0.26)	0.34 (0.26)	0.34 (0.26)
$Z_{\mathcal{G}}$	-0.03 (0.1)	-0.03 (0.11)	-0.01 (0.1)	0.02 (0.1)	0.03 (0.1)

Conclusions and perspectives

1. The kernel FANOVA decomposition explains the distribution of the whole vector-valued field of projected fields
2. Almost sure sparsity properties of GRF paths are characterized by k. Such sparsity properties are "stable" by conditioning.

Conclusions and perspectives

1. The kernel FANOVA decomposition explains the distribution of the whole vector-valued field of projected fields
2. Almost sure sparsity properties of GRF paths are characterized by k. Such sparsity properties are "stable" by conditioning.
3. Some results have been obtained towards posterior distributions of Sobol' indices. Calculating their moments is ongoing research.

Conclusions and perspectives

1. The kernel FANOVA decomposition explains the distribution of the whole vector-valued field of projected fields

2. Almost sure sparsity properties of GRF paths are characterized by k. Such sparsity properties are "stable" by conditioning.

3. Some results have been obtained towards posterior distributions of Sobol' indices. Calculating their moments is ongoing research.

4. ANOVA-like kernels and associated extracted ("sparsified") kernels seem to provide a flexible family for high-dimensional problems.

Conclusions and perspectives

1. The kernel FANOVA decomposition explains the distribution of the whole vector-valued field of projected fields
2. Almost sure sparsity properties of GRF paths are characterized by k. Such sparsity properties are "stable" by conditioning.
3. Some results have been obtained towards posterior distributions of Sobol' indices. Calculating their moments is ongoing research.
4. ANOVA-like kernels and associated extracted ("sparsified") kernels seem to provide a flexible family for high-dimensional problems.

Thank you :-)
Questions?