



**HAL**  
open science

# Optimal Design for Prediction in Random Field Models via Covariance Kernel Expansions

Bertrand Gauthier, Luc Pronzato

► **To cite this version:**

Bertrand Gauthier, Luc Pronzato. Optimal Design for Prediction in Random Field Models via Covariance Kernel Expansions. MODA 11, J. Kunert, Ch.H. Müller, Jun 2016, Hamminkeln-Dingden, Germany. hal-01280355

**HAL Id: hal-01280355**

**<https://hal.science/hal-01280355>**

Submitted on 29 Feb 2016

**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.

# Optimal Design for Prediction in Random Field Models via Covariance Kernel Expansions

Bertrand Gauthier and Luc Pronzato

**Abstract** We consider experimental design for the prediction of a realization of a second-order random field  $Z$  with known covariance function, or kernel,  $K$ . When the mean of  $Z$  is known, the integrated mean squared error of the best linear predictor, approximated by spectral truncation, coincides with that obtained with a Bayesian linear model. The machinery of approximate design theory is then available to determine optimal design measures, from which exact designs (collections of sites where to observe  $Z$ ) can be extracted. The situation is more complex in the presence of an unknown linear parametric trend, and we show how a Bayesian linear model especially adapted to the trend can be obtained via a suitable projection of  $Z$  which yields a reduction of  $K$ .

## 1 Introduction

We consider a centered second-order random field  $(Z_x)_{x \in \mathcal{X}}$  indexed over  $\mathcal{X}$ , a compact subset of  $\mathbb{R}^d$ ,  $d \geq 1$ , with covariance function  $E\{Z_x Z_{x'}\} = K(x, x') = K(x', x)$ . We also consider a  $\sigma$ -finite measure  $\mu$  on  $\mathcal{X}$  which will be used to define the Integrated Mean Squared Error (IMSE) for prediction, see (1). We assume that  $\mu(\mathcal{X}) = 1$  without any loss of generality. We suppose that  $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is continuous and thus belongs to  $L^2_{\mu \times \mu}(\mathcal{X} \times \mathcal{X})$ . This framework is now classical for computer experiments, where  $(Z_x)_{x \in \mathcal{X}}$  represents the output of a computer code with input variables  $x \in \mathcal{X}$ , in the sense that it forms a prior on the possible behavior of the code when  $x$  vary in  $\mathcal{X}$ , see in particular [10]. The design problem

---

B. Gauthier

KU Leuven, ESAT-STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Belgium, e-mail: bgauthie@esat.kuleuven.be

L. Pronzato

CNRS, Laboratoire I3S – UMR 7271 Université de Nice-Sophia Antipolis/CNRS, France, e-mail: pronzato@cnrs.fr

then consists in choosing  $n$  sites  $x_1, \dots, x_n$  where to observe  $Z_x$  *without error* (i.e.,  $n$  inputs values for the computer code) in order to predict at best the realization of  $Z_x$  over  $\mathcal{X}$ . Note that the objective concerns prediction of *one instance* of the random field, based on a finite set of error-free observations.

The best linear predictor of  $Z_x$ , unbiased and in terms of mean squared error (the simple kriging predictor), based on the vector of observations  $\mathbf{z}_n = (Z_{x_1}, \dots, Z_{x_n})^\top$  obtained with the design  $\mathcal{D}_n = (x_1, \dots, x_n)$ , is given by  $\mathbf{k}^\top(x)\mathbf{K}^{-1}\mathbf{z}_n$ , where  $\mathbf{k}(x) = (K(x_1, x), \dots, K(x_n, x))^\top$  and the matrix  $\mathbf{K}$  has elements  $\{\mathbf{K}\}_{ij} = K(x_i, x_j)$ . Its Mean Squared Error (MSE) is  $\text{MSE}(x; \mathcal{D}_n) = K(x, x) - \mathbf{k}^\top(x)\mathbf{K}^{-1}\mathbf{k}(x)$ .

The IMSE will be computed for the measure  $\mu$ , which weighs the importance given to the precision of predictions in different parts of  $\mathcal{X}$ ,

$$\text{IMSE}(\mathcal{D}_n) = \int_{\mathcal{X}} \text{MSE}(x; \mathcal{D}_n) d\mu(x). \quad (1)$$

Choosing a design  $\mathcal{D}_n$  with minimum IMSE is then a rather natural objective; however, IMSE-optimal designs are considered as difficult to compute, see [10]. The introduction of the following integral operator on  $L^2_\mu(\mathcal{X})$  yields a spectral representation of the IMSE and is at the core of the design approach presented in this paper.

For all  $f \in L^2_\mu(\mathcal{X})$  and  $x \in \mathcal{X}$ , we define

$$T_\mu[f](x) = \int_{\mathcal{X}} f(t)K(x, t) d\mu(t).$$

From Mercer's theorem, there exists an orthonormal set  $\{\tilde{\phi}_k, k \in \mathbb{N}\}$  of eigenfunctions of  $T_\mu$  in  $L^2_\mu(\mathcal{X})$  associated with nonnegative eigenvalues, the eigenfunctions corresponding to non-zero eigenvalues being continuous on  $\mathcal{S}_\mu$ , the support of  $\mu$ . Moreover, for  $x, x' \in \mathcal{S}_\mu$ ,  $K$  has the representation  $K(x, x') = \sum_{k=1}^{\infty} \lambda_k \tilde{\phi}_k(x) \tilde{\phi}_k(x')$ , where the  $\lambda_k$  denote positive eigenvalues that we assume to be sorted by decreasing values, and the convergence is uniform on compact subsets of  $\mathcal{S}_\mu$ . As shown below, see (2), the integration over  $\mathcal{X}$  in (1) can then be replaced by a summation over  $k$ , the index of eigenvalues. There is a slight technicality here, due to the fact that  $\mathcal{S}_\mu$  may be strictly included in  $\mathcal{X}$ , so that the eigenfunctions  $\tilde{\phi}_k$ , which are only defined  $\mu$ -almost everywhere, may not be defined at a general design point  $x_i \in \mathcal{X}$ . This is the case for instance when  $\mu$  is a discrete measure supported at a finite set of points  $\mathcal{X}_Q \subset \mathcal{X}$  (thus in particular when the IMSE is approximated via a quadrature rule) and  $\mathcal{D}_n$  is not a subset of  $\mathcal{X}_Q$ . A general way to avoid this difficulty is via the introduction of the canonical extensions

$$\phi_k = (1/\lambda_k)T_\mu[\tilde{\phi}_k], \quad \lambda_k > 0,$$

which are defined on the whole set  $\mathcal{X}$  and satisfy  $\phi_k(x) = \tilde{\phi}_k(x)$  on  $\mathcal{S}_\mu$ . Note that the  $\phi_k$  are continuous since we assumed that  $K(\cdot, \cdot)$  is continuous. One may refer to [4] for more details. Then,  $K(x, x') = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \phi_k(x')$  for all  $x, x'$  in  $\mathcal{X}$  and

$$\text{IMSE}(\mathcal{D}_n) = \tau - \sum_{k=1}^{\infty} \lambda_k^2 \underline{\phi}_k^\top \mathbf{K}^{-1} \underline{\phi}_k, \quad (2)$$

where  $\tau = \sum_{k=1}^{\infty} \lambda_k$  and  $\underline{\phi}_k = (\phi_k(x_1), \dots, \phi_k(x_n))^\top$ . A spectral truncation that uses only the  $n_{trc}$  largest eigenvalues yields an approximation of  $\text{IMSE}(\mathcal{D}_n)$ ,

$$\text{IMSE}_{trc}(\mathcal{D}_n) = \tau_{trc} - \sum_{k=1}^{n_{trc}} \lambda_k^2 \underline{\phi}_k^\top \mathbf{K}^{-1} \underline{\phi}_k = \text{trace}(\Lambda_{trc} - \Lambda_{trc} \Phi_{trc}^\top \mathbf{K}^{-1} \Phi_{trc} \Lambda_{trc}), \quad (3)$$

with  $\tau_{trc} = \sum_{k=1}^{n_{trc}} \lambda_k$ ,  $\Phi_{trc}$  the  $n \times n_{trc}$  matrix with elements  $\{\Phi_{trc}\}_{ik} = \phi_k(x_i)$  and  $\Lambda_{trc} = \text{diag}\{\lambda_1, \dots, \lambda_{n_{trc}}\}$ . This satisfies  $\text{IMSE}_{trc}(\mathcal{D}_n) \leq \text{IMSE}(\mathcal{D}_n) \leq \text{IMSE}_{trc}(\mathcal{D}_n) + \tau_{err}$ , where  $\tau_{err} = \tau - \tau_{trc} = \sum_{k > n_{trc}} \lambda_k$ , see [4, 6].

Following [2, 13], we recall below (Sect. 2) how, using its Karhunen-Loève decomposition,  $(Z_x)_{x \in \mathcal{X}}$  can be interpreted as a Bayesian Linear Model (BLM) *with correlated errors* having the eigenfunctions  $\phi_k$ ,  $k = 1, \dots, n_{trc}$ , as regressors, and for which the IMSE of prediction coincides with  $\text{IMSE}_{trc}(\mathcal{D}_n)$ . A similar decomposition has been used in [3], but for the estimation of trend parameters (or equivalently, for the prediction of the mean of  $Z_x$ ). The IMSE for a BLM *with uncorrelated errors* takes the form of a Bayesian A-optimality criterion, which opens the ways for the construction of approximate optimal designs (design measures) by convex optimization, see [8]. Exact designs  $\mathcal{D}_n$ , that is, collections of sites where to observe  $Z$ , can be extracted from these optimal measures, see Sect. 3. However, replacing correlated errors by uncorrelated ones, in order to be able to use the approximate design machinery, introduces an additional approximation (besides spectral truncation). Whereas the approximation in [13] relies on an homoscedastic model, the one we propose in Sect. 2 uses a more accurate heteroscedastic model. The situation is more complex when a linear parametric trend is present, and in Sect. 4 we show how a suitable projection of  $(Z_x)_{x \in \mathcal{X}}$  on the linear subspace spanned by the trend, equivalent to a kernel reduction, yields a BLM with smaller errors than the direct approach of [13] that uses the original kernel  $K$ .

## 2 Bayesian Linear Models

*Exact BLM* The Karhunen-Loève decomposition of  $Z_x$  yields

$$Z_x = \sum_{k=1}^{n_{trc}} \beta_k \phi_k(x) + \varepsilon_x, \quad (4)$$

where  $\varepsilon_x$ ,  $x \in \mathcal{X}$ , is a centered random field with covariance given by

$$\text{E}\{\varepsilon_x \varepsilon_{x'}\} = K_{err}(x, x') = K(x, x') - K_{trc}(x, x'),$$

with  $K_{trc}(x, x') = \sum_{k=1}^{n_{trc}} \lambda_k \phi_k(x) \phi_k(x')$  and where the  $\beta_k$  are mutually uncorrelated centered r.v., orthogonal to  $\varepsilon_x$  for all  $x \in \mathcal{X}$ , with  $\text{var}(\beta_k) = \lambda_k$ . No approximation is involved at this stage, and this BLM gives an exact representation of  $(Z_x)_{x \in \mathcal{X}}$ . Consider the predictor  $\widehat{Z}_x = \sum_{k=1}^{n_{trc}} \widehat{\beta}_k \phi_k(x) = \underline{\phi}_{trc}^\top(x) \widehat{\underline{\beta}}$ , with  $\underline{\phi}_{trc} = (\phi_1(x), \dots, \phi_{n_{trc}}(x))^\top$  and  $\widehat{\underline{\beta}} = (\Phi_{trc}^\top \mathbf{K}_{err}^{-1} \Phi_{trc} + \Lambda_{trc}^{-1})^{-1} \Phi_{trc}^\top \mathbf{K}_{err}^{-1} \mathbf{z}_n$  the estimator that minimizes the regularized LS criterion  $J(\underline{\beta}) = (\mathbf{z}_n - \Phi_{trc} \underline{\beta})^\top \mathbf{K}_{err}^{-1} (\mathbf{z}_n - \Phi_{trc} \underline{\beta}) + \underline{\beta}^\top \Lambda_{trc}^{-1} \underline{\beta}$  ( $\widehat{\underline{\beta}}$  coincides with the posterior mean of  $\underline{\beta}$  when  $Z_x$  is Gaussian). Direct calculation shows that the IMSE of  $\widehat{Z}_x$  equals  $\text{IMSE}_{trc}(\mathcal{D}_n)$  given by (3) and can also be written as

$$\text{IMSE}_{trc}(\mathcal{D}_n) = \text{trace}[(\Phi_{trc}^\top \mathbf{K}_{err}^{-1} \Phi_{trc} + \Lambda_{trc}^{-1})^{-1}].$$

*Approximate BLM with uncorrelated errors* In [13], the correlated errors  $\varepsilon_x$  of (4) are replaced by uncorrelated homoscedastic errors having variance  $\sigma^2 = \tau_{err} = \tau - \tau_{trc}$ . A more accurate approximation of the exact model (4) is obtained when using uncorrelated but heteroscedastic errors, with the same variance  $\sigma^2(x) = K_{err}(x, x)$  as  $\varepsilon_x$ . We shall call this model the *heteroscedastic BLM* and denote by  $\Sigma_{err}$  the diagonal matrix  $\text{diag}\{K_{err}(x_1, x_1), \dots, K_{err}(x_n, x_n)\}$ . The curve (solid line) on the top of Fig. 3-Right shows  $\sigma^2(x)$ ,  $x \in [0, 1]$ , for the Matérn 3/2 kernel  $K(x, x') = C_{3/2, 10}(|x - x'|)$ , where

$$C_{3/2, \vartheta}(t) = (\vartheta t + 1) \exp(-\vartheta t), \quad (5)$$

with  $n_{trc} = 10$ . The first 4 eigenfunctions  $\phi_k$  for this kernel are plotted in Fig. 2-Left. The strongly oscillating behavior of  $\sigma^2(x)$ , due to the form of eigenfunctions in  $K_{trc}$ , motivates the use of the heteroscedastic BLM instead of an approximate model with homoscedastic errors. This seems important within the framework of computer experiments, but would be less critical, however, in presence of additive uncorrelated measurement errors, as considered in [13]; in that case, the errors  $\varepsilon_x$  due to spectral truncation are even neglected in [3]. The IMSE for prediction with the heteroscedastic BLM is

$$\text{IMSE}_{trc}^{hBLM}(\mathcal{D}_n) = \text{trace}[(\Phi_{trc}^\top \Sigma_{err}^{-1} \Phi_{trc} + \Lambda_{trc}^{-1})^{-1}]. \quad (6)$$

### 3 Optimal Design

The IMSE (6), considered as a function of the exact design  $\mathcal{D}_n$ , has the form of an A-optimality criterion for the Bayesian information matrix  $\Phi_{trc}^\top \Sigma_{err}^{-1} \Phi_{trc} + \Lambda_{trc}^{-1}$ , see [8]. For  $\xi$  a probability measure on  $\mathcal{X}$  (i.e., a design measure), let

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} \sigma^{-2}(x) \underline{\phi}_{trc}(x) \underline{\phi}_{trc}^\top(x) d\xi(x),$$

which satisfies  $\text{IMSE}_{trc}^{hBLM}(\mathcal{D}_n) = \text{trace}[(n\mathbf{M}(\xi_n) + \Lambda_{trc}^{-1})^{-1}]$ , with  $\xi_n$  the empirical measure  $(1/n) \sum_{i=1}^n \delta_{x_i}$  associated with  $\mathcal{D}_n$ . For any  $\alpha \geq 0$ , the Bayesian A-

optimality criterion

$$\psi_\alpha(\xi) = \text{trace}[(\alpha \mathbf{M}(\xi_n) + \Lambda_{trc}^{-1})^{-1}]$$

is a convex function of  $\xi$  which can easily be minimized using an algorithm for optimal design, see, e.g., [9, Chap. 9]. The solution is much facilitated when (i)  $\mu$  is a discrete measure with finite support  $\mathcal{X}_Q$  (a quadrature approximation is used to compute the IMSE) and (ii) only design measures  $\xi$  dominated by  $\mu$  are considered (design points are chosen among quadrature points in  $\mathcal{X}_Q$ ). Indeed, only the values  $\tilde{\phi}_k(x_j)$  for  $x_j \in \mathcal{X}_Q$  are then used in all calculations, without having to compute the canonical extensions  $\phi_k = (1/\lambda_k)T_\mu[\tilde{\phi}_k]$ , see Sect. 1, and the  $\tilde{\phi}_k(x_j)$  are obtained by the spectral decomposition of a  $Q \times Q$  matrix, with  $Q = |\mathcal{X}_Q|$ , see [4, 5]. Once an optimal design measure  $\xi_\alpha^*$  minimizing  $\psi_\alpha(\cdot)$  has been determined, we still have to extract from it an exact design  $\mathcal{D}_n$  with  $n$  points (and no repetitions since there are no observation errors). We have tried several procedures; all are based on the selection of  $n$  points among the support of  $\xi_\alpha^*$ ,  $\mathcal{S}_{\xi_\alpha^*} = (x^{(1)}, \dots, x^{(m)})$  say, and involve some trial and error for tuning values of  $\alpha$  and  $n_{trc}$  in order to facilitate extraction of an  $n$ -point design when  $n$  is given *a priori*. They are listed below:

1. Keep the  $n$  points  $x^{(i)}$  of  $\mathcal{S}_{\xi_\alpha^*}$  with largest weight  $\xi_\alpha^*(x^{(i)})$ .
2. Perform a greedy (one-point-at-a-time) minimization of the IMSE using the finite set  $\mathcal{S}_{\xi_\alpha^*}$  as design space.
3. Perform a minimum-spanning-tree clustering of  $\mathcal{S}_{\xi_\alpha^*}$ , with the metric  $\Delta(x, x') = K(x, x) + K(x', x') - 2K(x, x')$  induced by  $K$ , with  $n$  clusters (the  $x_i$  then correspond to the barycenters of  $x^{(j)}$  in clusters, with weights the  $\xi_\alpha^*(x^{(j)})$ ).

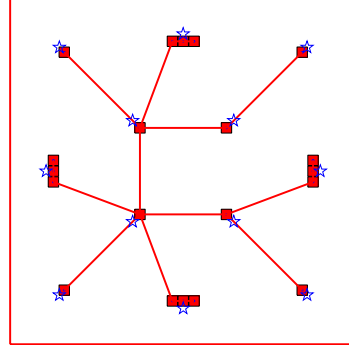
Choosing  $\alpha$  and  $n_{trc}$  of the same order of magnitude as  $n$  seems reasonable. Note that while 1. and 2. maintain the  $x_i$  within  $\mathcal{X}_Q$  when  $\mu$  has finite support  $\mathcal{X}_Q$  and  $\xi_\alpha^*$  is dominated by  $\mu$ , this is not the case for 3. and canonical extensions  $\phi_k = (1/\lambda_k)T_\mu[\tilde{\phi}_k]$  must then be computed (which requires summations over  $\mathcal{X}_Q$ ). After extraction of a  $n$ -point design  $\mathcal{D}_n$ , a local minimization of  $\text{IMSE}(\mathcal{D}_n)$  (a problem in  $\mathbb{R}^{n \times d}$ ) can be performed using any standard algorithm for unconstrained optimization (e.g., conjugate gradient, variable metric), since optimal points for observation of  $Z_x$  need to be as correlated as possible with points within  $\mathcal{X}$ , and therefore lie in the interior of  $\mathcal{X}$ . Here also canonical extensions must be computed.

An illustration of procedure 3. is presented in Fig. 1 for  $\mathcal{X} = [0, 1]^2$  and  $K(x, x') = C_{3/2, 10}(\|x - x'\|)$ , see (5), with  $n_{trc} = \alpha = 10$ ;  $\mu$  is the uniform measure on a regular grid  $\mathcal{X}_Q$  of  $Q = 33 \times 33$  points and a vertex-exchange algorithm with Armijo-type line search [1] is used to determine  $\xi_\alpha^*$  supported on  $\mathcal{X}_Q$ .

## 4 Kernel Reduction for Models with Parametric Trend

Consider now the random field  $Y_x = \mathbf{g}^\top(x)\underline{\theta} + Z_x$ , where  $Z_x$  is as in Sect. 1,  $\mathbf{g}(x) = (g_1(x), \dots, g_p(x))^\top$  is a vector of (known) real-valued trend functions defined on

**Fig. 1** The 20 squares correspond to the support of  $\xi_\alpha^*$  when the vertex-exchange algorithm is stopped (the directional derivative of  $\psi_\alpha(\xi)$  exceeds  $-10^{-4}$ ), the associated minimum spanning-tree identifies  $n = 12$  clusters, the stars indicate the optimal 12-point design obtained by local minimization of the IMSE



$\mathcal{X}$  and where  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$  is an unknown vector of parameters. The Best Linear Unbiased Predictor (BLUP) of  $Y_x$  (the universal kriging predictor), based on observations  $\mathbf{y}_n = (Y_{x_1}, \dots, Y_{x_n})^\top$ , is  $\mathbf{g}^\top(x)\hat{\boldsymbol{\theta}} + \mathbf{k}^\top(x)\mathbf{K}^{-1}(\mathbf{y}_n - \mathbf{G}\hat{\boldsymbol{\theta}})$ , where  $\hat{\boldsymbol{\theta}} = (\mathbf{G}^\top\mathbf{K}^{-1}\mathbf{G})^{-1}\mathbf{G}^\top\mathbf{K}^{-1}\mathbf{y}_n$  with  $\mathbf{G}$  the  $n \times p$  matrix with elements  $\{\mathbf{G}\}_{ij} = g_j(x_i)$  (we assume that  $\mathbf{G}$  has full column rank). Its MSE is

$$\begin{aligned} \text{MSE}(x; \mathcal{D}_n) &= K(x, x) - \mathbf{k}^\top(x)\mathbf{K}^{-1}\mathbf{k}(x) \\ &\quad + [\mathbf{g}(x) - \mathbf{G}^\top\mathbf{K}^{-1}\mathbf{k}(x)]^\top (\mathbf{G}^\top\mathbf{K}^{-1}\mathbf{G})^{-1} [\mathbf{g}(x) - \mathbf{G}^\top\mathbf{K}^{-1}\mathbf{k}(x)], \end{aligned}$$

see, e.g., [11, Chaps. 3,4], and  $\text{IMSE}(\mathcal{D}_n) = \int_{\mathcal{X}} \text{MSE}(x; \mathcal{D}_n) d\mu(x)$ .

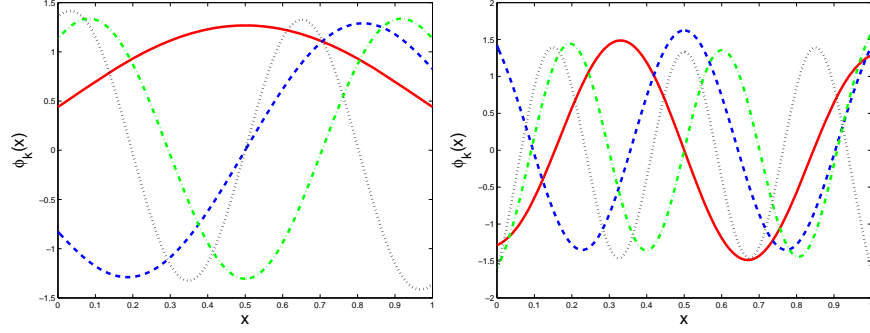
Similarly to Sect. 2, we can consider an exact BLM  $Y_x = \mathbf{f}^\top(x)\boldsymbol{\gamma} + \varepsilon_x$ , where  $\mathbf{f}^\top(x) = (\mathbf{g}^\top(x), \boldsymbol{\phi}_{trc}^\top(x))$  and  $\boldsymbol{\gamma} = (\boldsymbol{\theta}^\top, \boldsymbol{\beta}^\top)^\top$ . The predictor of  $Y_x$  is then  $\hat{Y}_x = \mathbf{f}^\top(x)\hat{\boldsymbol{\gamma}}$ , where  $\hat{\boldsymbol{\gamma}} = (\mathbf{F}^\top\mathbf{K}_{err}^{-1}\mathbf{F} + \Gamma^{-1})^{-1}\mathbf{F}^\top\mathbf{K}_{err}^{-1}\mathbf{y}$ , with  $\mathbf{F} = (\mathbf{G}, \boldsymbol{\Phi}_{trc})$  and  $\Gamma^{-1} = \text{diag}\{0, \dots, 0, \lambda_1^{-1}, \dots, \lambda_{n_{trc}}^{-1}\}$  where there are  $p$  zeros. Its IMSE coincides with the truncated version of  $\text{IMSE}(\mathcal{D}_n)$  and is given by

$$\text{IMSE}_{trc}(\mathcal{D}_n) = \text{trace}[(\mathbf{F}^\top\mathbf{K}_{err}^{-1}\mathbf{F} + \Gamma^{-1})^{-1}\mathbf{U}],$$

$$\text{where } \mathbf{U} = (\mathbf{f}|\mathbf{f}^\top)_{L_\mu^2} = \int_{\mathcal{X}} \mathbf{f}(x)\mathbf{f}^\top(x) d\mu(x) = \begin{pmatrix} (\mathbf{g}|\mathbf{g}^\top)_{L_\mu^2} & (\mathbf{g}|\boldsymbol{\phi}_{trc}^\top)_{L_\mu^2} \\ (\boldsymbol{\phi}_{trc}|\mathbf{g}^\top)_{L_\mu^2} & \text{Id}_{n_{trc}} \end{pmatrix},$$

with  $\text{Id}_{n_{trc}}$  the  $n_{trc}$ -dimensional identity matrix. Note that  $\mathbf{U}$  does not depend on  $\mathcal{D}_n$ . Approximation of  $Y_x$  by a heteroscedastic BLM where uncorrelated errors with variance  $\sigma^2(x) = K_{err}(x, x)$  are substituted for  $\varepsilon_x$ , gives a predictor with  $\text{IMSE}_{trc}^{hBLM}(\mathcal{D}_n) = \text{trace}[(\mathbf{F}^\top\boldsymbol{\Sigma}_{err}^{-1}\mathbf{F} + \Gamma^{-1})^{-1}\mathbf{U}]$ , i.e., a Bayesian L-optimality criterion which can be minimized following the same lines as in Sect. 3. However, in general the realizations of  $(Z_x)_{x \in \mathcal{X}}$  are not orthogonal to the linear subspace  $\mathcal{T}$  spanned by the trend functions  $\mathbf{g}$ , so that, roughly speaking, this direct approach involves unnecessarily large error variances  $\sigma^2(x)$ .

Denote by  $\mathfrak{p}$  the orthogonal projection of  $L_\mu^2(\mathcal{X})$  onto  $\mathcal{T}$ : for any  $f \in L_\mu^2(\mathcal{X})$ ,  $\mathfrak{p}f = \mathbf{g}^\top(\mathbf{g}|\mathbf{g}^\top)_{L_\mu^2}^{-1}(\mathbf{g}|f)_{L_\mu^2}$ . Suppose for simplicity that the sample realizations of



**Fig. 2** First 4 eigenfunctions  $\phi_k(x)$ ,  $x \in \mathcal{X} = [0, 1]$ , for the kernel  $K(x, x') = C_{3/2, 10}(|x - x'|)$  (Left) and for  $K^q(x, x')$  adapted to the trend functions  $\mathbf{g}(x) = (1, x, x^2)^\top$  (Right)

$(Z_x)_{x \in \mathcal{X}}$  belong to  $L^2_\mu(\mathcal{X})$  and define  $\mathfrak{p}Z_x = \mathbf{g}^\top(x)(\mathbf{g}|\mathbf{g}^\top)_{L^2_\mu}^{-1} \int_{\mathcal{X}} \mathbf{g}(t)Z_t d\mu(t)$ , so that  $Y_x = \mathbf{g}^\top(x)\underline{\theta} + Z_x = \mathbf{g}^\top(x)\underline{\theta} + \mathfrak{p}Z_x + \mathfrak{q}Z_x = \mathbf{g}^\top(x)\underline{\theta}' + \mathfrak{q}Z_x$ , where  $\mathfrak{q} = \text{id}_{L^2_\mu} - \mathfrak{p}$  and  $\underline{\theta}' = \underline{\theta} + (\mathbf{g}|\mathbf{g}^\top)_{L^2_\mu}^{-1} \int_{\mathcal{X}} \mathbf{g}(t)Z_t d\mu(t)$ . The covariance kernel of  $(\mathfrak{q}Z_x)_{x \in \mathcal{X}}$ , called reduction of the kernel  $K$  in [12], is

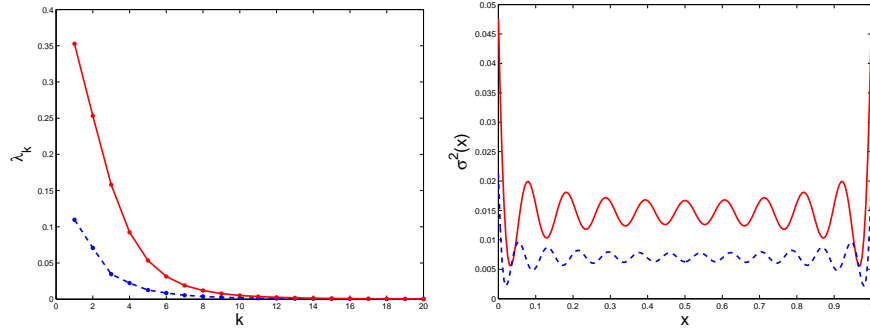
$$K^q(x, x') = E\{(\mathfrak{q}Z_x)(\mathfrak{q}Z_{x'})\} = K(x, x') + \mathbf{g}^\top(x)\mathbf{S}\mathbf{g}(x') - \mathbf{b}^\top(x)\mathbf{g}(x') - \mathbf{g}^\top(x)\mathbf{b}(x'),$$

where  $\mathbf{S} = (\mathbf{g}|\mathbf{g}^\top)_{L^2_\mu}^{-1}(T_\mu[\mathbf{g}|\mathbf{g}^\top])_{L^2_\mu}(\mathbf{g}|\mathbf{g}^\top)_{L^2_\mu}^{-1}$  and  $\mathbf{b}(x) = (\mathbf{g}|\mathbf{g}^\top)_{L^2_\mu}^{-1}T_\mu[\mathbf{g}](x)$ . A key property here is that, for a given design  $\mathcal{D}_n$ , the two random fields  $\mathbf{g}^\top(x)\underline{\theta} + Z_x$  and  $\mathbf{g}^\top(x)\underline{\theta}' + \mathfrak{q}Z_x$  with unknown  $\underline{\theta}$  yield the same BLUP and same IMSE. This can be related to results on intrinsic random functions, see [7]; notice, however, that the kernel reduction considered here applies to any linear parametric trend and is not restricted to polynomials in  $x$ .

Figure 2 presents the first four eigenfunctions for the kernels  $K(x, x')$  (Left) and  $K^q(x, x')$  (Right) for a one-dimensional random process on  $[0, 1]$  with Matérn 3/2 covariance function, see (5), and trend functions  $\mathbf{g}(x) = (1, x, x^2)^\top$ . The eigenfunctions oscillate at higher frequency for  $K^q(x, x')$  than for  $K(x, x')$ : in some sense, kernel reduction has removed from  $K$  the energy which can be transferred to the slowly varying trend functions. This is confirmed by Fig. 3-Left which shows the first 20 eigenvalues for both kernels. One can thus expect that for the same number  $n_{trc}$  of eigenfunctions, an heteroscedastic BLM will be more accurate when based on  $K^q$  than when based on  $K$ . This is confirmed by Fig. 3-Right where  $\sigma^2(x) = K_{err}(x, x)$  is plotted for both kernels when  $n_{trc} = 10$ .

**Acknowledgements** This work was partially supported by the ANR project DESIRE (DESIGns for spatial Random fElds), nb. 2011-IS01-001-01, joint with the Statistics Dept. of the JKU Linz (Austria). B. Gauthier has also been supported by the MRI Dept. of EdF Chatou from Sept. to Dec. 2014.





**Fig. 3** Left: first 20 eigenvalues (sorted by decreasing values) for the kernel  $K(x, x') = C_{3/2, 10}(|x - x'|)$  (top, solid line) and for  $K^q(x, x')$  adapted to the trend functions  $\mathbf{g}(x) = (1, x, x^2)^\top$  (bottom, dashed line). Right:  $\sigma^2(x)$ ,  $x \in \mathcal{X} = [0, 1]$ , for the heteroscedastic BLM with  $K(x, x')$  (top, solid line) and with  $K^q(x, x')$  (bottom, dashed line);  $n_{trc} = 10$  in both cases

## References

1. Armijo, L.: Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics* **16**(1), 1–3 (1966)
2. Fedorov, V.: Design of spatial experiments: model fitting and prediction. In: S. Gosh, C. Rao (eds.) *Handbook of Statistics*, vol. 13, chap. 16, pp. 515–553. Elsevier, Amsterdam (1996)
3. Fedorov, V., Müller, W.: Optimum design for correlated processes via eigenfunction expansions. In: J. López-Fidalgo, J. Rodríguez-Díaz, B. Torsney (eds.) *mODA'8 – Advances in Model-Oriented Design and Analysis, Proceedings of the 8th Int. Workshop, Almagro (Spain)*, pp. 57–66. Physica Verlag, Heidelberg (2007)
4. Gauthier, B., Pronzato, L.: Spectral approximation of the IMSE criterion for optimal designs in kernel-based interpolation models. *SIAM/ASA Journal on Uncertainty Quantification* **2**, 805–825 (2014)
5. Gauthier, B., Pronzato, L.: Approximation of IMSE-optimal designs via quadrature rules and spectral decomposition. *Communications in Statistics – Simulation and Computation* (2015). DOI:10.1080/03610918.2014.972518, to appear
6. Harari, O., Steinberg, D.: Optimal designs for gaussian process models via spectral decomposition. *Journal of Statistical Planning and Inference* **154**, 87–101 (2014)
7. Matheron, G.: The intrinsic random functions and their applications. *Advances in Applied Probability* pp. 439–468 (1973)
8. Pilz, J.: *Bayesian Estimation and Experimental Design in Linear Regression Models*, vol. 55. Teubner-Texte zur Mathematik, Leipzig (1983). (also Wiley, New York, 1991)
9. Pronzato, L., Pázman, A.: *Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties*. Springer, LNS 212, New York, Heidelberg (2013)
10. Sacks, J., Welch, W., Mitchell, T., Wynn, H.: Design and analysis of computer experiments. *Statistical Science* **4**(4), 409–435 (1989)
11. Santner, T., Williams, B., Notz, W.: *The Design and Analysis of Computer Experiments*. Springer, New York (2003)
12. Schaback, R.: Native Hilbert spaces for radial basis functions I. In: *New Developments in Approximation Theory*, pp. 255–282. Springer (1999)
13. Spöck, G., Pilz, J.: Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. *Stochastic Environmental Research and Risk Assessment* **24**(3), 463–482 (2010)