



# Bayesian adaptive reconstruction of profile optima and optimizers

David Ginsbourger, Jean Baccou, Clément Chevalier, Frédéric Perales, Nicolas Garland, Yann Monerie

## ► To cite this version:

David Ginsbourger, Jean Baccou, Clément Chevalier, Frédéric Perales, Nicolas Garland, et al.. Bayesian adaptive reconstruction of profile optima and optimizers. 2013. <hal-00920154>

HAL Id: hal-00920154

<https://hal.archives-ouvertes.fr/hal-00920154>

Submitted on 17 Dec 2013

**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.

# Bayesian adaptive reconstruction of profile optima and optimizers

David Ginsbourger<sup>\*†</sup>, Jean Baccou<sup>‡†</sup>, Clément Chevalier<sup>\*†</sup>, Frédéric Perales<sup>‡†</sup>, Nicolas Garland<sup>‡</sup>, Yann Monerie<sup>§</sup>

**Abstract.** Given a function depending both on decision parameters and nuisance variables, we consider the issue of estimating and quantifying uncertainty on profile optima and/or optimal points as functions of the nuisance variables. The proposed methods base on interpolations of the objective function constructed from a finite set of evaluations. Here the functions of interest are reconstructed relying on a kriging model, but also using Gaussian field conditional simulations, that allow a quantification of uncertainties in the Bayesian framework. Besides, we elaborate a variant of the Expected Improvement criterion, that proves efficient for adaptively learning the set of profile optima and optimizers. The results are illustrated on a toy example and through a physics case study on the optimal packing of polydisperse frictionless spheres.

**Key words.** Robust Design, Gaussian Process Regression, Uncertainty Quantification, Conditional Simulations, Active Learning, Expected Improvement

**1. Introduction.** A number of scientific and industrial questions boil down to optimization problems: Physical equilibria are found by minimizing adequate energy functionals, optimal engineering designs are sought by minimizing so-called fitness functions, etc. Mathematically, the deterministic system at hand (that can be defined also in the stochastic case, e.g., as an expectation) is classically seen as a real-valued objective function  $f : \mathbf{x} \in \mathcal{X} \rightarrow f(\mathbf{x}) \in \mathbb{R}$  with domain  $\mathcal{X} \subset E$ , where typically  $E = \mathbb{R}^d$  for some integer  $d \geq 1$ . Whenever  $f$  reaches a *global maximum*  $f^* = \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$ , values of  $\mathbf{x} \in \mathcal{X}$  such that  $f(\mathbf{x}) = f^*$  are usually called *points of (global) maximum* or *(global) maximizers*. Here we focus on cases where the input vector  $\mathbf{x}$  can be split into a group of *decision parameters*  $\alpha \in A$  and so-called *nuisance variables*  $v \in V$ , such that  $\mathcal{X} = A \times V$ . In such case, one can see the global maximum of  $f$  with respect to the nuisance variables  $v$  as a function of  $\alpha$ :

$$(1.1) \quad f^* : \alpha \in A \rightarrow f^*(\alpha) = \max_{v \in V} f(\alpha, v) \in \mathbb{R},$$

which we refer to as *profile optimum* or *conditional optimum* function. Assuming further that  $f(\alpha, \cdot)$  possesses a unique point of global maximum  $v_f^*(\alpha)$  for any given  $\alpha \in A$  (or defining one particular global optimizer in case there are several, e.g., the smallest one in dimension 1), we also consider the map of *profile* or *conditional optimizers*:

$$(1.2) \quad v_f^* : \alpha \in A \rightarrow v_f^*(\alpha) \in V.$$

Our main aim in the present work is to reconstruct  $f^*$  and  $v_f^*$  under drastically limited evaluation budget of the objective function  $f$ . An example of profile optimum and profile optimizer functions is shown on Figure 1.1, where the objective function  $f$  is the (negative) Branin-Hoo function [18].

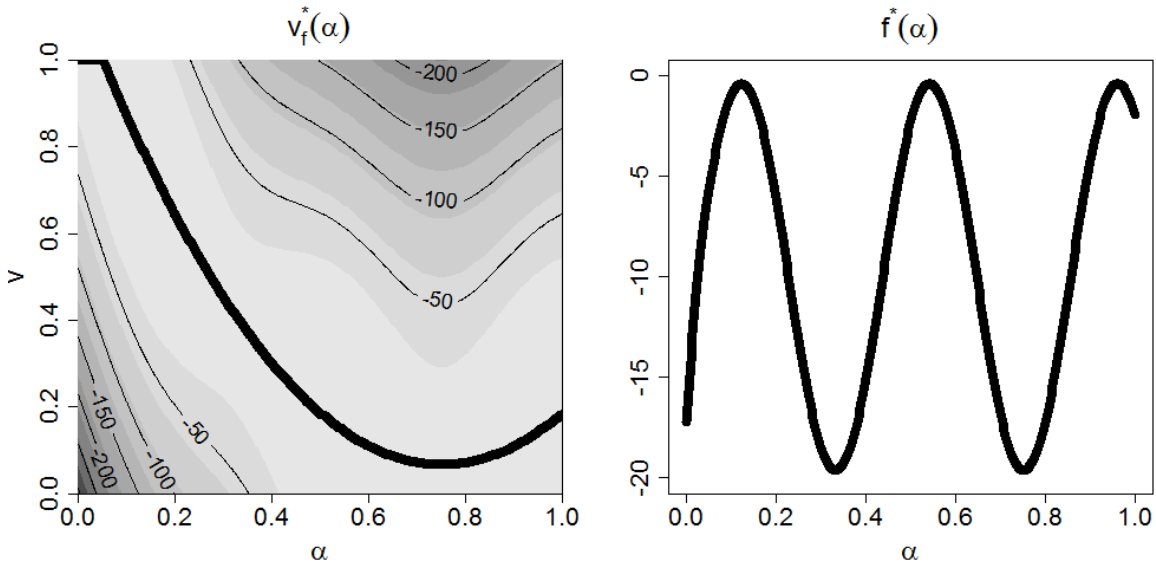
---

<sup>\*</sup>Department of Mathematics and Statistics, University of Bern, Switzerland

<sup>†</sup>ReDICE Consortium

<sup>‡</sup>Institut de Radioprotection et de Sûreté Nucléaire, Cadarache, France

<sup>§</sup>Université de Montpellier, France



**Figure 1.1.** Left: profile optimizer  $v_f^*$  (black curve) and contour lines of a toy function  $f$ . Right: Profile optimum function  $f^*$ .

Concrete problems involving the reconstruction of profile optima and/or optimizers occur in various contexts. Let us give a few examples:

- In reliability engineering, the dangerousness of a system may depend both on controllable design variables  $\alpha$  like materials and geometry, and also on uncontrollable variables  $v$  like future climatic conditions (wind characteristics, temperature trends, etc.). In such case, one may be interested in the evolution of the worst-case scenario with respect to the climatic conditions, seen as a function of the design variables. The profile optimum function then maps any design  $\alpha$  to the associated maximum level of danger with respect to the climate variable, while the profile optimizer function represents the worst climatic condition  $v_f$  (i.e. the one implying the highest level of danger) as a function of the design variables  $\alpha$ . A neighbouring problem referred to as robust inversion has been tackled in [5], with nuclear criticality safety applications.
- In statistical modelling, it often arises that some likelihood function may be analytically maximized as function of a subgroup of parameters, the other model parameters being considered as fixed. One then speaks of *profile Likelihood* [33]. E.g., this happens in the case of centred Gaussian vector with stationary covariance depending on a variance  $v$  (or its square root, often named scale) and on some correlation parameter  $\alpha$ . It is well-known that the maximum likelihood estimator of  $v$  can be derived analytically as a function of  $\alpha$ . Plotting the optimal  $v$  as a function of  $\alpha$  then amounts to representing what we call here a profile optimizer function.

In the present work the main test case comes from physics, more precisely from an optimal packing problem. A medium containing bidisperse frictionless spheres is considered, and the maximum occupation density is numerically investigated as a function of the radii and a coefficient tuning the volume fractions. A particular focus is put on the estimation of optimal

volume fraction coefficient ( $v_f^*$ ) seen as a function of the ratio of radii ( $\alpha$ ).

When addressing such problems in realistic conditions, each pointwise evaluation of the function  $f$  corresponds to a timely expensive experiment (be it physical or numerical), so that the number of calls to  $f$  is drastically limited by practical constraints. It is then mandatory to appeal to approximations of  $f$  for reconstructing the  $f^*$  and  $v_f^*$  functions. Here we concentrate on the use of kriging and Gaussian process models for estimating them, quantifying the uncertainties bond to these estimations, and reducing those uncertainties by appealing to adaptive evaluation strategies in the sequential Bayesian framework.

The article is organized as follows: in Section 2, we present plug-in and Bayesian approaches for reconstruction by kriging and uncertainty quantification on profile optimum and optimizer functions. In Section 3, we then propose a new infill sampling criterion, obtained as a variant of the Expected Improvement, that aims at reinforcing exploration in regions of profile optima. Detail is given on the computation and the numerical maximization of this new criterion, and a proof of consistency is provided for the associated sequential strategy under fixed mean and covariance hyperparameters. In Section 4, applications of the proposed methodology are presented on an optimal packing problem. Conclusions and perspectives of future research are given in Section 5.

## 2. Plug-in and Bayesian reconstructions in the static framework.

**2.1. Plug-in approach: Estimating  $f^*$  and  $v_f^*$  based on an approximation of  $f$ .** We now assume that the function  $f$  was already evaluated at a set of points, or *design of experiments*,

$$(2.1) \quad \mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} = \{(\alpha_1, v_1), \dots, (\alpha_n, v_n)\} \in (A \times V)^n.$$

Reconstructing  $f^*$  and  $v_f^*$  directly from such data is not straightforward. A natural idea, provided some prior knowledge on the function  $f$  is available, is to use an interpolator or an approximation model for  $f$ , and to base on this model to reconstruct the profile maps. A plethora of interpolation and approximation approaches may be envisaged [15], encompassing polynomials, methods based on Fourier and wavelets, splines, neural nets, and kernel methods. Here we mainly focus on kriging [22, 7, 34, 6], a method originally used for interpolating field data in geosciences. Assuming that noise-free evaluations of  $f$  at  $\mathbf{X}$  are available, and introducing the vector of observations

$$(2.2) \quad \mathbf{z} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))' \in \mathbb{R}^n,$$

where  $\mathbf{u}'$  denotes the transpose of the vector  $\mathbf{u}$ , the so-called Ordinary Kriging (OK) predictor writes

$$(2.3) \quad m(\mathbf{x}) = \hat{\mu} + \mathbf{k}(\mathbf{x})' K^{-1}(\mathbf{z} - \hat{\mu}\mathbf{1}),$$

where  $\hat{\mu} = (\mathbf{1}' K^{-1} \mathbf{1})^{-1} \mathbf{1}' K^{-1} \mathbf{z}$ ,  $\mathbf{k}(\mathbf{x}) = (k(\mathbf{x}_1, \mathbf{x}), \dots, k(\mathbf{x}_n, \mathbf{x}))'$ ,  $K = (k(\mathbf{x}_i, \mathbf{x}_j))_{(i,j) \in \{1, \dots, n\}^2}$ ,  $\mathbf{1} = (1, \dots, 1)'$ , and  $k$  is a kernel of positive type chosen *a priori* and/or tuned based on data. In this noise-free context, a notable fact is that the OK predictor interpolates the data:

$$(2.4) \quad m(\mathbf{x}_i) = f(\mathbf{x}_i) \quad (1 \leq i \leq n).$$

Note that in the case where the observations are corrupted by a Gaussian noise with mean 0 and variance  $\tau^2$ , the kriging equations are slightly modified through a substitution of  $K$  by  $K + \tau^2 I$ , and the interpolation property is lost. This variant (which slightly departs from kriging with a nugget effect) is referred to as kriging with homogeneous noise variance in the present work. Coming back to the profile functions and substituting  $f$  by  $m$ , we simply obtain *plug-in* estimates for  $f^*$  and  $v_f^*$ :

$$(2.5) \quad m^*(\alpha) = \max_{v \in V} m(\alpha, v),$$

and  $v_m^*(\alpha)$  such that

$$(2.6) \quad m(\alpha; v_m^*(\alpha)) \geq m(\alpha, v) \quad (\alpha \in A, v \in V).$$

Note that even if unicity of the profile maximizers is assumed for  $f$ , there is no reason why this unicity should propagate to  $m$ . Here, for the sake of simplicity, we assume that the profile maximizer of  $m$  is unique for any given  $\alpha \in A$ , a realistic assumption in practice. Concerning the quality of the approximation of  $f^*$  by  $m^*$ , it is straightforward that

$$(2.7) \quad \|f^* - m^*\|_\infty \leq \|f - m\|_\infty,$$

where the norms refer to function spaces on  $A$  and  $A \times V$ , respectively. In other words, if one controls the error of the kriging predictor uniformly, the same control is guaranteed on the map of profile maxima. The same doesn't hold, however, for the profile maximizers: whatever  $\|f - m\|_\infty$  may be, it is always possible to have the maximizers  $v_f^*(\alpha)$  and  $v_m^*(\alpha)$  at any given distance within the domain  $V$ .

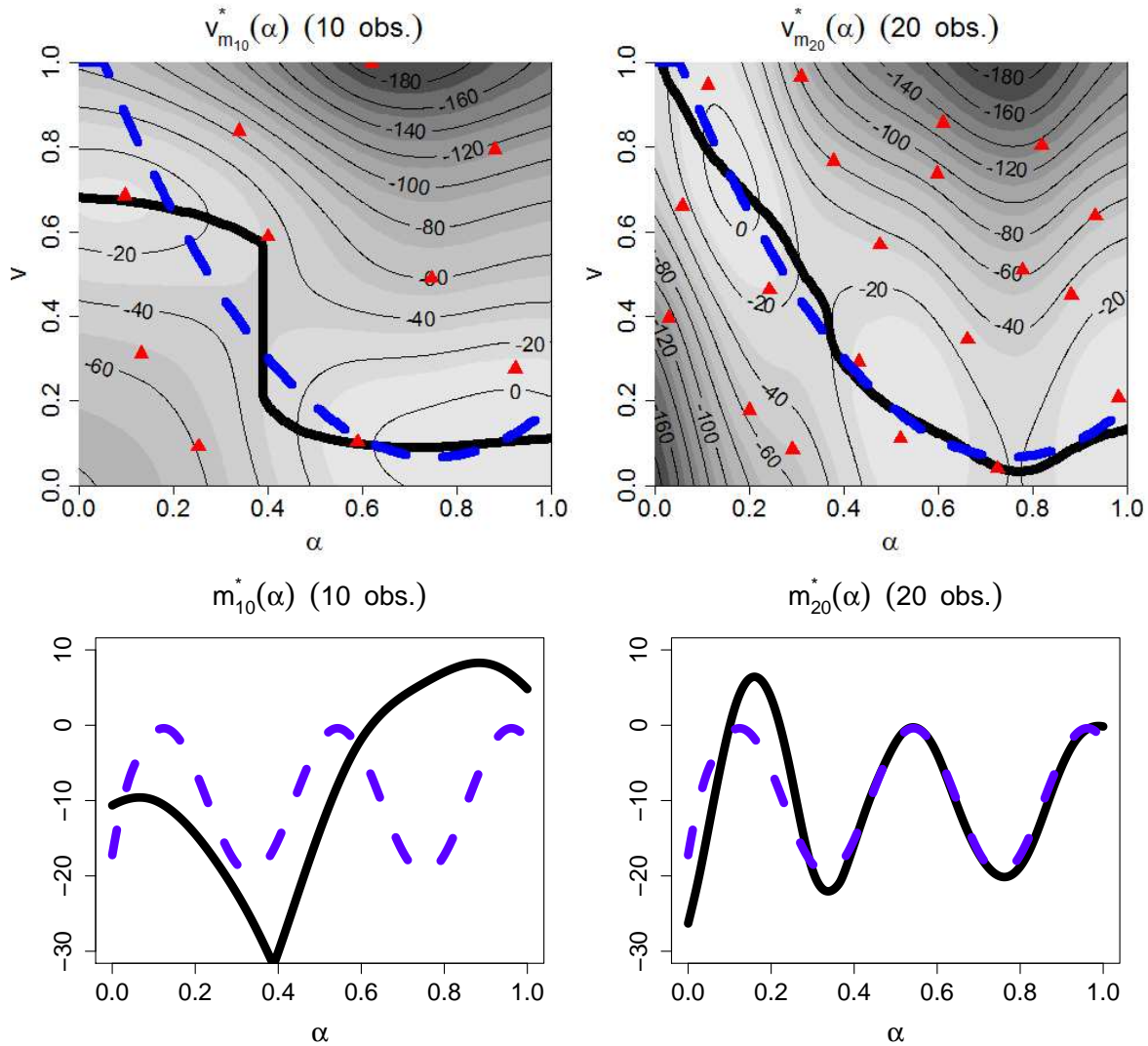
Below (see, Figure 2.1), we give some first illustrations on a toy test function. Here, two different designs of experiments are considered. In the first case,  $f$  is known at  $n = 10$  points only, forming a so-called Latin Hypercube Design (LHD). In the second case, the LHD is made of  $n = 20$  points. Here the LHDs are generated using the R package *lhs* [4] with "optimum" option and default settings. In both cases, an anisotropic Matérn kernel with  $\nu = 3/2$  is chosen for  $k$ , and the corresponding scale and range hyperparameters are estimated by maximum likelihood using the *DiceKriging* package [31]. One can see on Figure 2.1 that  $f^*$  and  $v_f^*$  are better reconstructed with the 20-point design, as expected. More specifically, Table 2.1 gives the bias obtained on the profile optimum and profile optimizer functions, measured as follows:

$$(2.8) \quad b^{v_f^*} := \|v_m^* - v_f^*\|$$

$$(2.9) \quad b^{f^*} := \|m^* - f^*\|,$$

where the considered norms are either  $L^2$  or  $L^\infty$  on appropriate spaces.

In applications, the use of the bias as a performance indicator is unrealistic as the true  $v_f^*$  and  $f^*$  are unknown. The next subsection discusses uncertainty quantification approaches in the case where a Gaussian random field prior is put on the objective function  $f$ .



**Figure 2.1.** Plug-in estimates  $v_m^*$  (top) and  $m^*$  (bottom) obtained from 10-point (left) and 20-point (right) experimental designs. The estimates are thick black curves. The blue dashed curves represent the actual  $v_f^*$  and  $f^*$  functions. The triangles stand for the design points.

**Table 2.1**

Model bias if 10 or 20 locations are evaluated

norm	indicator	10 obs.	20 obs.
$L^2$	$b^{f^*}$	13	4
$L^2$	$b^{v_f^*}$	0.13	0.04
$L^\infty$	$b^{f^*}$	25	10
$L^\infty$	$b^{v_f^*}$	0.32	0.08

**2.2. Bayesian approach, with Gaussian field priors.** An alternative approach proposed here, that complements the prediction by kriging presented in the last section, consists in assuming a Gaussian random field prior for the function  $f$ , and then in relying on the obtained posterior distribution to draw conclusions on the profile optima and optimizer functions. In such framework,  $f$  is hence seen as one possible realization (i.e.  $Z(\omega)$  for some  $\omega \in \Omega$ ,  $\Omega$  being the underlying probability space) of a Gaussian random field  $(Z_{\mathbf{x}})_{\mathbf{x} \in \mathcal{X}}$  with mean  $\mu$  and covariance kernel  $k$ . Then not only may  $f$  be approximated by the conditional expectation  $\widehat{Z}_{\mathbf{x}} = \mathbb{E}[Z_{\mathbf{x}}|Z_{\mathbf{X}}]$ , but the approximation error happens to be itself a centred random field with analytically tractable covariance. Under well-known technical conditions [27] (e.g., uniform improper prior on  $\mu$ ), it turns out that the conditional expectation coincides with the OK predictor of Eq.(2.3):

$$(2.10) \quad \mathbb{E}[Z_{\mathbf{x}}|Z_{\mathbf{X}} = \mathbf{z}] = m(\mathbf{x})$$

while the covariance of the prediction error writes

$$(2.11) \quad \text{Cov}[Z_{\mathbf{x}} - \widehat{Z}_{\mathbf{x}}, Z_{\mathbf{z}} - \widehat{Z}_{\mathbf{z}}] = k(\mathbf{x}, \mathbf{z}) - \mathbf{k}(\mathbf{x})' K^{-1} \mathbf{k}(\mathbf{z}) + \frac{(1 - \mathbf{k}(\mathbf{x})' K^{-1} \mathbf{1})(1 - \mathbf{k}(\mathbf{z})' K^{-1} \mathbf{1})}{\mathbf{1}' K^{-1} \mathbf{1}}.$$

Using this result, one may simulate Gaussian random field realizations that honor the data while giving a quantification of uncertainty on various quantities regarding  $f$ . Here we use that workflow to enhance the estimation of the  $f^*$  and  $v_f^*$  functions, and also provide associated measures of accuracy that do not depend on validation data. Instead of basing our estimates on a single predictor as in the plug-in approach, we rely on a set of conditional simulations

$$z_i(\cdot) = Z(\omega_i) \quad (1 \leq i \leq s)$$

In other words, the  $f^*$  and  $v_f^*$  functions are represented by the random process  $Z^*$  and the random field  $v_{Z^*}^*$ , being themselves approximated by the sets of functions  $\{z_1^*, \dots, z_s^*\}$  and  $\{v_{z_1}^*, \dots, v_{z_s}^*\}$ . An example of such sets of functions, obtained from conditional simulations, is shown in Figure 2.2. Here  $s = 25$  conditional simulations are computed on a  $50 \times 50$  grid, leading to 25 functions  $z_i^*$  and  $v_{z_i}^*$ ,  $1 \leq i \leq s$ . One clearly sees that the functions  $z_i^*$  and  $v_{z_i}^*$  have a higher dispersion when they are obtained from simulations conditioned on 10 observations (left plots). This suggests to use this dispersion as an indicator to quantify the uncertainty, given the observations, on the true unknown functions  $f^*$  and  $v_f^*$ .

How to quantify the variability associated with a set of curves remains an open question, to which various answers have been given throughout the literature of probability theory [25]. In the present work, we quantify this variability using the expectation of the distance between 2 (independently) simulated profile optimizers  $v_{Z_1}^*$ ,  $v_{Z_2}^*$  or profile optimum functions  $Z_1^*$ ,  $Z_2^*$ :

$$(2.12) \quad H^{v_f^*} := \mathbb{E} [\|v_{Z_1}^* - v_{Z_2}^*\| | Z_{\mathbf{X}} = \mathbf{z}]$$

$$(2.13) \quad H^{f^*} := \mathbb{E} [\|Z_1^* - Z_2^*\| | Z_{\mathbf{X}} = \mathbf{z}] ,$$

---

<sup>1</sup>Cf. [19] for a discussion on the almost sure unicity of the optimizer for suitable Gaussian random field models.



where the considered norms are either the  $L^2$  or  $L^\infty$  norm. Note that  $H_n^{v_f^*}$  and  $H_n^{f^*}$  can be computed without knowing the true functions  $f^*$  or  $v_f^*$ . In Section 4 we will also compute the so-called “risks”

$$(2.14) \quad r^{v_f^*} := \mathbb{E} [\|v_{Z_1}^* - v_f^*\| | Z_{\mathbf{X}} = \mathbf{z}],$$

$$(2.15) \quad r^{f^*} := \mathbb{E} [\|Z_1^* - f^*\| | Z_{\mathbf{X}} = \mathbf{z}],$$

which require to know  $f^*$  and  $v_f^*$ , respectively. Table 2.2 gives these indicators on the studied toy example using 1000 conditional simulations. It appears that the variability and the risk are lower for 20-point design than for the 10-point design, as could be expected. In general, one may expect the uncertainty to be reduced if more locations are evaluated. Under a limited evaluation budget, the next evaluation location must be chosen carefully. The next section tackles this issue and introduces a new sampling criterion, the *Profile Expected Improvement*.

**Table 2.2**

*Risk and uncertainty indicators on the profile optima and optimizer functions, for the  $L^2$  and  $L^\infty$  norms, if the Branin-Hoo function is evaluated at 10 or 20 design points. The set up is the same as for Fig. 2.2.*

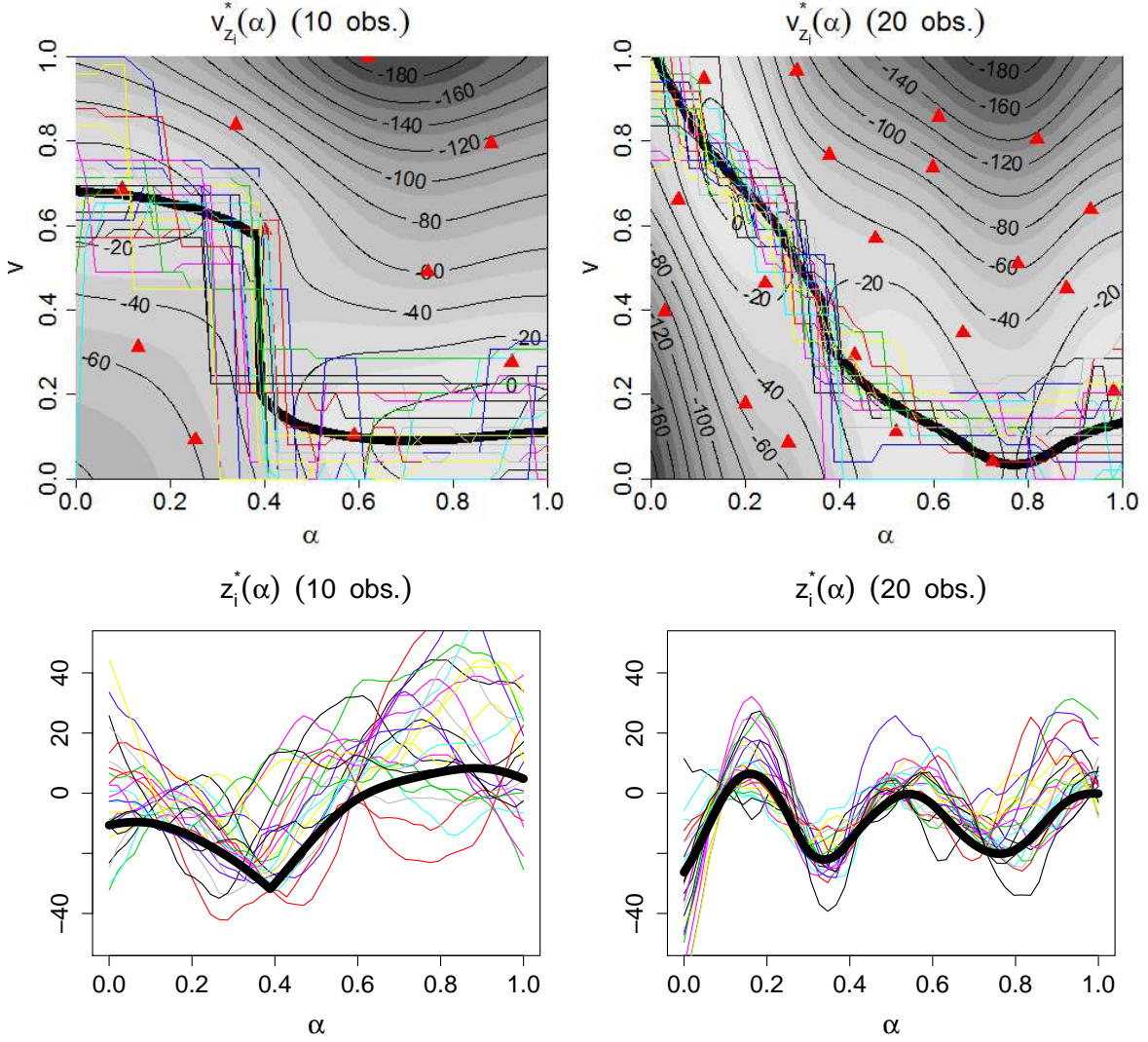
norm	indicator	10 obs.	20 obs.
$L^2$	$H^{f^*}$	21	12
$L^2$	$r^{f^*}$	22	10
$L^2$	$H^{v_f^*}$	0.18	0.13
$L^2$	$r^{v_f^*}$	0.17	0.10
$L^\infty$	$H^{f^*}$	47	33
$L^\infty$	$r^{f^*}$	46	26
$L^\infty$	$H^{v_f^*}$	0.43	0.28
$L^\infty$	$r^{v_f^*}$	0.39	0.21

### 3. Adaptive design of experiments for Bayesian profile curve reconstruction.

**3.1. Problem set-up: how to choose the next point(s)?** We observed previously on a toy example that quality of the profile functions reconstruction increased with the size of the design of experiments, a feature corroborated by the variability of the simulated profiles obtained by conditional simulations. Of course, no general conclusion may be drawn from a single comparison with arbitrary settings, but this reflects what may reasonably be expected: adding experiments generally tends to improve the profile estimates and reduce the associated uncertainty. Note that in the example above, the two designs of experiments are generated independently of each other. In practice, however, one rather needs to augment an existing design with new well-chosen points, be it in a sequential (one point after the other, like in [18]) or in a batch-sequential manner [14].

Several approaches have been explored for efficient sequential and batch-sequential strategies dedicated to various goals, in the Gaussian process framework. A standard problem is to design new experiments for reducing the integrated mean-squared error (IMSE), i.e. improving the predictivity of a Gaussian process model by treating all areas of the domain equally [32]. Variants of IMSE criteria have been proposed, notably for learning contour lines of a





**Figure 2.2.** Profile optimizers  $v_{z_i}^*$  and profile optima functions  $z_i^*$  of  $s = 25$  realizations  $z_i, 1 \leq i \leq s$  of a Gaussian process conditioned on 10 (left) and 20 (right) observations of the Branin-Hoo function.

function corresponding to a prescribed target level [29]. Another question that has inspired a number of sequential and batch-sequential strategies is the global optimization of  $f$ , i.e. both estimating the global maximum and locating the corresponding global maximizer(s). Various infill sampling criteria for sequential optimization strategies using Gaussian process models have been proposed [17], including the *Probability of Improvement* (PI), the *Expected Improvement* (EI), and the expected conditional entropy of the maximizer [37]. Due to its conceptual simplicity and its good empirical performances, EI has become very popular in the past 15 years, for solving medium-dimensional non-convex global optimization problem under a severely limited budget and a derivative-free framework. Even though it is known that the optimality properties of EI are only in the *one step lookahead* framework [24, 13], there exists some theoretical guarantees on the convergence of EI algorithms given a few restrictions on

the family of used kernels [35]. Several variants of EI for constrained or batch-sequential problems exist, and the criterion and its maximization are available in open source software [31, 28].

**3.2. Profile-EI: A variant of the expected improvement criterion.** In the present work, we aim at adapting EI in order to create an efficient infill sampling criterion for sequentially learning the profile maxima and/or profile maximizer functions. Before looking more precisely into the matter and presenting the so-called *Profile-EI* (PEI) criterion, let us briefly recall the basics of EI. In EI settings, the variable  $\mathbf{x}$  does not have to be split into two sub-groups ( $v$  and  $\alpha$ ); everything happens as if there was no parameter  $\alpha$ , so the domain of  $\mathbf{x}$  reduces to  $V$ . The goal being to globally maximize  $f$ , a quantity of interest after evaluation of  $f$  at  $n$  points  $\mathbf{x}_1, \dots, \mathbf{x}_n \in V$  is the *current maximum*  $\max_{1 \leq i \leq n} f(\mathbf{x}_i)$ . Sticking to the random process notations, the EI criterion at an arbitrary point  $\mathbf{x} \in \mathcal{X}$  is then the average departure above the current maximum of the response at the considered  $\mathbf{x}$ , under the Gaussian field model:

$$(3.1) \quad EI(\mathbf{x}) = \mathbb{E} \left[ \left( Z_{\mathbf{x}} - \max_{1 \leq i \leq n} Z_{\mathbf{x}_i} \right)^+ \mid Z_{\mathbf{X}} = \mathbf{z} \right] = \mathbb{E} \left[ \max \left( 0, Z_{\mathbf{x}} - \max_{1 \leq i \leq n} Z_{\mathbf{x}_i} \right) \mid Z_{\mathbf{X}} = \mathbf{z} \right].$$

A very practical aspect about the EI criterion defined in Eq. (3.1) is that a analytical formula exists, that expresses EI in terms of a simple function of the kriging predictor and the associated prediction variance at the point of interest only:

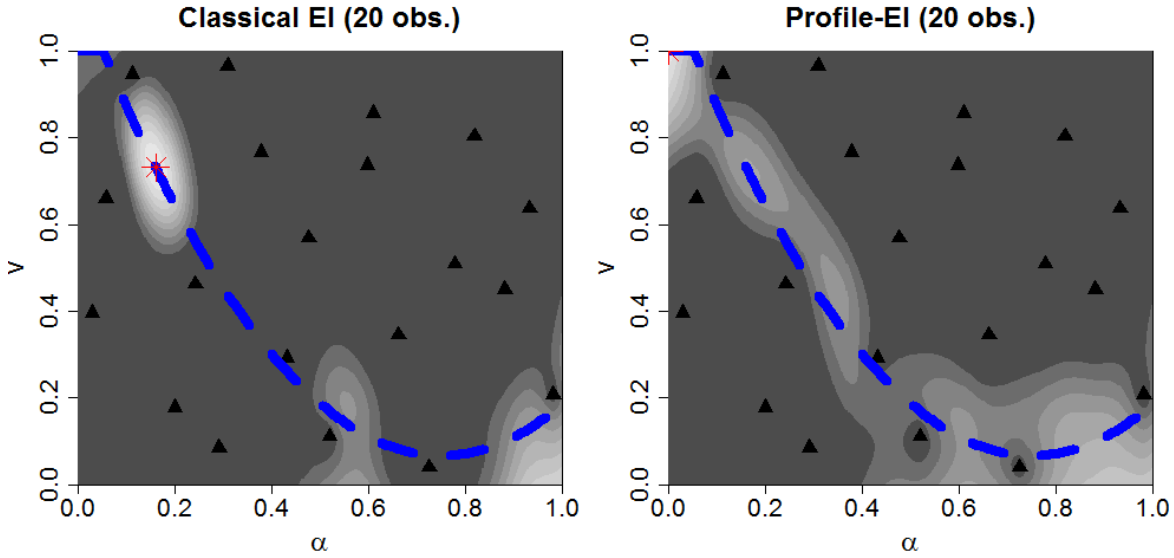
$$(3.2) \quad EI(\mathbf{x}) = \begin{cases} s(\mathbf{x}) (u(\mathbf{x})\Phi(u(\mathbf{x})) + \phi(u(\mathbf{x}))) & \text{if } s^2(\mathbf{x}) \neq 0 \\ 0 & \text{else} \end{cases}$$

where  $u(\mathbf{x}) = \frac{m(\mathbf{x}) - T}{s(\mathbf{x})}$ ,  $T = \max_{1 \leq i \leq n} f(\mathbf{x}_i)$ ,  $s^2(\mathbf{x})$  is the Ordinary Kriging prediction variance (given by Eq. (2.11) for  $\mathbf{z} = \mathbf{x}$ ), and  $\Phi, \phi$  respectively denote the cumulative distribution function and probability density function of the standard Gaussian distribution.

Back to our motivating problem of profile estimation, distinguishing between  $\alpha$  and  $v$  is needed. Indeed, we do not wish our sequential strategy to target only regions with highest response level over the whole domain  $A \times V$ . What is needed is rather a strategy that visits points having a potential for being points of profile maximum, for all values of the parameter  $\alpha$ . Hence, between two points corresponding to two different values of  $\alpha$ , we would favour the point that has the most potential of improvement *with respect to the current maximum* corresponding to its  $\alpha$  value. However, as no response value may have already obtained with exactly the same  $\alpha$ , our approach consists in calculating a *profile EI* (PEI), that compares  $Z_{(\alpha, v)}$  with  $\max_{w \in V} m(\alpha, w)$ , for any given  $v \in V$  and  $\alpha \in A$ :

$$(3.3) \quad PEI(\alpha, v) := \mathbb{E} \left[ \left( Z_{(\alpha, v)} - \min \left( \max_{w \in V} m(\alpha, w), \max_{1 \leq i \leq n} Z_{(\alpha_i, v_i)} \right) \right)^+ \mid Z_{\mathbf{X}} = \mathbf{z} \right]$$

The "cap" at the current maximum  $\max_{1 \leq i \leq n} Z_{(\alpha_i, v_i)}$  is added for preventing the criterion to vanish in regions where  $\max_{w \in V} m(\alpha, w)$  would artificially overshoot the data and prohibit



**Figure 3.1.**  $EI(\alpha, v)$  and  $PEI(\alpha, v)$  functions obtained from a 20-point initial design of experiments (black triangles). The blue dashed curves represents the actual  $v_f^*$  functions.  $EI$  and  $PEI$  are represented through grayscale contour lines (with colors ranging from dark to light gray with increasing criterion value).

global exploration of the input space (this cap is needed for the consistency proof given later in this section). In turn,  $PEI$  inherits the tractability of  $EI$ , leading to the formula:

$$(3.4) \quad PEI(\alpha, v) = \begin{cases} s(\alpha, v) (g(\alpha, v)\Phi(g(\alpha, v)) + \phi(g(\alpha, v))) & \text{if } s^2(\alpha, v) \neq 0 \\ 0 & \text{else} \end{cases}$$

where  $g(\alpha, v) := \frac{m(\alpha, v) - t(\alpha)}{s(\alpha, v)}$  and  $t(\alpha) := \min(\max_{w \in V} m(\alpha, w), \max_{1 \leq i \leq n} f(\alpha_i, v_i))$ .

Figure 3.1 highlights the differences between the Profile-EI and the classical EI, using our toy example and a 20-point experimental design. The key difference here is that the expectation of the improvement in the classical EI is usually computed with respect to the current maximum observation  $\max_{1 \leq i \leq n} Z_{\mathbf{x}_i}$  while, in the  $PEI$ , the threshold depends on  $\alpha$ . This feature allows a trade-off between global exploration and an intensified search of locations that are expected to be close to the profile optimizer curve  $v_f^*$ .

**3.3. Computation and maximization of Profile-EI.** From Eq. (3.3), one sees that each computation of  $PEI$  at one location  $(\alpha, v) \in A \times V$  involves a global optimisation over  $V$  to find the threshold  $t(\alpha)$ . A brute-force maximization of  $PEI$  over  $A \times V$  then appears to be computationally very expensive. On the other hand, if, for a given  $\alpha \in A$ , the threshold  $t(\alpha)$  is computed, then computing  $PEI$  at any  $(\alpha, v), v \in V$  simply requires computing kriging means and variances and to apply the (fast to evaluate) formula (3.4). To take advantage of this, instead of maximizing the  $PEI$  over  $A \times V$ , we equivalently maximize the profile  $PEI$ :

$$(3.5) \quad PEI^*(\alpha) := \max_{w \in V} PEI(\alpha, w),$$

which is defined on  $A$  (and not  $A \times V$ ) and which computation requires 2 global optimizations:

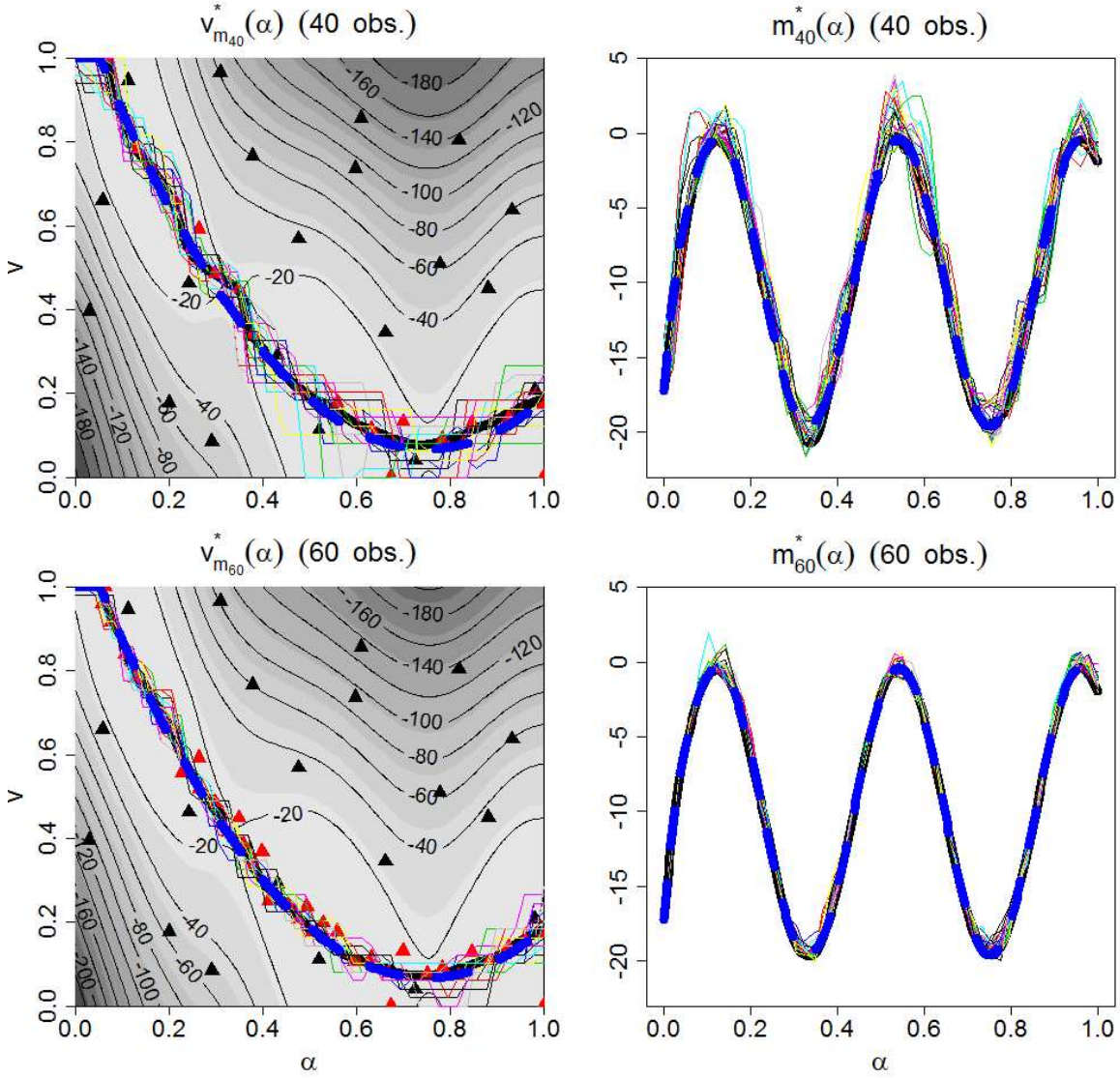
- one global optimization to find the threshold  $t(\alpha)$ ,
- one global optimization to maximize  $PEI(\alpha, \cdot)$  over  $V$ .

Once a global maximizer  $\alpha^*$  of  $PEI^*$  is identified, we obtain a global maximizer of PEI by taking  $(\alpha^*, v^*)$ , where  $v^*$  maximizes  $PEI(\alpha^*, \cdot)$ .

The optimizations involved in the computation of  $PEI^*$  only require to compute kriging means and variances. The latter can be computed in a large number of locations  $\mathbf{x}_1, \dots, \mathbf{x}_p \in A \times V$  by pre-computing the quantities that do not depend on them (in particular, the inverse  $K^{-1}$  of the covariance matrix at the observation locations, and the matrix-vector product between  $K^{-1}$  and the column vectors that do not involve  $\mathbf{x}$ , see Eqs. (2.3), (2.11)). For large  $p$  and with the proper precomputations, the computational cost is  $O(np)$  to compute  $p$  kriging means and  $O(n^2p)$  to compute  $p$  kriging variances, where  $n$  is the current number of observations. In the present article, given the relative simplicity of the examples, the two optimizations (over  $V$ ) involved in Eq. (3.5) and the optimization of  $PEI^*$  (over  $A$ ) are handled through a regular grid of size  $p$  for both  $\alpha$  and  $v$ , so that the optimal location  $(\alpha^*, v^*)$  can be found by computing (at the same time)  $p^2$  kriging mean and variances. However, given the clear limitation of this approach when the dimension of the problem increases, our code also permits to use other optimization methods, like the genetic algorithm implemented in the `genoud` R package [23].

Figure 3.2 exhibits the sequence of points evaluated by sequentially maximizing PEI. The 20-point initial design of Section 2.2 and 3.2 is used. Then, 40 points are added sequentially. One can see (bottom left and right plots) that after 40 iterations, both the true profile optima and optimizer functions are well estimated (small bias). In addition, conditional simulations exhibit a low variability, which means that the uncertainties  $H^{f^*}$  and  $H^{v^*}$  have been reduced. This uncertainty reduction is quantified (relying on 1000 conditional simulations) in Table 3.1, and represented on Figure 3.3.

**3.4. Convergence property with fixed mean and covariance.** Here we follow the route of [35] where convergence properties for the expected improvement algorithm under fixed mean and covariance functions are given. Since our criterion departs to some extent from the class of criteria considered in [35], we needed to revisit the results, which lead us to the proofs given in this section. To facilitate back and forth switching between the present article and [35], we temporarily (for the current subsection only) adopt their notations. A brief dictionary between our notations and their notations is now given. The random field is denoted by  $\xi$ , instead of  $Z$ , and  $Z_{\mathbf{x}}(\omega)$  becomes  $\xi(x, \omega) = \omega(x)$  (where  $\omega$  stands both for a probabilistic outcome and for a path of  $\xi$ ) while  $\xi(x)$  is used when the dependence on  $\omega$  is implicit. Besides the set of points  $(x_1, \dots, x_n) \in \mathbb{X}^n$  (instead of  $\mathbf{x}_1, \dots, \mathbf{x}_n \in D$ ) is denoted in short by  $\underline{x}_n$ . The kriging mean at point  $x$  of the function  $\omega(\cdot)$  after having observed  $\underline{x}_n$  is denoted  $\hat{\xi}_n(x, \omega; \underline{x}_n)$ ,  $\hat{\xi}_n(x; \underline{x}_n)$ , or  $\hat{\xi}_n(x)$  in short, depending on the context. The mean-square prediction error (or *kriging variance*) then writes  $\sigma_n^2(x; \underline{x}_n) := \mathbb{E} \left[ \left( \xi(x) - \hat{\xi}_n(x; \underline{x}_n) \right)^2 \right]$ . Finally, the considered class of criteria including the Expected Improvement as particular case writes  $\rho_n(x) = \gamma(\hat{\xi}_n(x) - M_n; \sigma_n^2(x))$  where  $M_n$  is the maximum of the responses available at step  $n$  (i.e.,  $\max(\omega(x_1), \dots, \omega(x_n))$ ), and  $\gamma : \mathbb{R} \times [0, +\infty) \rightarrow [0, +\infty)$  is a continuous function satisfying  $\forall z \leq 0, \gamma(z, 0) = 0$  and



**Figure 3.2.** Plug-in estimate (solid dark curves)  $v_m^*$  (left) and  $m^*$  (right) together with the true functions  $v_f^*$  and  $f^*$  (blue dashed curves) after having evaluated an initial 20-point design of experiments (black triangles) and 20 (top) or 40 (bottom) additional points (red triangles) chosen by sequentially maximizing PEI. In addition, 25 profile optima  $z_i^*$  and profile optimizers  $v_{z_i}^*$  are printed (thin lines in various colors), obtained from Gaussian process realizations conditioned on the available observations.

$\forall z \in \mathbb{R}, \forall s > 0, \gamma(z, s) > 0$ . For the special case of the Expected Improvement, the  $\gamma$  function is defined by

$$(3.6) \quad \gamma(z, s) = \begin{cases} \sqrt{s}\phi(z/\sqrt{s}) + z\Phi(z/\sqrt{s}) & \text{if } s > 0, \\ z^+ = \max(z, 0) & \text{if } s = 0. \end{cases}$$

Coming back to the criterion proposed in the present work, the profile EI now writes

$$(3.7) \quad \rho_n^*(x) = \gamma(\hat{\xi}_n(x) - M_n^*(x); \sigma_n^2(x))$$



Table 3.1

Uncertainty, risk and bias indicators on the profile optimizer and optima functions after having evaluated the Branin-Hoo function at a 20-point initial design of experiments, and after having added 20 and 40 additional points by PEI maximization.

norm	indicator	20 obs.	20 + 20 obs.	20 + 40 obs.
$L^2$	$H^{f^*}$	11.4	1.6	0.6
$L^2$	$r^{f^*}$	9.6	1.3	0.5
$L^2$	$b^{f^*}$	3.9	0.4	0.1
$L^2$	$H^{v_f^*}$	0.13	0.07	0.04
$L^2$	$r^{v_f^*}$	0.10	0.05	0.03
$L^2$	$b^{v_f^*}$	0.04	0.02	0.01
$L^\infty$	$H^{f^*}$	32.2	4.5	1.8
$L^\infty$	$r^{f^*}$	25.4	3.7	1.5
$L^\infty$	$b^{f^*}$	10.2	1.4	0.3
$L^\infty$	$H^{v_f^*}$	0.28	0.16	0.10
$L^\infty$	$r^{v_f^*}$	0.21	0.12	0.08
$L^\infty$	$b^{v_f^*}$	0.09	0.07	0.02

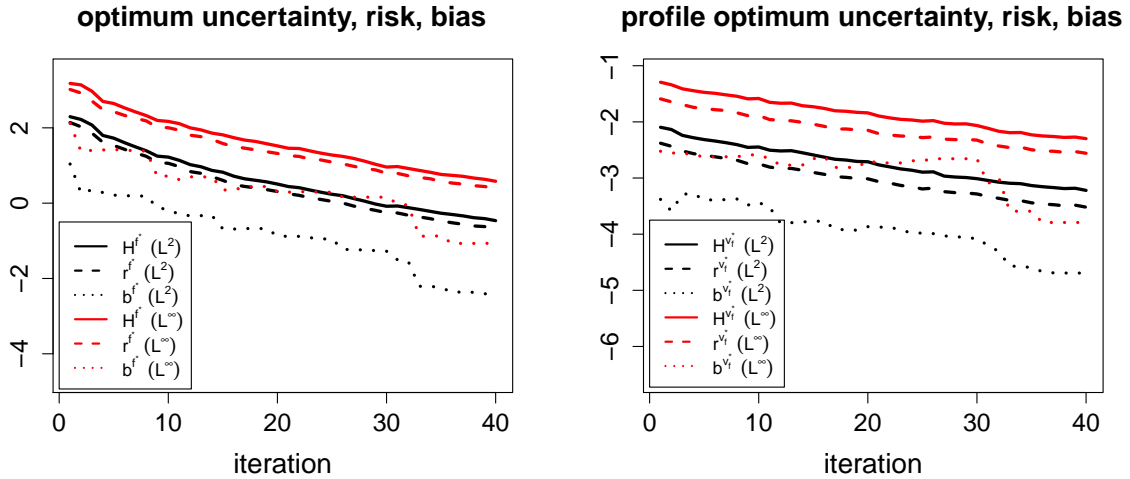


Figure 3.3. Evolution of the uncertainty, risk and bias indicators for the estimated profile optima (left) and optimizers functions (right) throughout the PEI algorithm (log. scale).

where  $M_n^*(x) = M_n^*(\alpha) = \min \left( M_n, \max_{v \in V} \widehat{\xi}_n((\alpha, v)) \right)$  ( $\alpha$  being the decision parameter part of  $x$ ) and  $\gamma$  is taken as in Eq. 3.6. We are now ready to state and prove our convergence result with fixed mean and covariance functions. Let us denote by  $\mathcal{H}$  the reproducing Hilbert space with kernel  $k$ , and write NEB for the No-Empty-Ball defined and discussed in [35] (Definition 3 and Section 4). Theorem 6 of [35] can now be adapted as follows:

**Theorem 3.1.** *Assume that the covariance function has the NEB property. Then, for all  $x_{init} \in \mathbb{X}$  and all  $\omega \in \mathcal{H}$ , the sequence  $(x_n(\omega))_{n \geq 1}$  generated by the algorithm*

$$(3.8) \quad \begin{cases} x_1 = x_{init} \\ x_{n+1} = \arg \max_{x \in \mathbb{X}} \rho_n^*(x) \end{cases}$$

is dense in  $\mathbb{X}$ .

As noticed in [35] (Remark 5) for the EI criterion, even if there is no guarantee that the maximizer of  $\rho_n^*$  is unique, considering algorithms that choose measurably  $x_{n+1}$  among the maximizers of  $\rho_n^*$  would lead to the same result.

Before proving Theorem 3.1, let us enounce and prove a lemma, analogue of the Lemma 12 of [35], that will play a crucial role in the proof of 3.1. Always following [35], let us denote  $\nu_n^* = \sup_{x \in \mathbb{X}} \rho_n^*(x)$  where  $\rho_n^*$  is the profile EI as recalled in Eq. 3.7, and remark that for all  $n \geq 1$ ,  $\nu_n^* = \rho_n^*(x_{n+1}) = \gamma(\widehat{\xi}_n(x_{n+1}) - M_n^*(x_{n+1}); \sigma_n^2(x_{n+1}))$ .

**Lemma 3.2.** *For all  $\omega \in \mathcal{H}$ ,  $\liminf_{n \rightarrow +\infty} \nu_n^*(\omega) = 0$ .*

*Proof.* For any arbitrary  $\omega \in \mathcal{H}$ , set  $s_n = \sigma_n^2(x_{n+1}, \omega)$  and  $z_n^* = \widehat{\xi}_n(x_{n+1}, \omega) - M_n^*(x_{n+1})$ , so that  $\nu_n^*(\omega) = \gamma(z_n^*, s_n)$ . Let  $y^* \in \mathbb{X}$  be an accumulation point of  $(x_n)$  and let  $(x_{\phi_n})$  be any subsequence of  $(x_n)$  converging to  $y^*$ . We are going to prove not that  $\nu_{\phi_n-1} \rightarrow 0$  as in [35] but rather that  $\nu_{\phi_{\psi_n}-1} \rightarrow 0$  for some further extraction (i.e. increasing) function  $\psi : \mathbb{N} \rightarrow \mathbb{N}$ . We know from [35] (Proof of lemma 12) that  $\widehat{\xi}_{\phi_n-1}(x_{\phi_n}, \omega) \rightarrow \omega(y^*)$  and that  $(M_{\phi_n-1}(\omega))$  is a bounded increasing sequence (and so converges) with a limit lower-bounded by  $\omega(y^*)$ . We will now prove that the sequence  $(M_{\phi_n-1}^*(x_{\phi_n})) = (\min(M_{\phi_n-1}, \max_{v \in V} \widehat{\xi}_{\phi_n-1}((\alpha_{\phi_n}, v), \omega)))$  is bounded. As it is clearly upper-bounded by any upper bound of  $(M_{\phi_n-1}(\omega))$ , it suffices to prove that  $(\max_{v \in V} \widehat{\xi}_{\phi_n-1}((\alpha_{\phi_n}, v), \omega))$  is lower-bounded. This last point follows from the fact that  $\max_{v \in V} \widehat{\xi}_{\phi_n-1}((\alpha_{\phi_n}, v)) \geq \widehat{\xi}_{\phi_n-1}(x_{\phi_n}) \rightarrow \omega(y^*)$ .  $(M_{\phi_n-1}^*(x_{\phi_n}))$  being bounded, we can extract from it a converging subsequence  $(M_{\phi_{\psi_n}-1}^*(x_{\phi_{\psi_n}}))$ . Since by continuity,  $\omega(x_{\phi_{\psi_n}-1}) \rightarrow \omega(y^*)$ ,  $M_{\phi_{\psi_n}-1}^*(x_{\phi_{\psi_n}}) \geq \min(\omega(x_{\phi_{\psi_n}-1}), \widehat{\xi}_{\phi_{\psi_n}-1}(x_{\phi_{\psi_n}})) \rightarrow \omega(y^*)$ , and we get that

$$\lim_{n \rightarrow +\infty} z_{\phi_{\psi_n}-1}^* = \lim_{n \rightarrow +\infty} \widehat{\xi}_{\phi_{\psi_n}-1}(x_{\phi_{\psi_n}}, \omega) - \lim_{n \rightarrow +\infty} M_{\phi_{\psi_n}-1}^*(x_{\phi_{\psi_n}}) \leq 0.$$

Finally, noting that  $s_{\phi_{\psi_n}-1} \rightarrow 0$  we conclude in a similar way than [35] that

$$\nu_{\phi_{\psi_n}-1} = \gamma(z_{\phi_{\psi_n}-1}^*, s_{\phi_{\psi_n}-1}) \rightarrow \gamma\left(\lim_{n \rightarrow +\infty} z_{\phi_{\psi_n}-1}^*, 0\right) = 0.$$

which completes the proof of the lemma. ■

*Proof.* [Proof of Theorem 3.1] We follow closely again the proof scheme of [35] (Theorem 6), with a slight adaptation due to specificities of PEI. Let us fix  $\omega \in \mathcal{H}$  and assume that  $(x_n(\omega))$  is not dense in  $\mathbb{X}$ . Then there exists a point  $y^* \in \mathbb{X}$  which is not adherent to  $(x_n(\omega))$ ,



and so by the NEB property  $\inf_{n \geq 1} \sigma_n^2(y^*, \omega) > 0$ . Besides, by the same argument as in [35], the sequence  $(\widehat{\xi}_n(y^*, \omega))$  is bounded. We now use the fact that the specific  $\gamma$  of Eq. 3.6 is a decreasing function of the first variable for any fixed positive value of the second one to get that, for any  $k \geq 1$ ,  $\gamma(\widehat{\xi}_k(y^*, \omega) - M_k^*(y^*, \omega), \sigma_k^2(y^*, \omega)) \geq \gamma(\widehat{\xi}_k(y^*, \omega) - M_k(\omega), \sigma_k^2(y^*, \omega))$  as  $M_k^* \leq M_k(\omega)$  and  $\sigma_k^2(y^*, \omega) > 0$ . We then obtain that

$$\rho_n^*(y^*, \omega) \geq \inf_{k \geq 1} \gamma \left( \widehat{\xi}_k(y^*, \omega) - M_k^*(y^*, \omega), \sigma_k^2(y^*, \omega) \right) \geq \inf_{k \geq 1} \gamma \left( \widehat{\xi}_k(y^*, \omega) - M_k(\omega), \sigma_k^2(y^*, \omega) \right) > 0,$$

which is in contradiction with Lemma 3.2, and hence concludes the proof. ■

## 4. Applications, industrial test case in physics.

### 4.1. Random close packing of granular matter.

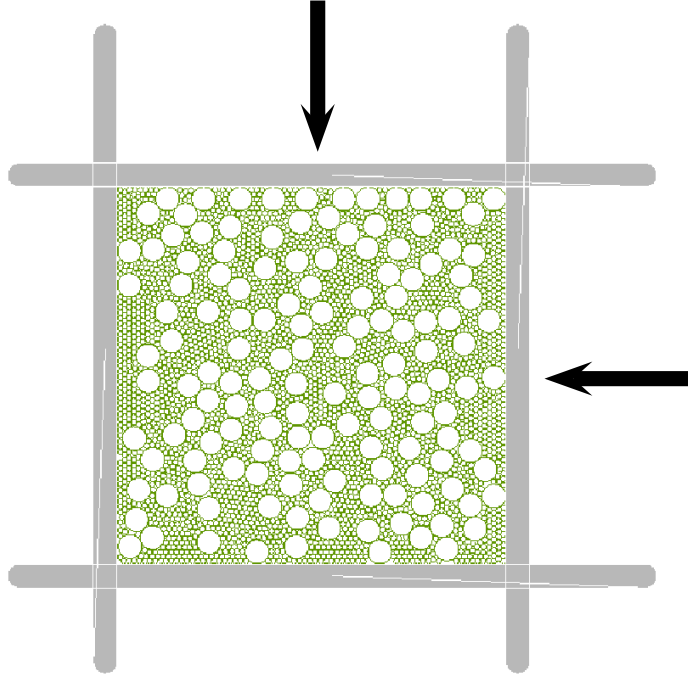
**4.1.1. General context.** Granular materials are large conglomerations of discrete macroscopic particles, a few examples include sand, powder, rocks, cereals or pharmaceutical pills. They play an important role in different industries such as civil engineering, agriculture, pharmaceutical engineering or energy production. In the framework of nuclear safety, the fuel pellets after many years in a nuclear reactor develop many cracks because of thermal stresses. In addition, during an hypothetical nuclear accident, additional fragmentation is postulated to occur because of the thermal - mechanical response to the transient. The fuel pellets stacked into tubes-can are thus considered as close-packed granular clusters.

The complex static and flow properties have been extensively studied during the last decades [10]. In particular, the arrangement of the grains has a significant influence on these properties. The density, which is the volume fraction filled by the particles, permits to characterize the state of these materials.

The test case deals with the random close packing limiting density. The granular material is represented here by a configuration of hard spheres subjected to unilateral contacts. In a monodisperse case, the close packing density in two dimensions is known and is about 0.907 (crystalline arrangements). In the literature, few studies focuses on polydispersity and the impact on mechanical behaviour. For a polydisperse granular media, the close packing density is not clear and in particular depends on the random geometry [9, 30]. It is called the random close packing limiting density and is defined statistically. The probability to obtain an ordered geometrical configuration as local crystalline arrangement is low. So, the close packing can be viewed as a maximum of a geometric problem. This state is numerically obtained by assuming granulars without friction and hydrostatic stress on the sample [21, 1].

In this paper, the previous statistical framework is applied to the study of model random close packing of a bidisperse sample in two dimensions. The contact interactions between the grains are treated using the Non Smooth Contact Dynamics approach [26, 16]. The numerical developments have been performed in the framework of the LMGC90 platform [20]. The main asset of LMGC90 is its large range of contact laws already implemented and its ability to consider more complex surface interactions (as cohesion, wear, etc) without neither regularization nor penalization schemes.

**4.1.2. Description of the test case.** The initial domain is a square randomly filled with 3000 hard disks of radius  $R_1 = 0.1m$  and  $R_2$ . The two input parameters are the radius ratio  $R_1/R_2$  and the density of large disks  $\frac{N_1 R_1^2}{N_1 R_1^2 + N_2 R_2^2}$  where  $N_1$  (resp.  $N_2$ ) is the number of spheres of radius  $R_1$  (resp.  $R_2 \leq R_1$ ). They are respectively denoted  $X$  and  $W$  and their range of variation is respectively  $[4, 10]$  and  $[0.1, 0.9]$ . Moreover, a pressure stress  $P = 10^6 Pa$  is prescribed on the top and the right faces and the left and bottom faces are blocked (see Figure 4.1).

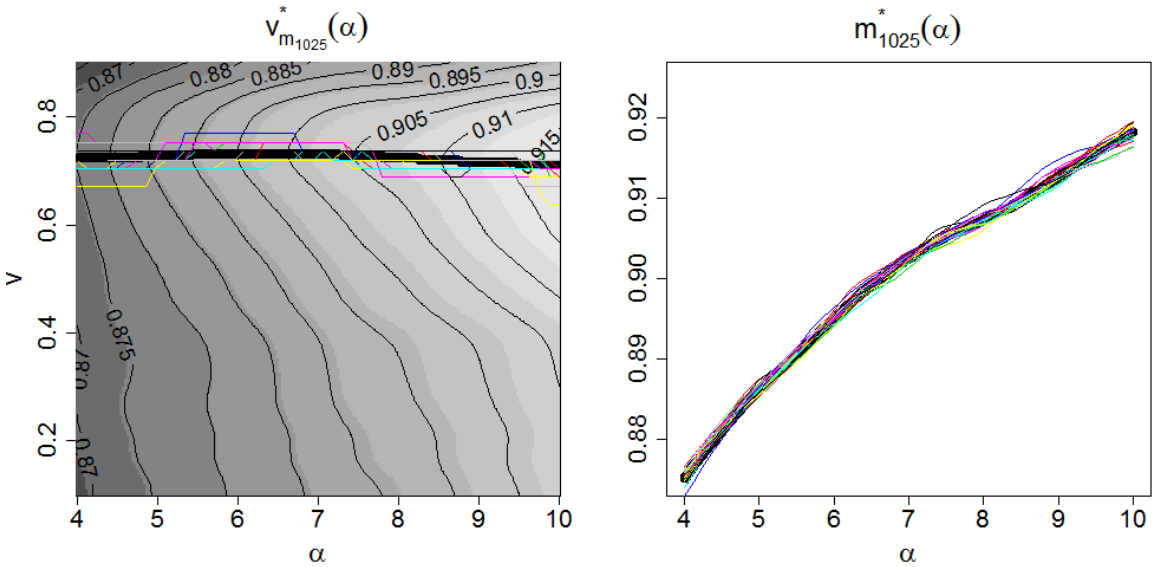


**Figure 4.1.** *Random close packing of 3000 hard disks*

We focus in this study on a topological optimization problem. More precisely, for any given ratio  $X$ , the goal is to exhibit the corresponding density  $W$  leading to the maximum of the code response. Coming back to Expression (1.2), it can be reformulated as a problem of reconstruction of profile optimizers where  $f$  is the maximal packing fraction,  $\alpha := X$  and  $v := W$ . The computational cost of one simulation is about 4 hours. Therefore, for industrial studies, an adaptive design of experiment appears to be an appropriate alternative to classical DoE since it is compulsory to reduce as much as possible the number of simulations while ensuring an accurate reconstruction of the curve of interest. Before considering the DoE construction, a first step in our analysis consists in estimating the map of profile optimizers from a large set of simulations. This will then be used to evaluate the capability of the adaptive DoE obtained by our procedure. All the numerical tests have been performed using the R statistical software and an application programming interface between LMGC90 and R developed in Python language thanks to the rPython R package [12]

**4.2. Estimation of the function of profile optimizers.** The range of variation associated with the input parameters is discretized using a full factorial design with 25 (resp. 41) levels in the  $\alpha$  (resp.  $v$ ) directions. Figure 4.2, left, displays the simulated maximum packing fraction obtained after kriging on a  $200 \times 200$  regular grid from the original data of this design. It points out a strong spatial variability due to the random initial configuration of the hard spheres [9, 30]. From a statistical point of view, the results are non-deterministic and lead to a perturbed curve of interest. Therefore, a post-processing is applied in order to integrate this variability in the data modelling by using kriging with homogeneous noise variance (recalled in Section 2). The noise variance  $\tau^2$  is assumed constant on the whole domain and has been estimated to  $10^{-5}$  using the nugget estimation procedure of the DiceKriging package [31].

The corresponding post-processed maximum packing fraction is depicted on Figure 4.2, right. As expected, the spatial variability in the reconstructed surface has been reduced, leading to a simple curve of profile optimizers that reads  $v_f^*(\alpha) \approx 0.7$  with a very small dependence on the dispersity ratio  $\alpha$ , which is consistent with the literature [11].



**Figure 4.2.** Maximum packing fraction obtained from a 1025-point full factorial design. The surfaces are reconstructed by kriging with homogeneous noise on the original data. 25 functions of profile optimizers  $v_{z_i}^*$  and profile optima  $z_i^*$  of  $s = 25$  conditional realizations  $z_i, 1 \leq i \leq s$  of a Gaussian random field conditioned on 1025 observations of the test function are also represented.

Since the reference map of profile optimizers and profile optimum function are estimated after kriging, they are subject to uncertainty. As an example, Figure 4.2 shows these two quantities obtained from  $s = 25$  realizations  $z_i, 1 \leq i \leq s$  of a Gaussian process conditioned on 1025 observations. Moreover, the third column of Table 4.1 displays the values of the associated uncertainty, risk and bias measures computed from 1000 simulations.

**4.3. Construction and performance evaluation of the adaptive DoE based on Profile-EI.** We now want to compare the performances of our criterion based on the PEI with other evaluation strategies like the Integrated Mean Squared Error (IMSE) minimization (see, e.g., [29]), the classical Expected Improvement maximization, or pure uniform random sampling.

A total of 100 random initial design of 20 points are used. They are generated with the optimumLHS function of the LHS R package. Each strategies is tested on these 100 designs and the results are averaged. We evaluate sequentially a total of 40 additional points and compare the results in terms of uncertainty, risk and bias. To do so, we assume that the reference map of profile optimizers and profile optimum function are the ones obtained after having evaluated the function on the full design of 1025 points. An ordinary kriging model is always used, with a separable Matérn(3/2) covariance kernel. The parameters of this covariance kernel are estimated, plugged-in and renewed at each iteration using maximum likelihood estimates.

**Table 4.1**

*Comparison of the decrease of the model uncertainty, risk and bias obtained from different sequential evaluation strategies.*

norm	indicator	full design (1025 points)	initial designs (20 points)	random	IMSE	EI	PEI
$L^2$	$H^{f^*}$	0.18	0.80	0.56	0.52	0.79	<b>0.39</b>
$L^2$	$r^{f^*}$	0.13	0.93	0.56	0.54	0.90	<b>0.43</b>
$L^2$	$b^{f^*}$	0	0.78	0.40	0.41	0.61	<b>0.32</b>
$L^2$	$H^{v_f^*}$	4.4	35	20	21	32	<b>18</b>
$L^2$	$r^{v_f^*}$	3.1	28	16	17	26	<b>15</b>
$L^2$	$b^{v_f^*}$	0	15.6	7.3	<b>6.8</b>	17.4	7.6
$L^\infty$	$H^{f^*}$	0.16	0.70	0.50	0.46	0.69	<b>0.33</b>
$L^\infty$	$r^{f^*}$	0.11	0.86	0.54	0.49	0.91	<b>0.36</b>
$L^\infty$	$b^{f^*}$	0	0.67	0.38	0.36	0.57	<b>0.27</b>
$L^\infty$	$H^{v_f^*}$	3.3	35	25	26	35	<b>23</b>
$L^\infty$	$r^{v_f^*}$	2.4	25	17	19	24	<b>16</b>
$L^\infty$	$b^{v_f^*}$	0	13.5	7.5	6.7	17.2	<b>4.5</b>

Many conclusions can be drawn from the results given on Table 4.1. It appears that the evaluation strategy based on the PEI maximization provides good performances for both estimating the profile optimizer and profile optimum functions. In particular, PEI clearly outperforms the other strategies for the estimation of  $f^*$ . However, although the performances of PEI remain good in terms of estimation of  $v_f^*$ , its domination over the are other criteria is less marked than for  $f^*$ . Similarly to the Expected Improvement - which is more suited to find the maximum of a function than the maximizer (see, e.g., [36]) - the PEI is more adapted to find the profile optimum function  $f^*$  than the profile optimizer. Another important remark is that the covariance parameters of our Matérn obtained after the 40 sequential evaluations are not the same for all strategies as they are always re-estimated from the new observations. This might introduce a bias for our measure of uncertainty, if, e.g. one strategy tends to evaluate locations leading to a higher range estimate. For completeness, we also provide in Appendix a replicate of Table 4.1 where the covariance parameters are estimated from the initial design of experiment of 20 points and not re-estimated afterwards (see, Table A.1). In this setup, it appears that the performances of all the tested strategies (except the random

sampling) are slightly worse than previously. However, this does not change the ranking of the performances and our latter conclusions are unchanged.

**Conclusion.** We proposed a novel approach for estimating and quantifying uncertainty on profile optima and optimizers relying on Gaussian random field models. The proposed approach gave very convincing results on the considered two-dimensional test cases, as it allowed to efficiently reconstruct the curves of interest in reasonable evaluation budget, and could additionally quantify the reconstruction error in a sensible way through conditional simulations. Of course, the relevance of the error quantification relies on the choice and the estimation of an adequate Gaussian field model. Bootstrapping on covariance parameters [8] or a full Bayesian approach [3] are possible enhancements, easily adaptable to the presented methods, for mitigating the risk of model misspecification. Note however that for the final test case, re-estimating the covariance parameters did not lead to significant differences (See Section A), so that basing parameter estimation solely on initial evaluations appeared to be enough for a successful model building.

An infill sampling criterion for sequentially learning the profile optima function was introduced, namely the *Profile Expected Improvement* (PEI) criterion, that generalizes the popular *Expected Improvement* [24, 18] criterion to our specific set-up. PEI partially inherits the analytical tractability from EI (up to the calculation of the kriging mean’s profile optimum), and also its consistency properties, as proven in Section 3. Besides, experimental results obtained on both a toy function from the literature and on a realistic physics case study illustrate that PEI clearly outperforms standard criteria such as IMSE or EI for the specific task of learning the curve of profile optima. A good surprise is that PEI performs well also for learning the curve of profile optimizers, even though it is not directly conceived for it. Significant improvements for learning profile optimizer function might be obtained, e.g. by applying the principles of *Sequential Uncertainty Reduction* (SUR) strategies [2] to an uncertainty measure directly defined on the notion of profile optimizer. From a practical perspective, PEI algorithms may also be developed further by proposing parallelization schemes, e.g. by simultaneously evaluating the objective function for different values of the nuisance variable at each iteration. Finally, tackling further case studies in higher dimensions (such as the polydisperse sphere problem with a higher number of different radii) will probably lead to new computational challenges, be it in terms of conditional simulations or of internal optimizations needed to calculate curves of profile optima associated with kriging means.

**Acknowledgements:** This work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Ecole des Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments. The authors would like to thank Tarek Mokhtari for his contribution to preliminary numerical experiments.

**Appendix A. Evaluation strategies comparison with no covariance parameters re-estimation.**

## REFERENCES

**Table A.1**

Comparison of the decrease of the model uncertainty, risk and bias obtained from different sequential evaluation strategies. Here, in each of the 100 runs, the covariance parameters are estimated from the initial design of experiments of 20 points and not re-estimated afterwards.

norm	indicator	full design (1025 points)	initial designs (20 points)	random	IMSE	EI	PEI
$L^2$	$H^{f^*}$	0.18	0.80	0.56	0.52	0.69	<b>0.39</b>
$L^2$	$r^{f^*}$	0.13	0.93	0.61	0.58	0.79	<b>0.45</b>
$L^2$	$b^{f^*}$	0	0.78	0.46	0.47	0.58	<b>0.36</b>
$L^2$	$H^{v_f^*}$	4.4	35	21	22	25	<b>19</b>
$L^2$	$r^{v_f^*}$	3.1	28	17	17	23	<b>17</b>
$L^2$	$b^{v_f^*}$	0	15.6	8.7	<b>6.9</b>	14.2	9.0
$L^\infty$	$H^{f^*}$	0.16	0.70	0.51	0.46	0.60	<b>0.34</b>
$L^\infty$	$r^{f^*}$	0.11	0.86	0.57	0.53	0.75	<b>0.38</b>
$L^\infty$	$b^{f^*}$	0	0.67	0.41	0.41	0.54	<b>0.29</b>
$L^\infty$	$H^{v_f^*}$	3.3	35	25	28	29	<b>24</b>
$L^\infty$	$r^{v_f^*}$	2.4	25	18	20	21	<b>18</b>
$L^\infty$	$b^{v_f^*}$	0	13.5	8.6	6.0	12.0	<b>5.6</b>

- [1] I. AGNOLIN AND J.N. ROUX, *Internal states of model isotropic granular packings.*, Physical Review E, 76 (2007).
- [2] J. BECT, D. GINSBOURGER, L. LI, V. PICHENY, AND E. VAZQUEZ, *Sequential design of computer experiments for the estimation of a probability of failure*, Statistics and Computing, 22(3) (2012), pp. 773–793.
- [3] R. BENASSI, *Nouvel algorithme d'optimisation bayésien utilisant une approche Monte-Carlo séquentielle*, PhD thesis, Supélec, 2013.
- [4] R. CARNELL, *lhs: Latin Hypercube Samples*, 2009. R package version 0.5.
- [5] C. CHEVALIER, *Fast uncertainty reduction strategies relying on Gaussian process models*, PhD thesis, University of Bern, 2013.
- [6] J.-P. CHILÈS AND P. DELFINER, *Geostatistics: Modeling Spatial Uncertainty*, Wiley, New York, second edition ed., 2012.
- [7] N. CRESSIE, *Statistics for spatial data*, 1993.
- [8] D. DEN HERTOOG, J. P. C. KLEIJNEN, AND A. Y. D. SIEM, *The correct kriging variance estimated by bootstrapping*, The Journal of the Operational Research Society, 57 (4) (2006), pp. 400–409.
- [9] M. DIJKSTRA, R. VON ROIJ, AND R. EVANS, *Phase diagram of highly asymmetric binary hard-sphere mixtures*, Physical Review E, 59 (1999), pp. 5744–5770.
- [10] J. DURAN, *Sands, Powders, and Grains: An Introduction To the Physics of Granular Materials*, Springer-Verlag New York Inc., 1999.
- [11] R.S. FARR AND R.D. GROTT, *Close packing density of polydisperse hard spheres*, Journal of Chemical Physics, 131, 244104 (2009).
- [12] C.J. GIL BELLOSTA, *rPython: Package allowing R to call Python*, 2013. R package version 0.0-4.
- [13] D. GINSBOURGER AND R. LE RICHE, *Towards gaussian process-based optimization with finite time horizon*, in mODa 9 — Advances in Model-Oriented Design and Analysis, Contributions to Statistics, Springer, 2010, pp. 89–96.
- [14] D. GINSBOURGER, R. LE RICHE, AND CARRARO L., *Kriging is well-suited to parallelize optimization*, in Computational Intelligence in Expensive Optimization Problems, Springer, 2010.
- [15] T. HASTIE, R. TIBSHIRANI, AND J. FRIEDMAN, *The Elements of Statistical Learning*, Springer, 2001.
- [16] M. JEAN, *The non-smooth contact dynamics method*, Computer methods in applied mechanics and engi-

- neering, 177 (1999), pp. 235–257.
- [17] D. R. JONES, *A taxonomy of global optimization methods based on response surfaces*, Journal of Global Optimization, 21 (2001), pp. 345–383.
- [18] D. R. JONES, M. SCHONLAU, AND J. WILLIAM, *Efficient global optimization of expensive black-box functions*, Journal of Global Optimization, 13 (1998), pp. 455–492.
- [19] J. KIM AND D. POLLARD, *Cube root asymptotics*, The Annals of Statistics, 18 (1) (1990), pp. 191–219.
- [20] LMGC LABORATORY, *Lmgc90*. <http://www.lmgc.univ-montp2.fr/LMGC90>.
- [21] H. A. MAKSE, D. L. JOHNSON, AND M. SCHWARTZ, *Packing of compressible granular materials*, Physical Review Letters, 84 (2000), pp. 4160–4163.
- [22] G. MATHERON, *The intrinsic random functions, and their applications*, Advances in Applied Probability, 5 (1973), pp. 439–468.
- [23] W. MEBANE AND J. SEKHON, *Genetic optimization using derivatives: The rgenoud package for r*, Journal of Statistical Software, Vol. 42, Issue 11 (2011), pp. 1–26.
- [24] J. MOCKUS, *Bayesian Approach to Global Optimization. Theory and Applications*, Kluwer Academic Publisher, Dordrecht, 1989.
- [25] I. MOLCHANOV, *Theory of Random Sets*, Springer, 2005.
- [26] J.J. MOREAU, *Unilateral contact and dry friction in finite freedom dyanmics*, in Non Smooth Mechanics and Applications, CISM Courses and Lectures, vol. 302, 1988, pp. 1–82.
- [27] H. OMRE AND K.B. HALVORSEN, *The bayesian bridge between simple and universal kriging*, Mathematical Geology, 22 (7) (1989), pp. 767–786.
- [28] V. PICHENY AND D. GINSBOURGER, *Noisy kriging-based optimization methods: a unified implementation within the DiceOptim package*, Computational Statistics and Data Analysis, 71 (2014), pp. 1035–1053.
- [29] V. PICHENY, D. GINSBOURGER, O. ROUSTANT, R. T. HAFTKA, AND N.-H. KIM, *Adaptive designs of experiments for accurate approximation of target regions*, Journal of Mechanical Design, 132 (2010).
- [30] P. M. REIS AND T. MULLIN, *Granular segregation as a critical phenomenon*, Physical Review Letters, 89 (2002).
- [31] O. ROUSTANT, D. GINSBOURGER, AND Y. DEVILLE, *DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodelling and optimization*, Journal of Statistical Software, 51 (2012).
- [32] T. J. SANTNER, B. J. WILLIAMS, AND W. NOTZ, *The Design and Analysis of Computer Experiments*, Springer Verlag, 2003.
- [33] T. SEVERINI AND W.H. WONG, *Profile likelihood and conditionally parametric models*, Annals of Statistics, 20 (4) (1992), pp. 1768–1802.
- [34] M. L. STEIN, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York, 1999.
- [35] E. VAZQUEZ AND J. BECT, *Convergence properties of the experted improvement algorithm with fixed mean and covariance functions*, Journal of Statistical Planning and Inference, 140 (11) (2010), pp. 3088–3095.
- [36] J. VILLEMONTAIX, *Optimisation de fonctions coûteuses*, PhD thesis, Université Paris-Sud XI, Faculté des Sciences d’Orsay, 2008.
- [37] J. VILLEMONTAIX, E. VAZQUEZ, AND E. WALTER, *An informational approach to the global optimization of expensive-to-evaluate functions*, Journal of Global Optimization, 44 (2009), pp. 509–534.