HAL
open science

Kernels and designs for modelling invariant functions: From group invariance to additivity

David Ginsbourger, Nicolas Durrande, Olivier Roustant

To cite this version:

David Ginsbourger, Nicolas Durrande, Olivier Roustant. Kernels and designs for modelling invariant functions: From group invariance to additivity. MODA-10, Physica-Verlag HD, 2013. hal-00731657

HAL Id: hal-00731657

https://hal.science/hal-00731657v1

Submitted on 13 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Kernels and designs for modelling invariant functions: From group invariance to additivity

David Ginsbourger, Nicolas Durrande, and Olivier Roustant

Abstract We focus on kernels incorporating different kinds of prior knowledge on functions to be approximated by Kriging. A recent result on random fields with paths invariant under a group action is generalised to combinations of composition operators, and a characterisation of kernels leading to random fields with additive paths is obtained as a corollary. A discussion follows on some implications on design of experiments, and it is shown in the case of additive kernels that the so-called class of "axis designs" outperfoms latin hypercubes in terms of the IMSE criterion.

1 Introduction

Models based on Random Fields (RFs), and especially on Gaussian RFs, have been increasingly used in the last decades for designing and analysing costly deterministic experiments [9, 7]. In most popular implementations of such models, a constant or linear trend and a stationary covariance kernel are assumed. However, there seems to be an enormous potential in designing kernels reflecting different kinds of prior knowledge. Recently, a class of kernels leading to RFs with additive paths have been considered in [2] and [3]. Calling $f \in \mathbb{R}^D$ ($D = \prod_i^d D_i$ where $D_i \subset \mathbb{R}$ $(1 \leq i \leq d)$) additive when there exists $f_i \in \mathbb{R}^{D_i}$ $(1 \leq i \leq d)$ such that $\forall \mathbf{x} = (x_1, \ldots, x_d) \in D,\ f(\mathbf{x}) = \sum_{i=1}^d f_i(x_i)$, it was shown in [2] that

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

Nicolas Durrande
Department of Computer Science, The University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP, UK, e-mail: n.durrande(at)sheffield.ac.uk

Olivier Roustant
Ecole Nationale Supérieure des Mines de Saint-Etienne, FAYOL-EMSE, LSTI, F-42023 Saint-Etienne, France, e-mail: roustant@emse.fr

Proposition 1. *If a centered RF Z possesses a kernel of the form*

$$k(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{d} k_i(x_i,x_i') \tag{1}$$

where the k_i's are arbitrary positive definite kernels over the D_i's, then Z is additive up to a modification, i.e. there exists a random field A which paths are additive functions such that $\forall \mathbf{x} \in D, \mathbb{P}(Z_{\mathbf{x}} = A_{\mathbf{x}}) = 1$.

Are the kernels of the form $k(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{d} k_i(x_i,x_i')$ the only ones giving birth to RFs with additive paths? For a different question, [4] proposed a characterization of kernels which associated centered RFs have their trajectories invariant under the action of a finite group. Let G be a finite group acting on D via

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

Proposition 2. *Z has invariant paths under Φ (up to a modification)* ***if and only if*** *k is argumentwise invariant: $\forall \mathbf{x} \in D, \forall g \in G, k(g.\mathbf{x},.) = k(\mathbf{x},.)$.*

We show in Section 2 that both Propositions 1 and 2 are subcases of a general result on RFs invariant under the class of *combination of composition operators*, defined above. As a corollary, a characterization of kernels leading to RFs have additive paths is given in Section 3, and it is shown that having the form of Eq. 1 is not necessary. Sections 4 and 5 are dedicated to a discussion on designs of experiments for RF models with invariant kernels, with examples in the additive case.

2 Invariances and combinations of composition operators

2.1 Composition operators and their combinations

Definition 1. Let us consider an arbitrary function $v : \mathbf{x} \in D \longrightarrow v(\mathbf{x}) \in D$. The *composition operator T_v with symbol v* is defined as follows:

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

Remark 1. Such operators can be naturally extended to random fields indexed by D:

$$\forall \mathbf{x} \in D,\ T_v(Z)_{\mathbf{x}} := Z_{v(\mathbf{x})}$$

Definition 2. We call *combination of composition operators with symbols $v_i \in D^D$ and weights $\alpha_i \in \mathbb{R}$ $(1 \leq i \leq q)$* the operator

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

2.2 Invariance under a combination of composition operators

Proposition 3. *Let Z be a centered RF with kernel k. Then k is T-invariant, i.e.*

$$\forall \mathbf{x}' \in D,\ T(k(.,\mathbf{x}')) = k(.,\mathbf{x}') \tag{2}$$

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

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

Proof. $\Leftarrow$: let us fix arbitrary $\mathbf{x}$ and $\mathbf{x}'$. Since $Z_{\mathbf{x}}$ is a modification of $T(Z)_{\mathbf{x}}$, we have $\mathrm{cov}(Z_{\mathbf{x}}, Z_{\mathbf{x}'}) = \mathrm{cov}(T(Z)_{\mathbf{x}}, Z_{\mathbf{x}'}) = \mathrm{cov}(\sum_{i=1}^{q} \alpha_i Z_{v_i(\mathbf{x})}, Z_{\mathbf{x}'})$, so:

$$k(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{q} \alpha_i k(v_i(\mathbf{x}),\mathbf{x}') = T(k(.,\mathbf{x}'))(\mathbf{x})$$

$\Rightarrow$: Using $\forall \mathbf{x}' \in D\ T(k(.,\mathbf{x}')) = k(.,\mathbf{x}')$, $\mathrm{var}(T(Z)_{\mathbf{x}}) = \mathrm{cov}(Z_{\mathbf{x}}, T(Z)_{\mathbf{x}}) = \mathrm{var}(Z_{\mathbf{x}})$, so $\mathrm{var}(Z_{\mathbf{x}} - T(Z)_{\mathbf{x}}) = 0$. Since Z is centered, so is $T(Z)$, and hence $Z_{\mathbf{x}} \overset{a.s.}{=} T(Z)_{\mathbf{x}}$.

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

3 Kernels characterizing centered fields with additive paths

3.1 Additivity as invariance under a combination of compositions

Proposition 4. *Assuming $\mathbf{a} \in D$, a function $f : D \to \mathbb{R}$ is additive* ***if and only if*** *f is invariant under the following combination of composition operators:*

$$\forall \mathbf{x} \in D,\ T(f)(\mathbf{x}) = \sum_{i=1}^{d} f(v_i(\mathbf{x})) - (d-1) f(v_{d+1}(\mathbf{x})) \tag{3}$$

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

Proof. $\Leftarrow$: from $T(f) = f$, $f(\mathbf{x}) = \sum_{j=1}^{d} f_j(x_j)$ with $f_j(x_j) := f(v_j(\mathbf{x})) - \frac{d-1}{d} f(\mathbf{a})$.

$\Rightarrow$: $f(\mathbf{x}) = \sum_{j=1}^{d} f_j(x_j)$ implies $f(v_i(\mathbf{x})) = f_i(x_i) + \sum_{j=1, j\neq i}^{d} f_j(a_j)$, and so

$$T(f)(\mathbf{x}) = \sum_{i=1}^{d} f(v_i(\mathbf{x})) - (d-1)f(\mathbf{a})$$
$$= \sum_{i=1}^{d} f_i(x_i) + \underbrace{\sum_{i=1}^{d} \sum_{j=1, j\neq i}^{d} f_j(a_j) - (d-1)f(\mathbf{a})}_{0} = f(\mathbf{x})$$

3.2 Kernels characterizing centered fields with additive paths

Corollary 1. *A centered RF Z possessing a covariance kernel k has additive paths (up to a modification) if and only if k is a positive definite kernel of the form*

$$\boxed{k(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{d} \sum_{j=1}^{d} k_{ij}(x_i, x'_j)} \tag{4}$$

Proof. If Z has additive paths up to a modification, there exists a RF $(A_{\mathbf{x}})_{\mathbf{x}\in D}$ with additive paths such that $\forall \mathbf{x} \in D$ $\mathbb{P}(Z_{\mathbf{x}} = A_{\mathbf{x}}) = 1$, and so Z and A have the same covariance kernel. Now, A having additive paths, Proposition 4 implies that $A_{\mathbf{x}} = \sum_{i=1}^{d} A_{v_i(\mathbf{x})} - (d-1)A_{v_{d+1}(\mathbf{x})} = \sum_{i=1}^{d} A^i_{x_i}$, where $A^i_{x_i} := A_{v_i(\mathbf{x})} - \frac{(d-1)}{d} A_{v_{d+1}(\mathbf{x})}$, so Equation 4 holds with $k_{ij}(x_i, x'_j) := \mathrm{cov}(A^i_{x_i}, A^j_{x'_j})$. Reciprocally, from Proposition 3, we know that it sufficies for Z to have additive paths that $k(.,\mathbf{x}')$ is additive $\forall \mathbf{x}' \in D$. For a kernel k such as in Eq. 4 and an arbitrary $\mathbf{x}' \in D$, setting

$$\forall x_i \in D_i,\ \widetilde{k}_i(x_i, \mathbf{x}') := \sum_{j=1}^{d} k_{ij}(x_i, x'_j) \quad (1 \leq i \leq d)$$

we get $\forall \mathbf{x} \in D,\ k(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{d} \widetilde{k}_i(x_i, \mathbf{x}')$, and so $k(.,\mathbf{x}')$ is additive.

4 Kriging-equivalent designs: Generalities and invariant case

We now focus on cases where two designs $X \in D^n, X' \in D^{n'}$ bring the same information on Z. We first give general results, and then specify to invariant RFs.

Definition 3. X' and X are said *Kriging-equivalent*, denoted by $X' \equiv X$, iif the Kriging mean and variance of Z based on $\{Z_{\mathbf{x}}, \mathbf{x} \in X\}$ or $\{Z_{\mathbf{x}'}, \mathbf{x}' \in X'\}$ coincide.

In particular two equivalent designs lead to the same Integrated Mean Squared Error (IMSE). We now give a sufficient condition for two designs to be equivalent.

Proposition 5. *If $span(Z_{\mathbf{x}'}, \mathbf{x}' \in X') = span(Z_{\mathbf{x}}, \mathbf{x} \in X)$, then $X' \equiv X$.*

Proof. Kriging is characterized by the *linear* conditional expectations $E_L(Z_{\mathbf{u}}|Z_{\mathbf{x}}, \mathbf{x} \in X)$ and $E_L(Z_{\mathbf{u}}Z_{\mathbf{v}}|Z_{\mathbf{x}}, \mathbf{x} \in X)$ ($\mathbf{u}, \mathbf{v}$ in D), and hence depends only on $\text{span}(Z_{\mathbf{x}}, \mathbf{x} \in X)$.

Remark 2. If Z is also Gaussian, the equality of the two linear spans guarantees that the whole conditional processes $Z|Z_{\mathbf{x}}, \mathbf{x} \in X$ and $Z|Z_{\mathbf{x}'}, \mathbf{x}' \in X'$ have the same distribution. In particular conditional simulations performed with X or X' coincide.

Corollary 2 (An exchangeability condition). *Assume that* $\exists \mathbf{x}' \notin X, Z_{\mathbf{x}'} = \sum_{\mathbf{x} \in X} \alpha_{\mathbf{x}} Z_{\mathbf{x}}$, *with* $\alpha_{\mathbf{x}} \neq 0, \forall \mathbf{x} \in X$. *Then* $\forall \mathbf{x} \in X, X - \{\mathbf{x}\} + \{\mathbf{x}'\} \equiv X \equiv X + \{\mathbf{x}'\}$.

Remark 3. The condition "$\alpha_{\mathbf{x}} \neq 0, \forall \mathbf{x} \in X$" is one way to guarantee that the dimension of $\text{span}(Z_{\mathbf{x}}, \mathbf{x} \in X)$ does not decrease when exchanging one $Z_{\mathbf{x}}$ by $Z_{\mathbf{x}'}$.

Corollary 3 (Invariant kernels). *Let us consider a combination of composition operators* $T = \sum_{j=1}^{q} \alpha_j T_{v_j}$, *with* $\forall j = 1, \ldots, q, \alpha_j \neq 0$, *and assume that k is T-invariant. For* $\mathbf{x} \in D$, *denote* $X_v(\mathbf{x}) := (v_1(\mathbf{x}), \ldots, v_q(\mathbf{x}))$. *Then for* $j = 1, \ldots, q$, *we have:*

$$X_v(\mathbf{x}) \equiv X_v(\mathbf{x}) - v_j(\mathbf{x}) + \{\mathbf{x}\} \equiv X_v(\mathbf{x}) + \{\mathbf{x}\}$$

Proof. This results from the fact that $T(Z) = Z$ (Proposition 3) and Corollary 2.

Example 2 (Equivalent designs for additive kernels). Let us assume that Z is 2-dimensional, with an additive kernel k, and consider a *rectangle* design:

$$X = \{\mathbf{x}^{(1)} = (a_1, a_2), \mathbf{x}^{(2)} = (b_1, a_2), \mathbf{x}^{(3)} = (a_1, b_2), \mathbf{x}^{(4)} = (b_1, b_2)\}$$

with $\mathbf{a}, \mathbf{b} \in D, \mathbf{a} \neq \mathbf{b}$. Then, all 3-point designs contained in X are equivalent:

$$(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}) \equiv (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(4)}) \equiv (\mathbf{x}^{(1)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}) \equiv (\mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}) \equiv X$$

Indeed, consider the operator T defined by $T(f)(\mathbf{x}) = f(a_1, x_2) + f(x_1, a_2) - f(a_1, a_2)$, as in Proposition 4 ($d = 2$). Relying on this proposition, k is T-invariant. The result follows by applying Corollary 3 with $q = 3$ ($\alpha_1 = \alpha_2 = 1, \alpha_3 = -1$) and $\mathbf{x} = (b_1, b_2)$.

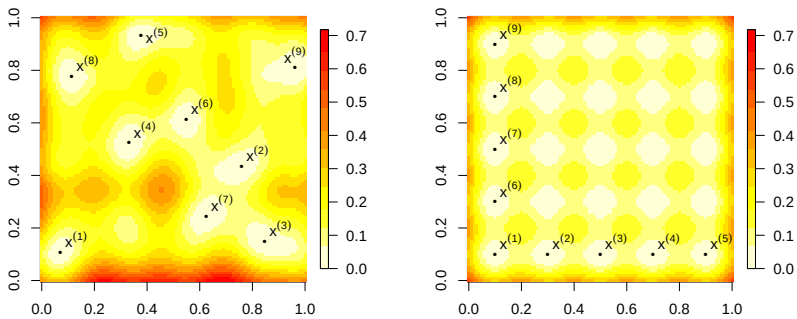
Example 3 (Equivalent designs for a group-invariant kernel). The case where Z is invariant under the action of a group G is degenerate since the condition $T(Z)_{\mathbf{x}} = Z_{\mathbf{x}}$ is equivalent to: $\forall g \in G, Z_{\mathbf{x}} = Z_{g.\mathbf{x}}$. Then it comes directly from Definition 3 that replacing one point in X by any other point of its orbit gives an equivalent design:

$$\forall g_1, \ldots, g_n \in G, \quad (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) \equiv (g_1.\mathbf{x}^{(1)}, \ldots, g_n.\mathbf{x}^{(n)})$$

5 On choosing designs for RF models with an additive kernel

Example 2 shows that additive kernels can lead to points with zero variance outside of the design. For LHS designs, such configuration cannot occur since points of the design are never aligned vertically nor horizontally. On the other hand, designs

where points are distributed parallel to the axis with a shared point at the intersection, hereafter called "axis designs", take advantage of this property since they imply zero variance on a whole grid. This property is illustrated on Figure 1.

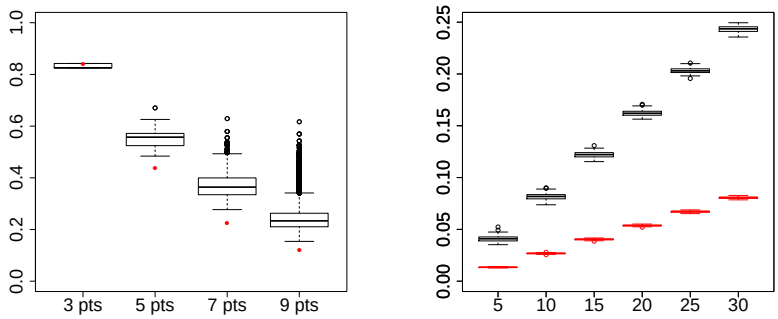


(a) 9 points LHS design (b) 9 points axis design

Fig. 1 MSE for two 9-point designs. The univariate kernels are Matérn 5/2 with parameters $\sigma^2 = 1$, $\theta = 0.2$. For the (scrambled) LHS, the IMSE is $I = 0.196$ wheras for the axis design $I = 0.116$.

Figure 2 compares, for different values of the number of points n, the IMSE of all possible LHS compared to the one of an axis design. In all cases, the space is divided in an appropriate number of square cells (n^2 for the LHS and $(\frac{n-1}{2}+1)^2$ for the axis design) and the design points are located in the center of some cells. It appears that, except for $n = 3$, the tested configurations are always in favour of the axis design. As the total number of possible LHS is $n!$ for $d = 2$, it was practically unfeasible to run the exhaustive comparison for more than 9 points. However, the right panel of Figure 2 shows the comparison between 100 maximin LHS generated with the *lhs* R package [1] and the axis designs for various values of the dimension d. As advocated in [6], the number of points is taken equal to 10 times the dimension: $n = 10 \times d + 1$. This graph suggests that the axis designs are more and more superior to LHS when the dimension increases. These numerical investigations show that axis designs seem particularly adapted for fitting additive Kriging models.

However, axis designs are likely to perform poorly for non additive functions since they do not fill the space. Fortunately, a direct application of Corollary 2 shows that any design point can be moved to any other point where it induces a zero variance without introducing any change in the resulting Kriging equations. A straightforward application of this property is that the points distributed over one axis can be scattered in a LHS fashion without modifying the IMSE. This approach is illustrated on Figure 3. If the function to approximate has an additive component but also some interaction terms, the design presented on the right panel may allow to capture efficiently not only the additive component but also the interaction terms.



(a) Influence of n for $d = 2$

(b) Influence of d with $n = 10 \times d + 1$

Fig. 2 Comparisons of the IMSE of LH designs (black) and axis designs (red). On the left panel, all possible LHS are enumerated and the integral in the IMSE expression is approximated by a Riemann sum based on 51^2 points. On the right panel, 100 maximin-LHS are compared with the axis design. The variability in the IMSE of axis designs is due to the use of Monte Carlo methods for integration. In both cases, the settings of the covariance are the same as in Figure 1.

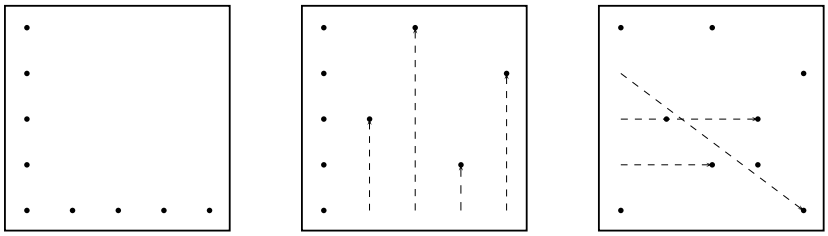


Fig. 3 Example of Kriging-equivalent modifications of a design in the additive case. The axis design is transformed into the sum of a 5-point LHS and another 5-point LHS deprived of one point. These three designs lead to identical Kriging models when using a kernel satisfying Eq. 4.

6 Concluding remarks and perspectives

As (generalized) additive models [5] and related sparsity assumptions for high-dimensional data seem to have reached a golden age, designing adapted kernels for high-dimensional Kriging is still at its beginning. Following the recent article [2] proposing a particular kind of kernels leading to centered RFs with additive paths, we give here a complete characterization of such kernels. This characterization appears in fact as a particular case of a property involving so-called combinations of composition operators. This property also generalizes another recent characterization of covariance kernels leading to RFs with paths invariant under the action of a finite group on the index set [4].

Some implications concerning the design of experiments are discussed, and it is illustrated with an empirical study that the so-called class of *axis designs* outperforms latin hypercubes in terms of IMSE for most configurations (for $n > 3$) in the case of an additive kernel. Furthermore, thanks to a proven exchangeability property in the case of Kriging modelling with an invariant kernel, such designs can be modified so as to lead to better performances when interactions exist while preserving exactly the same features as the axis design in the additive case.

Future work include an extended study of optimal designs for Kriging with an invariant kernel. In addition, further generalisations of the present result on combination of composition operators may be worth looking at for understanding better what kind of prior knowledge can (or can't) be injected in a RF model (in the Gaussian or in the general case), and how to practically implement kernels [8] incorporating given functional properties, with adapted parameter estimation procedures.

References

1. Carnell, R.: lhs: Latin Hypercube Samples (2009). R package version 0.5
2. Durrande, N., Ginsbourger, D., Roustant, O.: Additive covariance kernels for high-dimensional gaussian process modeling. Annales de la Faculte des Sciences de Toulouse (accepted) (2012). URL http://hal.inria.fr/hal-00446520/en
3. Duvenaud, D., Nickisch, H., Rasmussen, C.: Additive Gaussian Processes (2011). NIPS conference
4. Ginsbourger, D., Bay, X., Roustant, O., Carraro, L.: Argumentwise invariant kernels for the approximation of invariant functions. Annales de la Faculte des Sciences de Toulouse (in press) **21 (3)**, 501–527 (2012). URL http://hal.archives-ouvertes.fr/hal-00632815/
5. Hastie, T., Tibshirani, R.: Generalized additive models. Chapman & Hall/CRC (1990)
6. Loeppky, J., Sacks, J., Welch, W.: Choosing the sample size of a computer experiment: A practical guide. Technometrics **51**(4), 366–376 (2009)
7. Rasmussen, C.R., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press (2006)
8. Roustant, O., Ginsbourger, D., Deville, Y.: DiceKriging: Kriging methods for computer experiments (2012). URL http://CRAN.R-project.org/package=DiceKriging. R package version 1.4
9. Santner, T., Williams, B., Notz, W.: The design and analysis of computer experiments. Springer Verlag (2003)