Bayesian Adaptive Reconstruction of Profile Optima and Optimizers

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

SIAM/ASA J. Uncertainty Quantification
Vol. 2, pp. 490–510

Bayesian Adaptive Reconstruction of Profile Optima and Optimizers*

David Ginsbourger†, Jean Baccou‡, Clément Chevalier†, Frédéric Perales‡, Nicolas Garland‡, and Yann Monerie§

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 are based 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 random field conditional simulations that allow a quantification of uncertainties in the Bayesian framework. Besides this, we introduce a variant of the expected improvement criterion, which proves efficient for adaptively learning the set of profile optima and optimizers. The results are illustrated with 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 simulation, active learning, expected improvement

AMS subject classifications. 62K20, 62L05, 65C20, 65K10, 90C56

DOI. 10.1137/130949555

1. Introduction.

A number of scientific and industrial questions boil down to optimization problems: Physical equilibria are found by minimizing appropriate 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 $f : \mathbf{x} \in \mathcal{X} \longrightarrow f(\mathbf{x}) \in \mathbb{R}$ with domain $\mathcal{X} \subset E$, where typically $E = \mathbb{R}^d$ for some integer $d \geq 1$. Whenever f reaches a *global maximum*, $f^\star = \max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$, values of $\mathbf{x} \in \mathcal{X}$ such that $f(\mathbf{x}) = f^\star$ 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 a case, one can view the global maximum of f with respect to the nuisance variables v as a function of α:

$$f^\star : \alpha \in A \longrightarrow f^\star(\alpha) = \max_{v\in V} f(\alpha, v) \in \mathbb{R}, \tag{1.1}$$

which we refer to as a *profile optimum* or *conditional optimum* function. Assuming further that $f(\alpha, \cdot)$ possesses a unique point of global maximum $v^\star_f(\alpha)$ for any given $\alpha \in A$ (or defining

*Received by the editors December 17, 2013; accepted for publication (in revised form) June 25, 2014; published electronically September 23, 2014.
http://www.siam.org/journals/juq/2/94955.html

†IMSV, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland (david.ginsbourger@stat.unibe.ch, clement.chevalier@stat.unibe.ch), and the ReDICE Consortium (http://www.redice-project.org/).
‡Institut de Radioprotection et de Sûreté Nucléaire, PSN-RES, SEMIA, Centre de Cadarache, 13115 Saint-Paul-lès-Durance, France (jean.baccou@irsn.fr, frederic.perales@irsn.fr, nicolas-garland@club-internet.fr).
§Laboratoire de Mécanique et Génie Civil (LMGC - UMR 5508), Université Montpellier 2, 34095 Montpellier cedex 5, France (yann.monerie@univ-montp2.fr).

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*:

$$v_f^\star : \alpha \in A \longrightarrow v_f^\star(\alpha) \in V. \tag{1.2}$$

Our main aim in the present work is to reconstruct $f^\star$ and $v_f^\star$ under a drastically limited evaluation budget of the objective function f. An example of profile optimum and profile optimizer functions is shown in Figure 1, where the objective function f is the negative rescaled Branin–Hoo function (see [21] and (2.8)).

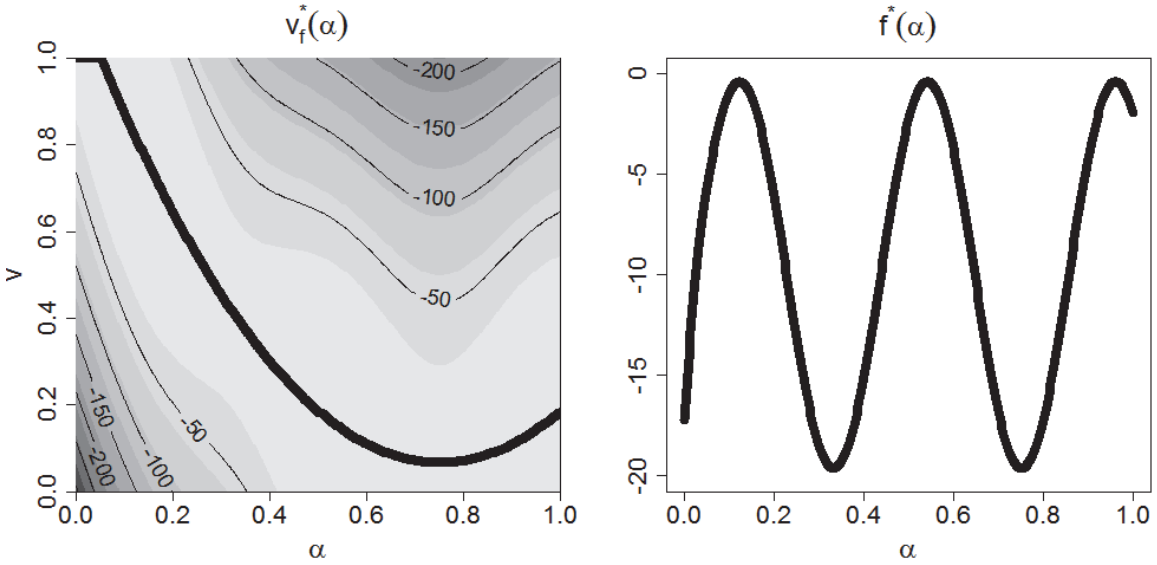

Figure 1. *Left: Profile optimizer $v_f^\star$ (black curve) and contour lines of the negative rescaled Branin–Hoo function f defined in section 2 relying on (2.8). Right: Corresponding profile optimum function $f^\star$.*

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

- In reliability engineering, the vulnerability of a system may depend both on controllable design variables α like materials and geometry, and also on uncontrollable variables v like future climatic conditions (wind characteristics, temperature trends, etc.). In such a 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 α 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 α. A neighboring problem referred to as robust inversion has been tackled in [8], involving nuclear criticality safety applications.
- In statistical modeling, it often arises that some likelihood function may be analytically maximized as a function of a subgroup of parameters, the other model parameters being considered fixed. One then speaks of *profile likelihood* [36]. For example, this happens in the case of a centered Gaussian vector with stationary covariance depending on a variance v (or its square root, often named scale) and on some correlation

parameter α. It is then well known that the maximum likelihood estimator of v can be derived analytically as a function of α. Plotting the optimal v as a function of α 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 ($v_f^\star$), seen as a function of the ratio of radii (α).

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 $f^\star$ and $v_f^\star$ functions. Here we concentrate on the use of kriging and Gaussian random field models for estimating them, quantifying the uncertainties associated with these estimations and reducing those uncertainties by appealing to adaptive evaluation strategies in the sequential Bayesian framework.

This paper 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, which 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 theorem 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.

Preliminary to the main contributions of the paper, about adaptive experimental design, we first set the notation and present the proposed profile reconstruction and associated uncertainty quantification methods in the static framework. Let us now briefly discuss a crude "plug-in" approach and then review the Bayesian approach under a Gaussian random field prior.

2.1. Plug-in approach: Estimating $f^\star$ and $v_f^\star$ based on an approximation of f.

We now assume that the function f was already evaluated at a set of points, or *experimental design*,

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

Reconstructing $f^\star$ and $v_f^\star$ directly from such data is not straightforward. A natural idea, provided some prior knowledge about the function f is available, is to use an interpolator or an approximation model for f and to reconstruct the profile maps based on this model. A vast set of interpolation and approximation approaches may be envisaged [18], encompassing polynomials, methods based on Fourier and wavelet analysis, splines, neural nets, and kernel methods. Here we mainly focus on kriging [25, 10, 37, 9], 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

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

where $\mathbf{u}'$ denotes the transpose of the vector $\mathbf{u}$, the so-called ordinary kriging (OK) predictor is written as

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

where $\widehat{\mu} = (\mathbf{1}'K^{-1}\mathbf{1})^{-1}\mathbf{1}'K^{-1}\mathbf{z}$, $\mathbf{k}(\mathbf{x}) = (k(\mathbf{x}_1, \mathbf{x}), \ldots, k(\mathbf{x}_n, \mathbf{x}))'$, $K = (k(\mathbf{x}_i, \mathbf{x}_j))_{(i,j)\in\{1,\ldots,n\}^2}$, $\mathbf{1} = (1, \ldots, 1)'$, and k is a positive definite kernel 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:

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

Note that in the case where the observations are corrupted by Gaussian noise with mean 0 and variance τ^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 $f^\star$ and $v_f^\star$,

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

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

$$m(\alpha; v_m^\star(\alpha)) \geq m(\alpha, v) \quad (\alpha \in A, v \in V). \tag{2.6}$$

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 $f^\star$ by $m^\star$, it is straightforward that

$$||f^\star - m^\star||_\infty \leq ||f - m||_\infty, \tag{2.7}$$

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 does not hold, however, for the profile maximizers: whatever $||f - m||_\infty$ may be, it is always possible to have the maximizers $v_f^\star(\alpha)$ and $v_m^\star(\alpha)$ at any given distance within the domain V.

In Figure 2, we give some first illustrations based on the negative rescaled Branin–Hoo function f, defined over $[0, 1]^2$ by $f(x_1, x_2) = -f_{\mathrm{BH}}(15x_1 - 5, 15x_2)$, where

$$f_{\mathrm{BH}} : (x_1, x_2) \in [-5, 10] \times [0, 15] \longrightarrow a(x_2 - bx_1^2 + cx_1 - r) + s(1 - t)\cos(x_1) + s, \tag{2.8}$$

with $a = 1, b = 5/(4\pi^2), c = 5/\pi, r = 6, s = 10$, and $t = 1/(8\pi)$.

Here two different experimental designs 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* [7] with "optimum" option and default settings. In both cases, an anisotropic Matérn kernel with

smoothness parameter fixed to $\nu = 3/2$ is chosen for k (See Table 1 in [34], freely available online) and the corresponding scale and range hyperparameters are estimated by maximum likelihood using the *DiceKriging* R package [34]. One can see in Figure 2 that $f^\star$ and $v_f^\star$ are better reconstructed with the 20-point design, as expected. More specifically, Table 1 gives the bias obtained on the profile optimum and profile optimizer functions measured as follows:

$$b^{v_f^\star} := \|v_m^\star - v_f^\star\| \,, \tag{2.9}$$

$$b^{f^\star} := \|m^\star - f^\star\| \,, \tag{2.10}$$

where the considered norms are either L^2 or L^∞ on appropriate spaces.

In applications the use of the bias as a performance indicator is unrealistic, as the true $v_f^\star$ and $f^\star$ 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.

2.2. Bayesian approach under a Gaussian random field prior. An alternative approach proposed here, which complements the prediction by kriging presented in the last section, consists in assuming a Gaussian random field prior for the function f and then relying on the obtained posterior predictive distribution to draw conclusions about the profile optima and optimizer functions. In such a framework, f is seen as one possible realization (i.e., $Z.(\omega)$ for some $\omega \in \Omega$, Ω being the underlying sample space) of a Gaussian random field $(Z_{\mathbf{x}})_{\mathbf{x}\in\mathcal{X}}$ with mean μ 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 itself be a centered random field with an analytically tractable covariance. Under well-known technical conditions [30], e.g., a uniform improper prior on μ, it turns out that the conditional expectation coincides with the OK predictor of (2.3):

$$\mathbb{E}[Z_{\mathbf{x}}|Z_{\mathbf{X}} = \mathbf{z}] = m(\mathbf{x}), \tag{2.11}$$

while for any pair of points $(\mathbf{x}, \mathbf{u}) \in \mathcal{X}^2$ the covariance of the prediction error is

$$\mathrm{Cov}[Z_{\mathbf{x}} - \widehat{Z}_{\mathbf{x}}, Z_{\mathbf{u}} - \widehat{Z}_{\mathbf{u}}] = k(\mathbf{x},\mathbf{u}) - \mathbf{k}(\mathbf{x})' K^{-1} \mathbf{k}(\mathbf{u}) + \frac{(1 - \mathbf{k}(\mathbf{x})' K^{-1}\mathbf{1})(1 - \mathbf{k}(\mathbf{u})' K^{-1}\mathbf{1})}{\mathbf{1}' K^{-1}\mathbf{1}}. \tag{2.12}$$

Using this result, one may simulate Gaussian random field realizations that honor the data while giving a quantification of uncertainty for various quantities regarding f. Here we use that workflow to enhance the estimation of the $f^\star$ and $v_f^\star$ functions and also to 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

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

In other words, the $f^\star$ and $v_f^\star$ functions are represented by the random process $Z^\star$ and the random field[1] $v_Z^\star$, themselves being approximated by the sets of functions $\{z_1^\star, \ldots, z_s^\star\}$ and $\{v_{z_1}^\star, \ldots, v_{z_s}^\star\}$. An example of such sets of functions, obtained from conditional simulations,

[1] See [22] for a discussion on the almost sure unicity of the optimizer for suitable Gaussian random field models.

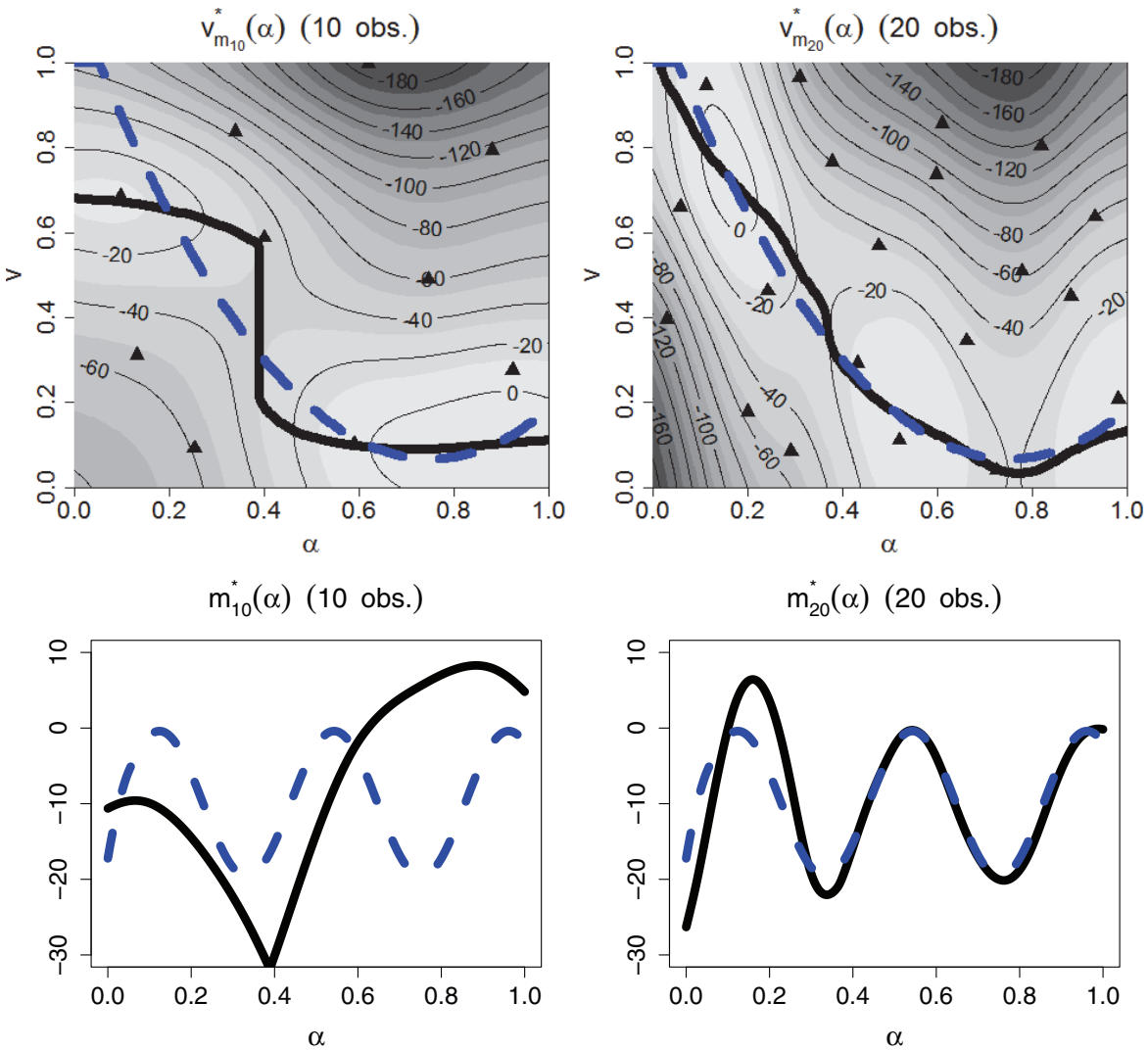

Figure 2. *Plug-in estimates $v^\star_m$ (top) and $m^\star$ (bottom) obtained from 10-point (left) and 20-point (right) experimental designs. The estimates are thick black curves. The blue dashed curves represent the actual $v^\star_f$ and $f^\star$ functions. The triangles stand for the design points.*

Table 1

Model bias if 10 or 20 locations are evaluated.

Norm	Indicator	10 obs.	20 obs.
L^2	$b^{f^\star}$	12.97	3.94
L^2	$b^{v^\star_f}$	0.13	0.04
L^∞	$b^{f^\star}$	25.46	10.17
L^∞	$b^{v^\star_f}$	0.32	0.09

is shown in Figure 3. Here $s = 25$ conditional simulations are computed on a 50×50 grid, leading to 25 functions $z^\star_i$ and $v^\star_{z_i}$, $1 \le i \le s$. Clearly, the functions $z^\star_i$ and $v^\star_{z_i}$ have a

higher dispersion when they are obtained from simulations conditioned on 10 observations (left plots). This suggests using this dispersion as an indicator to quantify the uncertainty, given the observations, on the true unknown functions $f^\star$ and $v_f^\star$.

How to quantify the variability associated with a set of curves remains an open question, to which various answers have been detailed throughout the literature of probability theory [28]. In the present work, we quantify this variability using the expectation of the distance between two (independently) simulated profile optimizers $v^\star_{Z_1}$, $v^\star_{Z_2}$ or profile optimum functions $Z_1^\star$, $Z_2^\star$:

$$H^{v_f^\star} := \mathbb{E}\left[\|v^\star_{Z_1} - v^\star_{Z_2}\| | Z_{\mathbf{X}} = \mathbf{z}\right], \tag{2.13}$$

$$H^{f^\star} := \mathbb{E}\left[\|Z_1^\star - Z_2^\star\| | Z_{\mathbf{X}} = \mathbf{z}\right], \tag{2.14}$$

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

$$r^{v_f^\star} := \mathbb{E}\left[\|v^\star_{Z_1} - v_f^\star\| | Z_{\mathbf{X}} = \mathbf{z}\right], \tag{2.15}$$

$$r^{f^\star} := \mathbb{E}\left[\|Z_1^\star - f^\star\| | Z_{\mathbf{X}} = \mathbf{z}\right], \tag{2.16}$$

which require us to know $f^\star$ and $v_f^\star$, respectively. Coming back to the Branin–Hoo toy example, estimates of these indicators relying on 1000 conditional simulations are displayed in Table 2. It appears that the variability and the risk are lower for 20-point design than for the 10-point design, as could be anticipated. 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*.

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

3.1. Problem setup: How to choose the next point. We previously observed in a toy example that the quality of the profile function reconstructions increased with the size of the experimental design, 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 experimental designs 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, as in [21]) or in a batch-sequential manner [17].

Several approaches have been explored for efficient 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 [35]. Variants of IMSE criteria have been proposed, notably for learning the contour lines of a function corresponding to a prescribed target level [32, 6]. Another question that has inspired a number of sequential

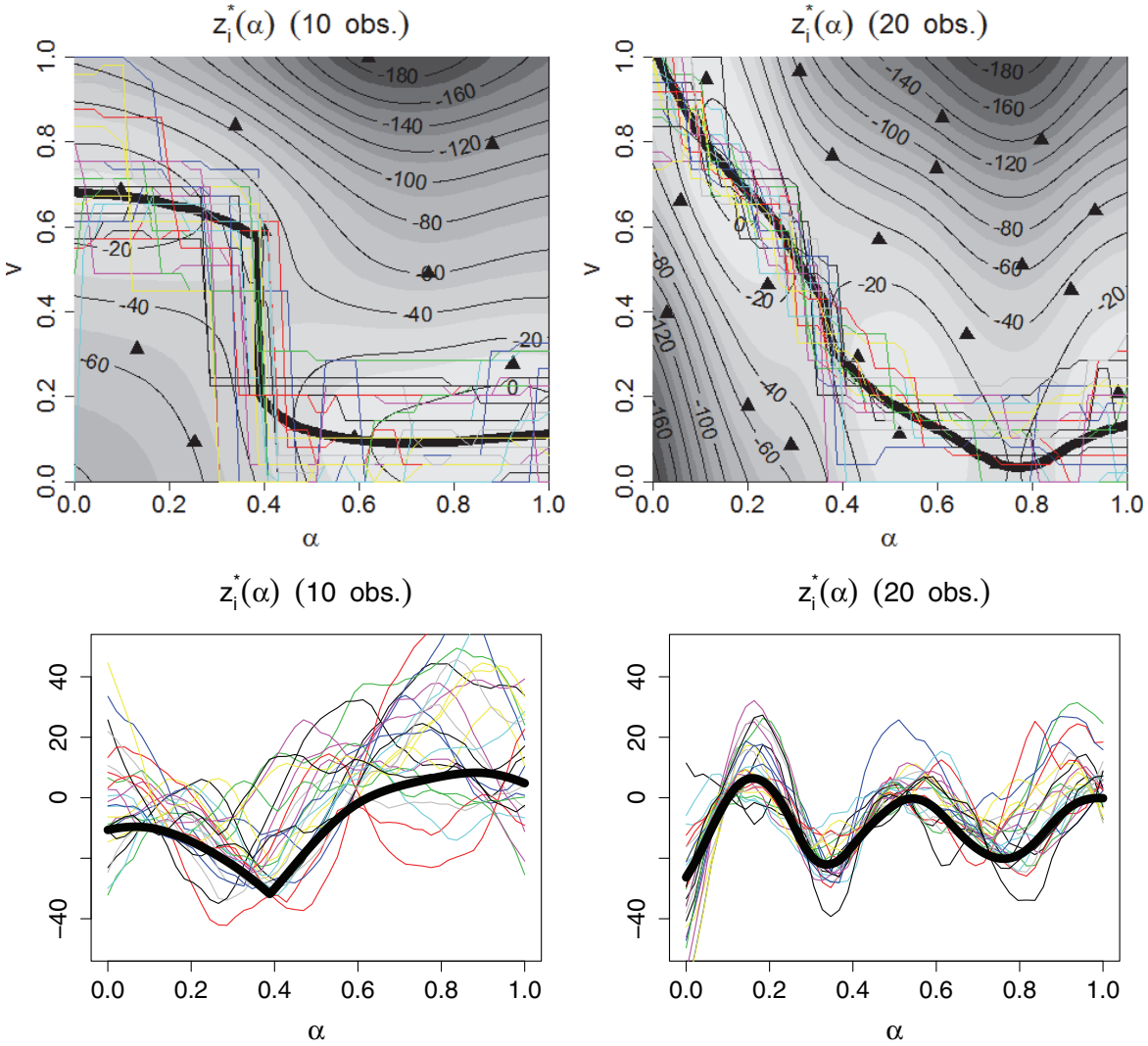

Figure 3. *Profile optimizers $v^\star_{z_i}$ and profile optima functions $z_i^\star$ of $s = 25$ realizations $z_i, 1 \le i \le s$, of a Gaussian random field conditioned on 10 (left) and 20 (right) observations of the test function f (see (2.8)).*

Table 2

Risk and uncertainty indicators for the profile optima and optimizer functions in terms of L^2 and L^∞ norms if the test function f is evaluated at 10 or 20 design points. The setup is the same as for Figure 3.

Norm	Indicator	10 obs.	20 obs.
L^2	$H^{f^\star}$	20.49	11.60
L^2	$r^{f^\star}$	22.32	10.17
L^2	$H^{v_f^\star}$	0.18	0.11
L^2	$r^{v_f^\star}$	0.17	0.09
L^∞	$H^{f^\star}$	44.27	31.03
L^∞	$r^{f^\star}$	45.84	26.79
L^∞	$H^{v_f^\star}$	0.49	0.27
L^∞	$r^{v_f^\star}$	0.40	0.21

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 [20], including the *probability of improvement* (PI), the *expected improvement* (EI), and the expected conditional entropy of the maximizer [40]. Due to its conceptual simplicity and its good empirical performances, EI has become very popular in the past 15 years for solving medium-dimensional derivative-free nonconvex global optimization problems under a severely limited budget. Even though it is known that the optimality properties of EI are only in the *one step look-ahead* framework [27, 16], there exist some theoretical guarantees on the convergence of EI algorithms given a few restrictions on the family of kernels used [38]. Several variants of EI for constrained or batch-sequential problems exist, and both the criterion and its maximization are coded in open source programs [34, 31].

3.2. Profile EI: A variant of the EI 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 standard EI settings the variable $\mathbf{x}$ does not have to be split into two subgroups (v and α); everything happens as if there were no parameter α so the domain of $\mathbf{x}$ reduces to V. The EI criterion at an arbitrary point $\mathbf{x}$ is defined as average departure of $f(\mathbf{x})$ above the *current maximum* $\max_{1\le i\le n} f(\mathbf{x}_i)$ under the Gaussian field model:

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

A very practical aspect about the EI criterion defined in (3.1) is that an 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:

$$EI(\mathbf{x}) = \begin{cases} \sigma(\mathbf{x})\left(u(\mathbf{x})\Phi(u(\mathbf{x})) + \phi(u(\mathbf{x}))\right) & \text{if } \sigma^2(\mathbf{x}) \neq 0 \\ 0 & \text{else,} \end{cases} \tag{3.2}$$

where $u(\mathbf{x}) = (m(\mathbf{x}) - \max_{1\le i\le n} f(\mathbf{x}_i))/\sigma(\mathbf{x})$ and $\sigma^2(\mathbf{x})$ is the OK prediction variance (given by (2.12) for $\mathbf{u} = \mathbf{x}$). Φ and ϕ denote, respectively, the cumulative distribution function and probability density function of the standard Gaussian distribution.

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

$$PEI(\alpha, v) := \mathbb{E}\left[\left(Z_{(\alpha,v)} - T(\alpha)\right)^+ \middle| Z_{\mathbf{X}} = \mathbf{z}\right], \tag{3.3}$$

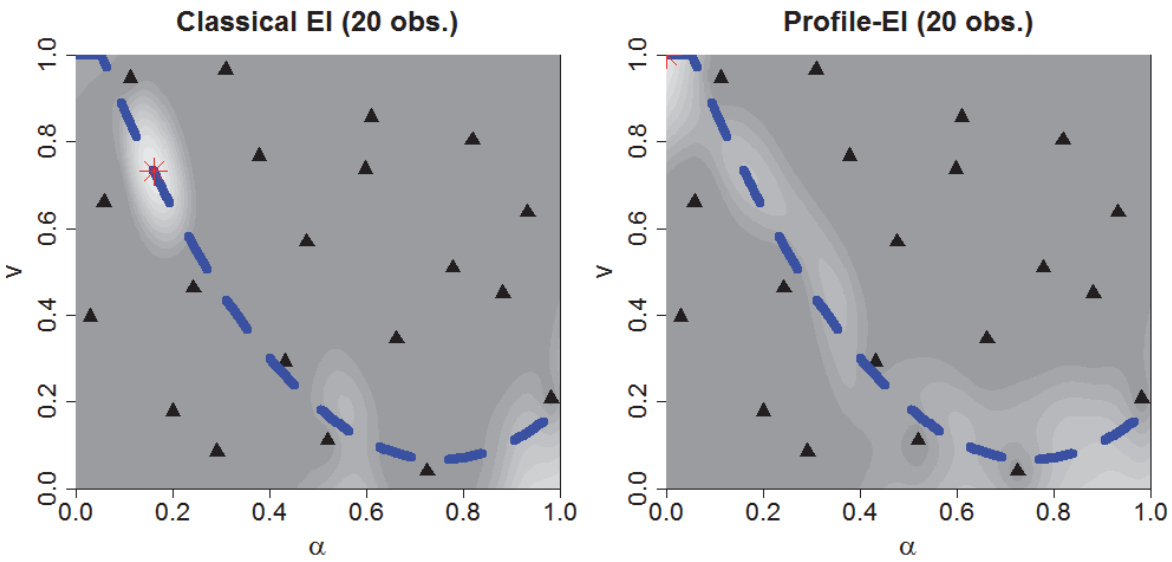

Figure 4. *$EI(\alpha, v)$ and $PEI(\alpha, v)$ functions obtained from a 20-point initial design of experiments (black triangles). The blue dashed curves represent the actual $v_f^\star$ function. EI and PEI are represented through grayscale contour lines (with colors ranging from dark to light gray with increasing criterion value). The red stars (see the top left corner for the right graph) stand for the points of global maximum of the criteria.*

where the target is defined by

$$T(\alpha) := \min\left(\max_{w\in V} m(\alpha, w), \max_{1\le i\le n} Z_{(\alpha_i, v_i)}\right). \tag{3.4}$$

Let us remark that the "cap" at the current maximum $\max_{1\le i\le n} Z_{(\alpha_i,v_i)}$ was added for preventing the criterion from vanishing in regions where $\max_{w\in V} m(\alpha, w)$ would artificially overshoot the data and prohibit global exploration of the input space. This cap is needed for the consistency proof given later in this section. In turn, PEI straightforwardly inherits the tractability of EI, leading to the formula

$$PEI(\alpha, v) = \begin{cases} \sigma(\alpha, v)\,(g(\alpha, v)\Phi(g(\alpha, v)) + \phi(g(\alpha, v))) & \text{if } \sigma(\alpha, v) \neq 0, \\ 0 & \text{else,} \end{cases} \tag{3.5}$$

where $g(\alpha, v) = (m(\alpha, v) - T(\alpha))/\sigma(\alpha, v)$.

Figure 4 highlights the differences between the PEI 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 $\max_{1\le i\le n} f(\mathbf{x}_i)$, while with the PEI criterion the threshold depends on α. 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 $v_f^\star$. In terms of computational complexity, it is reasonable to anticipate that the most penalizing dimension will be that of the space V. Indeed, when the dimension of V increases, the computation of PEI at a given (α, v) becomes more difficult because of the optimization over V required to compute $T(\alpha)$. Moreover, the optimization of PEI over $A \times V$ is also more difficult. In the next section, we will show that it is possible to partially address this issue.

3.3. Computation and maximization of PEI. We are now interested in algorithms to compute and maximize the PEI criterion over $A \times V$. As explained previously, a deeper look at (3.3) shows that each computation of PEI at a location $(\alpha, v) \in A \times V$ involves a global optimization over V to find the threshold $T(\alpha)$. Thus, a brute-force maximization of PEI over $A \times V$ appears to be computationally expensive and highly sensitive to the dimension of V. If, on the other hand, for a given $\alpha \in A$, the threshold $T(\alpha)$ is computed, then computing PEI at (α, v) for any $v \in V$ simply requires computing kriging means and variances and applying the (fast to evaluate) formula (3.5). To take advantage of this instead of maximizing the PEI over $A \times V$, we suggest to equivalently maximize the profile PEI:

$$\alpha \in A \longrightarrow PEI^\star(\alpha) := \max_{w \in V} PEI(\alpha, w), \tag{3.6}$$

whose evaluation at any $\alpha \in A$ requires two 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 $\alpha^\star$ of $PEI^\star$ is identified, we obtain a global maximizer of PEI by taking $(\alpha^\star, v^\star)$, where $v^\star$ maximizes $PEI(\alpha^\star, \cdot)$. The advantage of the proposed method is that we now need to maximize a criterion over A and not over $A \times V$. In addition, this criterion is not much more expensive than the PEI: it involves only two optimizations over V instead of one. The maximization of $PEI^\star$ thus becomes much more convenient if the dimension of V increases. In fact, the maximization of $PEI^\star$ now has a cost which is "symmetric" in the sense that it depends similarly on the dimension of A and V.

The optimizations involved in the computation of $PEI^\star$ only require computing kriging means and variances. The latter can be evaluated at a large number of locations $\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+p} \in A \times V$ by precomputing the quantities that do not depend on them, including in particular the inverse of the covariance matrix at the observation locations, K^{-1}, and the matrix-vector product between K^{-1} and the column vectors that do not involve $\mathbf{x}$ (see (2.3) and (2.12)). For large p and with the proper precomputations, the cost for computing kriging means and variances at p locations are, respectively, in $O(np)$ and $O(n^2 p)$, where n is the number of available observations. In the present paper, given the relative simplicity of the examples, the two optimizations (over V) involved in (3.6) and the optimization of $PEI^\star$ (over A) are handled through a regular grid of size p for both α and v, so that the optimal location $(\alpha^\star, v^\star)$ can be found by simultaneously computing p^2 kriging means and variances. However, given the clear limitation of this approach when the dimension of the problem increases, our code also permits the use of other optimization methods, such as the genetic algorithm implemented in the *genoud* R package [26]. A budget too low for the optimizations over V or over A can have different effects. Regarding A first, a low optimization budget might cause the algorithm to pick a point which does not have the largest $PEI^\star$. For V it is the $PEI^\star$ itself which might not be correctly computed, as the threshold $T(\alpha)$ might be underestimated.

Figure 5 exhibits the sequence of points evaluated by sequentially maximizing PEI. The 20-point initial design of sections 2.2 and 3.2 is used. Then 40 points are added sequentially. One can see that after 40 iterations (bottom left and right plots), 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 $H^{f^\star}$ and $H^{v_f^\star}$ have been reduced.

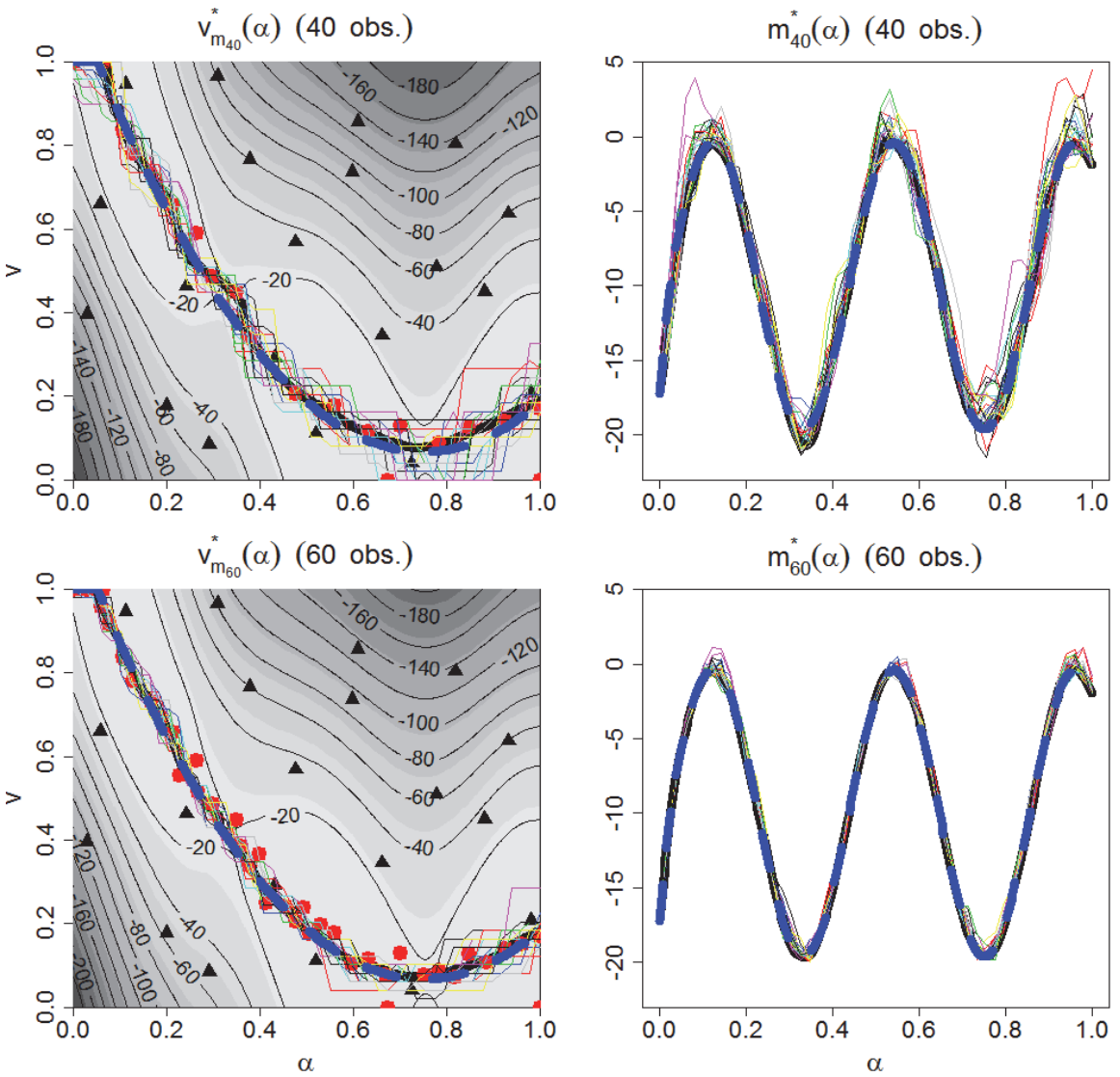

Figure 5. *Plug-in estimate (solid dark curves) $v^\star_m$ (left) and $m^\star$ (right) together with the true functions $v^\star_f$ and $f^\star$ (blue dashed curves) after having evaluated an initial* 20*-point design of experiments (black triangles) and* 20 *(top) or* 40 *(bottom) additional points (red triangles) chosen by sequentially maximizing PEI. In addition,* 25 *profile optima $z^\star_i$ and profile optimizers $v^\star_{z_i}$ are printed (thin lines in various colors), obtained from Gaussian process realizations conditioned on the available observations.*

This uncertainty reduction is quantified (relying on 1000 conditional simulations) in Table 3 and represented in Figure 6.

3.4. Convergence property with fixed mean and covariance. Here we follow the route of [38], where convergence properties are proven for the expected improvement algorithm under fixed mean and covariance functions. Throughout the section, the input space $\mathcal{X} \subset \mathbb{R}^d$ is assumed compact, and the covariance kernel k used for kriging is assumed to enjoy the so-called no-empty-ball (NEB) property.

Table 3

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	$H^{f^\star}$	11.47	1.62	0.63
L^2	$r^{f^\star}$	9.54	1.34	0.51
L^2	$b^{f^\star}$	3.92	0.42	0.09
L^2	$H^{v_f^\star}$	0.13	0.07	0.04
L^2	$r^{v_f^\star}$	0.10	0.05	0.03
L^2	$b^{v_f^\star}$	0.04	0.02	0.01
L^∞	$H^{f^\star}$	32.27	4.46	1.80
L^∞	$r^{f^\star}$	25.32	3.69	1.48
L^∞	$b^{f^\star}$	10.17	1.35	0.36
L^∞	$H^{v_f^\star}$	0.28	0.16	0.10
L^∞	$r^{v_f^\star}$	0.21	0.12	0.08
L^∞	$b^{v_f^\star}$	0.09	0.07	0.02

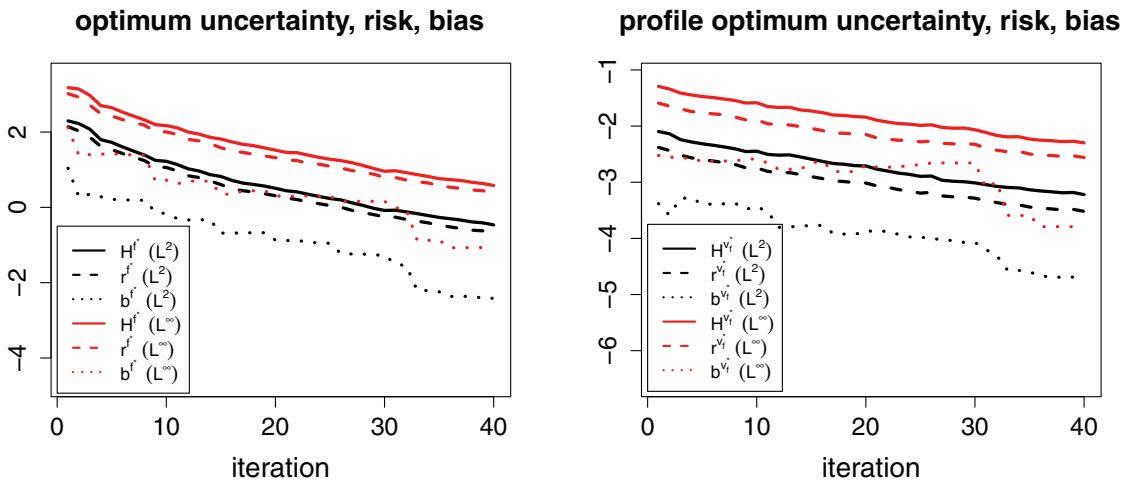

Figure 6. *Evolution of the uncertainty, risk, and bias indicators for the estimated profile optima (left) and optimizer functions (right) in the course of the PEI algorithm (log. scale).*

Definition 3.1. *The covariance kernel has the NEB property if, for all sequences* $(\mathbf{x}_n)_{n\geq 1}$ *in* $\mathcal{X}$ *and all* $\mathbf{y} \in \mathcal{X}$*, the following assertions are equivalent:*

(i) *y is an adherent point of the set* $\{\mathbf{x}_n, n \geq 1\}$.

(ii) $\sigma^2(\mathbf{y}; \mathbf{X}_n) \to 0$ *when* $n \to +\infty$,

where $\sigma^2(\mathbf{y}; \mathbf{X}_n)$ *refers to the kriging variance (a.k.a. mean-square prediction error) at the point* $\mathbf{y}$ *knowing the objective function at* $\mathbf{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{X}^n$.

Note that a sufficient condition for the NEB property is obtained in [38], where it is proved that stationary kernels with a spectral density converging to zero sufficiently slowly have the NEB property. In particular, the exponential covariances (with exponent smaller than two)

and the class of Matérn covariances have the NEB property [38], but the Gaussian covariance does not have it. This may cause inconsistency in EI algorithms, as illustrated in [41].

Another important assumption made in what follows is that the objective function f belongs to $\mathcal{H}$, the *reproducing kernel Hilbert space* (RKHS) associated with the kernel k [2, 5].

Here and in the following, $\rho_n^\star(\alpha, v) := PEI_n(\alpha, v) := \mathbb{E}\big[\big(Z_{(\alpha,v)} - T_n(\alpha)\big)^+ | Z_{\mathbf{X}_n} = \mathbf{z}_n\big]$ with $T_n(\alpha) := \min\big(\max_{w\in V} m_n(\alpha, w), \max_{1\le i\le n} Z_{(\alpha_i,v_i)}\big)$, where $\mathbf{z}_n$ refers to the objective function values at $\mathbf{X}_n$ and m_n refers to the kriging mean based on $\mathbf{X}_n, \mathbf{z}_n$. Note that subscripts are used in the same way for further quantities, such as in the following theorem. Note finally that here and in what follows, the dependence of various quantities on the objective function f may be stressed or not depending on the context. Before going further into technical detail, let us now state the main theorem of the section.

Theorem 3.2. *Assume that the covariance kernel k has the NEB property and that $\mathcal{X} = A \times V$ is compact. Then, for all $\mathbf{x}_{init} = (\alpha_{init}, v_{init}) \in A \times V$ and all $f \in \mathcal{H}$, the sequence $(\mathbf{x}_n)_{n\ge1} = (\alpha_n, v_n)_{n\ge1}$ generated by the algorithm*

$$\left\{ \begin{array}{l} \mathbf{x}_1 = \mathbf{x}_{init}, \\ \mathbf{x}_{n+1} = (\alpha_{n+1}, v_{n+1}) = \arg\max_{(\alpha,v)\in A\times V} \rho_n^\star(\alpha, v) \end{array} \right. \tag{3.7}$$

is dense in $\mathcal{X}$.

In other words, a sequence of points $(\mathbf{x}_n)_{n\ge1}$ generated by sequentially maximizing PEI will eventually fill the space. Note, furthermore, that as noticed in [38, Remark 5] in the case of the EI criterion, even if there is no guarantee that the maximizer of $\rho_n^\star$ is unique, considering algorithms that measurably choose $\mathbf{x}_{n+1}$ among the maximizers of $\rho_n^\star$ would lead to the same result. A crucial point for proving Theorem 3.2 is that the PEI criterion actually writes as

$$\rho_n^\star(\mathbf{x}) = \gamma(m_n(\mathbf{x}) - T_n(\mathbf{x}), \sigma_n^2(\mathbf{x})), \tag{3.8}$$

where $T_n(\mathbf{x}) := T_n(\alpha)$ for $\mathbf{x} = (\alpha, v)$ and

$$\gamma : (z, s) \in \mathbb{R} \times [0, +\infty) \longrightarrow \gamma(z, s) = \left\{ \begin{array}{ll} \sqrt{s}\phi(z/\sqrt{s}) + z\Phi(z/\sqrt{s}) & \text{if } s > 0, \\ z^+ = \max(z, 0) & \text{if } s = 0. \end{array} \right. \tag{3.9}$$

In particular, γ is known to be a continuous function satisfying the following: for all $z \le 0$, $\gamma(z, 0) = 0$ and for all $z \in \mathbb{R}$, for all $s > 0$, $\gamma(z, s) > 0$.

Before proving Theorem 3.2, let us state and prove a very useful lemma, an analogue for the PEI criterion of Lemma 12 of [38]. Always following [38], let us denote $M_n := \max(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))$ and $\nu_n^\star := \sup_{\mathbf{x}\in\mathcal{X}} \rho_n^\star(\mathbf{x})$ $(n \ge 1)$. In particular, let us remark that for all $n \ge 1$, $\nu_n^\star = \rho_n^\star(\mathbf{x}_{n+1}) = \gamma(m_n(\mathbf{x}_{n+1}) - T_n(\mathbf{x}_{n+1}), \sigma_n^2(\mathbf{x}_{n+1}))$.

Lemma 3.3. *For all $f \in \mathcal{H}$, $\liminf_{n\to+\infty} \nu_n^\star(f) = 0$.*

Proof. For any arbitrary $f \in \mathcal{H}$, set $s_n = \sigma_n^2(\mathbf{x}_{n+1}, f)$ and $z_n^\star = m_n(\mathbf{x}_{n+1}, f) - T_n(\mathbf{x}_{n+1}, f)$, so that $\nu_n^\star(f) = \gamma(z_n^\star, s_n)$. Let $\mathbf{y}^\star \in \mathcal{X}$ be an accumulation point of $(\mathbf{x}_n)$, and let $(\mathbf{x}_{\phi_n})$ be a subsequence of $(\mathbf{x}_n)$ converging to $\mathbf{y}^\star$. We are not exactly going to prove that $\nu_{\phi_n-1}^\star \to 0$, as done in [38] for an analogue quantity, but rather that $\nu_{\phi_{\psi_n}-1}^\star \to 0$ for some further extraction (i.e., increasing) function $\psi : \mathbb{N} \to \mathbb{N}$. We know from [38, Proof of Lemma 12]

that $m_{\phi_n-1}(\mathbf{x}_{\phi_n}, f) \longrightarrow f(\mathbf{y}^\star)$ and that $(M_{\phi_n-1}(f))$ is a bounded increasing sequence (and so converges) with a limit lower-bounded by $f(\mathbf{y}^\star)$. We will now prove that the sequence $(T_{\phi_n-1}(\mathbf{x}_{\phi_n})) = (\min(M_{\phi_n-1}(f), \max_{v\in V} m_{\phi_n-1}((\alpha_{\phi_n}, v), f)))$ is bounded.

As $(T_{\phi_n-1}(\mathbf{x}_{\phi_n}))$ is clearly upper-bounded by any upper bound of $(M_{\phi_n-1}(f))$, it suffices to prove that $(\max_{v\in V} m_{\phi_n-1}((\alpha_{\phi_n}, v), f))$ is lower-bounded. This last point follows from the fact that $\max_{v\in V} m_{\phi_n-1}(\alpha_{\phi_n}, v) \geq m_{\phi_n-1}(\mathbf{x}_{\phi_n}) \longrightarrow f(\mathbf{y}^\star)$. $(T_{\phi_n-1}(\mathbf{x}_{\phi_n}))$ being bounded, we can extract from it a converging subsequence $\left(T_{\phi_{\psi_n}-1}(\mathbf{x}_{\phi_{\psi_n}})\right)$. Since, by continuity, $f(\mathbf{x}_{\phi_{\psi_n-1}}) \longrightarrow f(\mathbf{y}^\star)$, we get that $T_{\phi_{\psi_n}-1}(\mathbf{x}_{\phi_{\psi_n}}) \geq \min\left(f(\mathbf{x}_{\phi_{\psi_n-1}}), m_{\phi_{\psi_n-1}}(\mathbf{x}_{\phi_{\psi_n}})\right) \longrightarrow f(\mathbf{y}^\star)$, and so

$$\lim_{n\to+\infty} z^\star_{\phi_{\psi_n}-1} = \lim_{n\to+\infty} m_{\phi_{\psi_n}-1}(\mathbf{x}_{\phi_{\psi_n}}, f) - \lim_{n\to+\infty} T_{\phi_{\psi_n}-1}(\mathbf{x}_{\phi_{\psi_n}}) \leq 0.$$

Finally, noting that $s_{\phi_{\psi_n}-1} \to 0$ we conclude similarly as in [38] that

$$\nu^\star_{\phi_{\psi_n}-1} = \gamma\left(z^\star_{\phi_{\psi_n}-1}, s_{\phi_{\psi_n}-1}\right) \to \gamma\left(\lim_{n\to+\infty} z^\star_{\phi_{\psi_n}-1}, 0\right) = 0,$$

which completes the proof of the lemma. ∎

Proof of Theorem 3.2. Again, we closely follow the proof scheme of [38, Theorem 6], with a slight adaptation due to specifics of the PEI criterion. Let us fix $f \in \mathcal{H}$ and assume that $(\mathbf{x}_n(f))$ is not dense in $\mathcal{X}$. Then there exists a point $\mathbf{y}^\star \in \mathcal{X}$ which is not adherent to $(\mathbf{x}_n(f))$, and so by the NEB property, $\inf_{n\geq 1} \sigma_n^2(\mathbf{y}^\star, f) > 0$. Besides, by the same argument as in [38], the sequence $(m_n(\mathbf{y}^\star, f))$ is bounded. We now use the fact that the specific γ of (3.9) is a decreasing function of the first variable for any fixed positive value of the second one to get that, for any $k \geq 1$, $\gamma(m_k(\mathbf{y}^\star, f) - T_k(\mathbf{y}^\star, f), \sigma_k^2(\mathbf{y}^\star, f)) \geq \gamma(m_k(\mathbf{y}^\star, f) - M_k(f), \sigma_k^2(\mathbf{y}^\star, f))$ as $T_k(\cdot, f) \leq M_k(f)$ and $\sigma_k^2(\mathbf{y}^\star, f) > 0$. We then obtain that

$$\begin{aligned}\rho_n^\star(\mathbf{y}^\star, f) &\geq \inf_{k\geq 1} \gamma\left(m_k(\mathbf{y}^\star, f) - T_k(\mathbf{y}^\star, f), \sigma_k^2(\mathbf{y}^\star, f)\right) \\ &\geq \inf_{k\geq 1} \gamma\left(m_k(\mathbf{y}^\star, f) - M_k(f), \sigma_k^2(\mathbf{y}^\star, f)\right) > 0,\end{aligned}$$

which contradicts Lemma 3.3 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 varied industries such as civil engineering, agriculture, pharmaceutical engineering, and energy production. In the framework of nuclear safety, after many years the fuel pellets in a nuclear reactor develop cracks because of thermal stresses. In addition, during a hypothetical nuclear accident, additional fragmentation is postulated to occur because of the thermal-mechanical response to the transient loading conditions. The fuel pellets stacked into tubes can thus be considered as close-packed granular clusters.

The complex static and flow properties of granular materials have been extensively studied during the last decades [13]. 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, is a key quantity to characterize the state of granular materials.

In the test case considered here, the granular material is represented by a configuration of hard spheres subjected to unilateral contacts. Note that in the monodisperse case, the close packing density in two dimensions is known and is about 0.907 (crystalline arrangements). However, few studies focus on polydispersity and its impact on mechanical behavior. For polydisperse granular media, the close packing density depends on the geometry of grains. The close packing density can hence be viewed as a maximum of a geometric problem, which is often tackled using stochastic methods [12, 33]. For any initial grain configuration, the corresponding optimal state can be approximated numerically by assuming grains without friction and hydrostatic stress on the sample [24, 1]. In practice, the close packing density is typically estimated by taking the maximum among numerically estimated close packing densities obtained from a number of random initial grain configurations.

In this paper, the previous Gaussian random field framework is applied to the study of the close packing density of a bidisperse sample in two dimensions. The contact interactions between the grains are treated using the nonsmooth contact dynamics approach [29, 19]. The numerical developments have been performed in the framework of the LMGC90 platform [23]. The main assets of LMGC90 are the large range of contact laws already implemented and its ability to consider more complex surface interactions (cohesion, wear, etc.) with neither regularization nor penalization schemes.

4.1.2. Description of the test case. The initial domain is a square box randomly filled (uniformly, with rejection in case of overlap) 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., $R_2 \leq R_1$). These input parameters are denoted by X and W, and their ranges of variation are $[4, 10]$ and $[0.1, 0.9]$, respectively. Moreover, a pressure stress $P = 10^6 Pa$ is exerted on the top and the right sides, and the left and bottom sides are blocked (see Figure 7).

For any given ratio X, the goal is to determine the corresponding density W leading to the maximum of the code response. Returning 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 time of one simulation is about 4 hours using in-house computing facilities. Therefore, an adaptive experimental design appears to be an appropriate alternative to classical experimental designs because it is compulsory to keep an affordable number of simulations, while ensuring an accurate reconstruction of the curve of interest. Before considering the design construction, a first step in our analysis consists in estimating the map of profile optimizers from a large set of simulations. The obtained estimate will then be used to evaluate the capability of the adaptive experimental designs obtained by our procedure. All the numerical tests have been performed using R statistical software and an application programming interface between LMGC90 and R developed in Python thanks to the rPython R package [15].

4.2. Estimation of the function of profile optimizers. The range of variation associated with the input parameters is discretized using a full factorial experimental design with 25

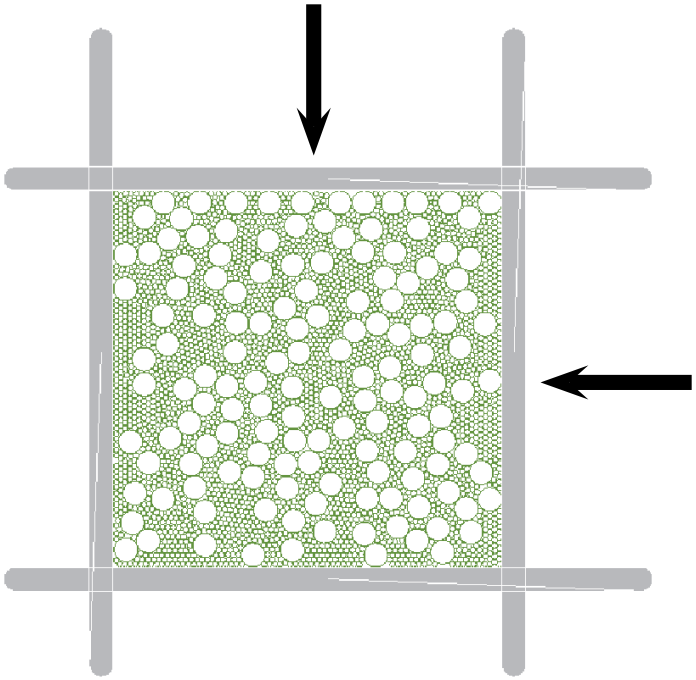

Figure 7. *Random close packing of* 3000 *hard disks.*

(resp., 41) levels in the α (resp., v) directions. It turns out that the simulated maximum packing fraction exhibits a strong spatial variability due to the random initial configuration of the hard spheres [12, 33]. As nondeterministic results lead to a perturbated curve of interest, kriging with homogeneous noise variance (recalled in section 2) is applied on a 200×200 regular grid in order to integrate the variability within the data modeling. The noise variance τ^2 is assumed constant on the whole domain and has been estimated to 10^{-5} using the nugget estimation procedure of the DiceKriging package [34].

The corresponding postprocessed maximum packing fraction is depicted in Figure 8. As expected, the spatial variability in the reconstructed surface has been reduced, leading to a simple curve of profile optimizers ($v_f^\star(\alpha) \approx 0.7$) with a very small dependence on the dispersity ratio α, which is consistent with the literature [14].

Since the reference map of profile optimizers and the profile optimum function are estimated after kriging, they are subject to uncertainty. As an example, Figure 8 shows these two quantities obtained from $s = 25$ realizations, $z_i, 1 \leq i \leq s$, of a Gaussian random field conditioned on 1025 observations. Moreover, the third column of Table 4 lists the values of the associated uncertainty, risk, and bias measures computed from 1000 simulations.

4.3. Construction and performance evaluation of adaptive experimental designs based on PEI. We now want to compare the performances of our criterion based on sequential PEI maximization with other evaluation strategies, such as sequential IMSE minimization (see, e.g., [32]), sequential EI maximization, or pure uniform random sampling. A total of 100 random initial experimental designs of 20 points each are used. They are generated with the

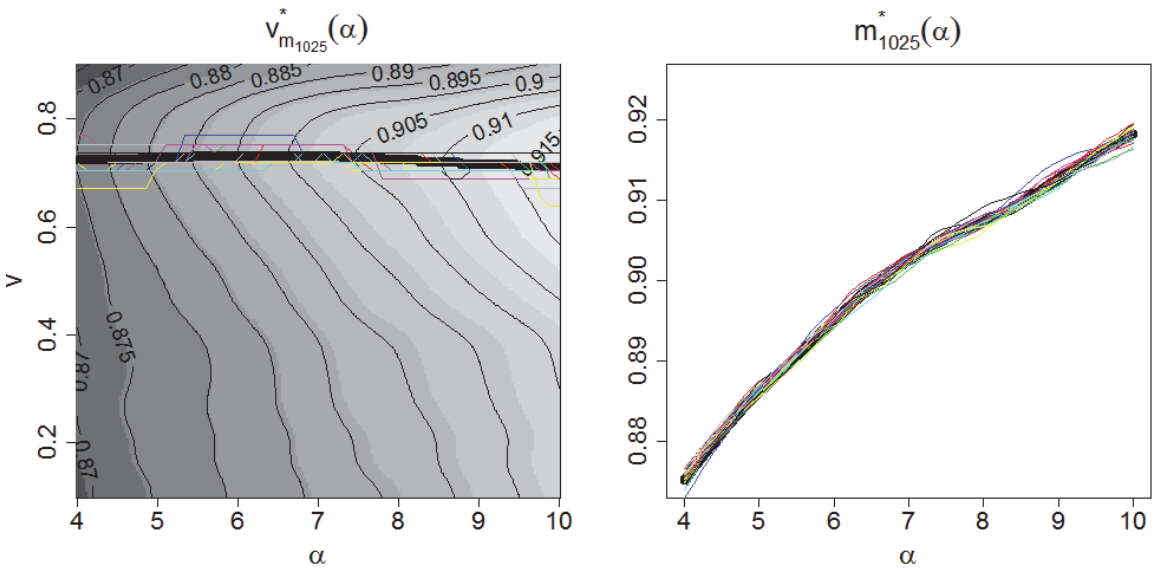

Figure 8. *Maximum packing fraction obtained from a* 1025*-point full factorial experimental design. The surfaces are reconstructed by kriging with homogeneous noise on the original data.* 25 *functions of profile optimizers* $v^\star_{z_i}$ *and profile optima* $z_i^\star$ *of* $s = 25$ *conditional realizations* $z_i, 1 \leq i \leq s$*, of a Gaussian random field conditioned on* 1025 *observations of the test function are also represented.*

optimumLHS function of the LHS R package. Each strategy is tested using all 100 initial designs. For each strategy and initial design, we sequentially evaluate a total of 40 additional points and compare the results in terms of uncertainty, risk, and bias. The results are then averaged. At each iteration, the objective function is evaluated at one point only. Concerning the Gaussian random field model, ordinary kriging with a separable Matérn covariance kernel (with smoothness parameter $\nu = 3/2$) is used. The parameters of this covariance kernel are estimated, plugged in, and renewed at each iteration using maximum likelihood estimation. To calculate the bias and the risk, we assume that the reference profile optimizer and profile optimum functions are those obtained after having evaluated the function on the full 1025-point experimental design.

A number of conclusions can be drawn from the results given in Table 4. It appears that the evaluation strategy based on the PEI maximization provides good performance for estimating both the profile optimizer and the profile optimum functions. In particular, PEI clearly outperforms the other strategies for the estimation of $f^\star$. However, although the performances of PEI remain good in terms of estimation of $v^\star_f$, its domination over the other criteria is less marked than for $f^\star$. Similarly to the EI, which is more suited to find the maximum of a function than the maximizer (see, e.g., [39]), the PEI is more adapted to find the profile optimum function $f^\star$ than the profile optimizer. Another important remark is that the covariance parameters 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 of our measures of uncertainty if, e.g., one strategy tends to evaluate locations leading to a higher range estimate. For completeness, we also conducted an analogue experiment where the covariance parameters were estimated from the initial design of experiment of 20

Table 4

Comparison of the decrease of the model uncertainty, risk, and bias obtained from different sequential evaluation strategies. The results are averaged over a total of 100 *runs. Values in brackets are the associated standard deviations (Monte Carlo estimator of the mean).*

Norm	Ind.	Full design	20-pt. initial design	Uniform design	IMSE design	EI design	PEI design
L^2	$H^{f^\star}$	0.18	0.80 (0.012)	0.56 (0.007)	0.52 (0.005)	0.79 (0.016)	**0.39 (0.004)**
L^2	$r^{f^\star}$	0.13	0.93 (0.022)	0.56 (0.011)	0.54 (0.011)	0.90 (0.024)	**0.43 (0.010)**
L^2	$b^{f^\star}$	0	0.78 (0.033)	0.40 (0.015)	0.41 (0.018)	0.61 (0.021)	**0.32 (0.015)**
L^2	$H^{v^\star_f}$	4.4	34.9 (1.19)	20.4 (0.78)	20.9 (0.75)	32.2 (1.24)	**18.2 (0.37)**
L^2	$r^{v^\star_f}$	3.1	28.3 (1.01)	16.0 (0.65)	16.7 (0.63)	26.4 (1.04)	**15.4 (0.40)**
L^2	$b^{v^\star_f}$	0	15.6 (1.18)	7.3 (0.74)	**6.8 (0.63)**	17.4 (1.36)	7.6 (0.44)
L^∞	$H^{f^\star}$	0.16	0.70 (0.013)	0.50 (0.007)	0.46 (0.005)	0.69 (0.016)	**0.33 (0.004)**
L^∞	$r^{f^\star}$	0.11	0.86 (0.020)	0.54 (0.012)	0.49 (0.007)	0.91 (0.027)	**0.36 (0.009)**
L^∞	$b^{f^\star}$	0	0.67 (0.026)	0.38 (0.013)	0.36 (0.012)	0.57 (0.021)	**0.27 (0.012)**
L^∞	$H^{v^\star_f}$	3.3	34.6 (1.27)	24.7 (1.18)	26.4 (1.18)	34.5 (1.25)	**22.9 (0.60)**
L^∞	$r^{v^\star_f}$	2.4	24.9 (0.99)	17.5 (0.90)	19.1 (0.91)	24.4 (1.02)	**16.5 (0.43)**
L^∞	$b^{v^\star_f}$	0	13.5 (1.57)	7.5 (1.42)	6.7 (1.29)	17.2 (2.21)	**4.5 (0.24)**

points and not re-estimated afterwards. It appeared that the performance of all the tested strategies (except the random sampling) was slightly worse than previously. However, this did not change the ranking of the performances, so our latter conclusions remained unchanged.

5. Conclusions.

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 in the considered two-dimensional test cases, as it allowed us to efficiently reconstruct the curves of interest within a 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 the covariance parameters [11] or appealing to a full Bayesian approach [4] 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 or not did not lead to significant differences, so that basing parameter estimation solely on initial evaluations appeared to be sufficient 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* (EI) [27, 21] criterion to our specific setup. 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 proved in section 3. In addition, experimental results obtained with both a toy function from the literature and in 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 welcome surprise is that PEI also performs well for learning the curve of profile optimizers, even though it was not directly conceived

for it. Significant improvements for learning profile optimizer functions might be obtained, e.g., by applying the principles of *sequential uncertainty reduction* (SUR) strategies [3] 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.

Acknowledgments. This work was conducted within the framework of the ReDICE Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, and 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 and Robert Komara for having kindly accepted proofreading the article. They also warmly thank the two anonymous referees for their constructive comments and suggestions.

REFERENCES

[1] I. Agnolin and J. N. Roux, *Internal states of model isotropic granular packings*, Phys. Rev. E (3), 76 (2007), 061304.
[2] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., 68 (1950), pp. 337–404.
[3] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez, *Sequential design of computer experiments for the estimation of a probability of failure*, Stat. Comput., 22 (2012), pp. 773–793.
[4] R. Benassi, *Nouvel algorithme d'optimisation bayésien utilisant une approche Monte-Carlo séquentielle*, Ph.D. thesis, Supélec, Gif-sur-Yvette, France, 2013.
[5] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers, Boston, MA, 2004.
[6] L. Buslig, J. Baccou, and V. Picheny, *Construction and efficient implementation of adaptive objective-based designs of experiments*, Math. Geosci., 46 (2013), pp. 285–313.
[7] R. Carnell, *lhs: Latin Hypercube Samples*, R package version 0.5, 2009.
[8] C. Chevalier, *Fast Uncertainty Reduction Strategies Relying on Gaussian Process Models*, Ph.D. thesis, University of Bern, Bern, Switzerland, 2013.
[9] J.-P. Chilès and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed., Wiley, New York, 2012.
[10] N. Cressie, *Statistics for Spatial Data*, John Wiley and Sons, New York, 1993.
[11] D. den Hertog, J. P. C. Kleijnen, and A. Y. D. Siem, *The correct kriging variance estimated by bootstrapping*, J. Oper. Res. Soc., 57 (2006), pp. 400–409.
[12] M. Dijkstra, R. von Roij, and R. Evans, *Phase diagram of highly asymmetric binary hard-sphere mixtures*, Phys. Rev. E (3), 59 (1999), pp. 5744–5770.
[13] J. Duran, *Sands, Powders, and Grains: An Introduction to the Physics of Granular Materials*, Springer-Verlag, New York, 1999.
[14] R. S. Farr and R. D. Grott, *Close packing density of polydisperse hard spheres*, J. Chem. Phys., 131 (2009), 244104.
[15] C. J. Gil Bellosta, *rPython: Package allowing R to call Python*, R package version 0.0-4, 2013.
[16] 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, Berlin, Heidelberg, 2010, pp. 89–96.

[17] D. Ginsbourger, R. Le Riche, and L. Carraro, *Kriging is well-suited to parallelize optimization*, in Computational Intelligence in Expensive Optimization Problems, Springer, Berlin, Heidelberg, 2010, pp. 131–162.
[18] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, 2001.
[19] M. Jean, *The non-smooth contact dynamics method*, Comput. Methods Appl. Mech. Engrg., 177 (1999), pp. 235–257.
[20] D. R. Jones, *A taxonomy of global optimization methods based on response surfaces*, J. Global Optim., 21 (2001), pp. 345–383.
[21] D. R. Jones, M. Schonlau, and J. Welch, *Efficient global optimization of expensive black-box functions*, J. Global Optim., 13 (1998), pp. 455–492.
[22] J. Kim and D. Pollard, *Cube root asymptotics*, Ann. Statist., 18 (1990), pp. 191–219.
[23] LMGC Laboratory, *Lmgc*90, https://subver.lmgc.univ-montp2.fr/trac_LMGC90v2/.
[24] H. A. Makse, D. L. Johnson, and M. Schwartz, *Packing of compressible granular materials*, Phys. Rev. Lett., 84 (2000), pp. 4160–4163.
[25] G. Matheron, *The intrinsic random functions, and their applications*, Advances in Appl. Probability, 5 (1973), pp. 439–468.
[26] W. Mebane and J. Sekhon, *Genetic optimization using derivatives: The rgenoud package for R*, J. Stat. Software, 42 (2011), pp. 1–26.
[27] J. Mockus, *Bayesian Approach to Global Optimization: Theory and Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989.
[28] I. Molchanov, *Theory of Random Sets*, Springer, London, 2005.
[29] J. J. Moreau, *Unilateral contact and dry friction in finite freedom dynamics*, in Non Smooth Mechanics and Applications, CISM Courses and Lectures 302, Springer, Vienna, 1988, pp. 1–82.
[30] H. Omre and K. B. Halvorsen, *The Bayesian bridge between simple and universal kriging*, Math. Geol., 22 (1989), pp. 767–786.
[31] V. Picheny and D. Ginsbourger, *Noisy kriging-based optimization methods: A unified implementation within the DiceOptim package*, Comput. Statist. Data Anal., 71 (2014), pp. 1035–1053.
[32] V. Picheny, D. Ginsbourger, O. Roustant, R. T. Haftka, and N.-H. Kim, *Adaptive designs of experiments for accurate approximation of target regions*, J. Mech. Des., 132 (2010), 071008.
[33] P. M. Reis and T. Mullin, *Granular segregation as a critical phenomenon*, Phys. Rev. Lett., 89 (2002), 244301.
[34] O. Roustant, D. Ginsbourger, and Y. Deville, *DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodelling and optimization*, J. Stat. Software, 51 (2012), pp. 1–55.
[35] T. J. Santner, B. J. Williams, and W. Notz, *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York, 2003.
[36] T. Severini and W. H. Wong, *Profile likelihood and conditionally parametric models*, Ann. Statist., 20 (1992), pp. 1768–1802.
[37] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York, 1999.
[38] E. Vazquez and J. Bect, *Convergence properties of the expected improvement algorithm with fixed mean and covariance functions*, J. Statist. Plann. Inference, 140 (2010), pp. 3088–3095.
[39] J. Villemonteix, *Optimisation de fonctions coûteuses*, Ph.D. thesis, Faculté des Sciences d'Orsay, Université Paris-Sud XI, Orsay, France, 2008.
[40] J. Villemonteix, E. Vazquez, and E. Walter, *An informational approach to the global optimization of expensive-to-evaluate functions*, J. Global Optim., 44 (2009), pp. 509–534.
[41] D. Yarotsky, *Examples of inconsistency in optimization by expected improvement*, J. Global Optim., 56 (2012), pp. 1773–1790.