GOAL-ORIENTED ADAPTIVE SAMPLING UNDER RANDOM FIELD MODELLING OF RESPONSE PROBABILITY DISTRIBUTIONS

Athénaïs Gautier[1], David Ginsbourger[1] and Guillaume Pirot[2]

Abstract. In the study of natural and artificial complex systems, responses that are not completely determined by the considered decision variables are commonly modelled probabilistically, resulting in response distributions varying across decision space. We consider cases where the spatial variation of these response distributions does not only concern their mean and/or variance but also other features including for instance shape or uni-modality versus multi-modality. Our contributions build upon a non-parametric Bayesian approach to modelling the thereby induced fields of probability distributions, and in particular to a spatial extension of the logistic Gaussian model. The considered models deliver probabilistic predictions of response distributions at candidate points, allowing for instance to perform (approximate) posterior simulations of probability density functions, to jointly predict multiple moments and other functionals of target distributions, as well as to quantify the impact of collecting new samples on the state of knowledge of the distribution field of interest. In particular, we introduce adaptive sampling strategies leveraging the potential of the considered random distribution field models to guide system evaluations in a goal-oriented way, with a view towards parsimoniously addressing calibration and related problems from non-linear (stochastic) inversion and global optimisation.

1. Introduction

Many problems in science and engineering boil down to studying the effect on a response of interest of varying some *decision* or *control* variables $\mathbf{x}$. Yet it is typically unrealistic to assume a deterministic relationship between $\mathbf{x}$ and the response, be it for instance because of uncertainty in other input variables or because the assumed response and/or observation generating processes themselves involve some randomness. Addressing optimisation and related tasks in such frameworks comes with both conceptual and implementation challenges. While a number of contributions around stochastic and "noisy" optimisation have been developed, they often rely on homoscedasticity or specific distributional assumptions. In contrast, expensive-to-evaluate systems for which various features of response distributions evolve in decision space call for versatile modelling approaches.

Here we consider the situation where responses follow probability distributions $\mu_{\mathbf{x}}$ indexed by decision variables $\mathbf{x} \in D$ and with a common support $\mathcal{I}$. Without loss of generality, the index set –or *decision space*– D is assumed to be a compact subset of $\mathbb{R}^d$ ($d = 1$ or 2 in forthcoming examples). We furthermore assume that an objective function $g(\mathbf{x}) = \rho(\mu_{\mathbf{x}})$ depending on $\mathbf{x}$ through $\mu_{\mathbf{x}}$ is to be minimised, where ρ returns for any considered probability distribution a real-valued quantity such as a moment or a quantile with some given level. One of the main challenges in the considered framework is that typically, $g(\mathbf{x})$ cannot be exactly evaluated as one only observes draws from $\mu_{\mathbf{x}}$. We use the letter t (resp. T in random form) to denote such draws.

---

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

[2] Centre for Exploration Targeting, The University of Western Australia, Australia
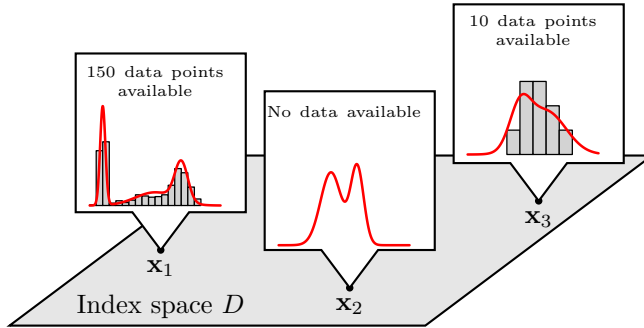
FIGURE 1. Schematic representation of a field of probability distributions $\mu_{\mathbf{x}}$ (represented by their probability density functions, in red) to be predicted based on scattered samples (with associated histograms in grey). Here the end goal is to collect additional samples towards the minimisation of $g(x) = \rho(\mu_{\mathbf{x}})$.

In stochastic optimisation (Ruszczyński and Shapiro [42]; Birge and Louveaux [7]; Prékopa [36]), having adapted assumptions for $\mu_{\mathbf{x}}$ is important to achieve good results. Approximation procedures have been studied, including but not limited to the Robbins Monro procedures and further developments in Bayesian Approximation (Robbins and Monro [39]; Mandt et al. [27]) or the multi-armed bandit paradigm (Thompson [45]; Cesa-Bianchi and Lugosi [9]; Bubeck and Cesa-Bianchi [8]). The most natural choice for the functional ρ is to consider the expectation. Yet, many other choices for ρ have been considered. These choices include quantiles (Rostek [41]; Torossian et al. [49]), the conditional value at risk (Rockafellar and Uryasev [40]) or expectiles (Bellini and Di Bernardino [3]). It is also possible to learn the unknown distribution $\mu_{\mathbf{x}}$ (Hall et al. [17]; Efromovich [13]; Moutoussamy et al. [30]; Zhu and Sudret [50]) but it generally requires a large sample size.

In the context of global optimisation under expensive objective function evaluations, Bayesian Optimisation (BO) leverages the potential of meta-modelling, especially Gaussian Processes (GP) (Rasmussen and Williams [37]), to keep a memory of explored points with the aim to explore decision space while keeping a parsimonious evaluation budget. First introduced by Močkus et al. [29] and further developed and popularized by Jones et al. [21] in the noise free setting, it has latter been extended to stochastic black box optimisation (Frazier et al. [15], Frazier [16]; Srinivas et al. [43]; Picheny et al. [33]; Hernández-Lobato et al. [18]; Jalali et al. [19] and references therein) and further sequential strategies have been studied (Risk and Ludkovski [38]; Binois et al. [6]). However, existing approaches typically assume Gaussian response distributions $\mu_{\mathbf{x}}$. Another branch of study pertaining to geostatistics and that does not rely on Gaussian or specific distributional assumptions is the so-called distributional Kriging (Aitchison [1]; Egozcue et al. [14]; Talská et al. [44]) but such approaches are ill-suited in the case of moderate sample size heterogeneously scattered across space.

In this paper, we investigate and leverage a non-parametric Bayesian model for fields of probability distributions that generalizes the logistic Gaussian process model. We focus on the situation where, for each $\mathbf{x}$, the distribution $\mu_{\mathbf{x}}$ admits a density $p_{\mathbf{x}}$ with respect to a prescribed dominating measure. Our approach is adapted to the case of moderate and heterogeneously scattered across space sample size. It delivers a probabilistic prediction of the field of probability distributions which allows us to derive sampling acquisition schemes. We introduce the aforementioned model in the Section 2. Then, we study an application of this model in sequential design in Section 3 and demonstrate its applicability on toy probability density fields and on a contaminant localization problem in Section 4.

The main contribution of this paper lies in the exploitation of a spatial extension of the logistic Gaussian process model for the sequential design of stochastic simulations towards optimisation and inversion.

2. Sample-based modelling of density valued fields

2.1. The logistic Gaussian process for density estimation

Before considering the full scale problem of density field estimation, let us briefly consider the simpler problem of (univariate) density estimation without any indexing on a spatial variable.

Our focus is on a class of Bayesian density estimation methods where prior distributions of probability density functions are constructed by rescaling positive random processes in such a way that resulting realisations integrate to one. When the considered random processes are obtained by taking the logistic density transform (exponentiation and normalisation) of a Gaussian process, the resulting model is called Logistic Gaussian Process (LGP). LGP for density estimation were established and studied by Leonard [26] and Lenk [24, 25]. More recently, Tokdar and Ghosh [47] demonstrated that a large class of LGP achieves strong and weak consistency (under some regularity assumptions) at true densities that are either continuous or piecewise continuous.

Using this prior within a Markov Chain Monte Carlo framework allows approximate sampling from the posterior distribution over the space of densities on $\mathcal{I}$, as illustrated in Figure 2.
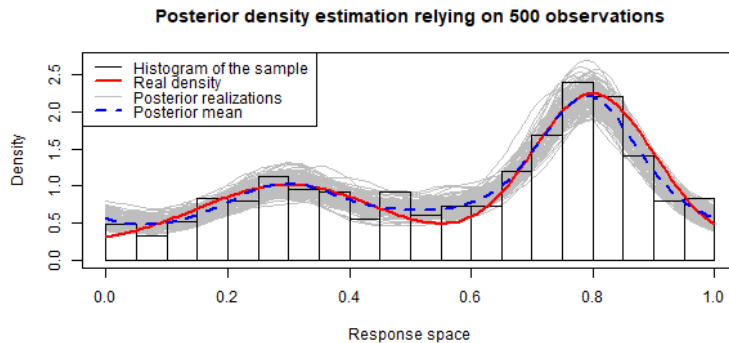

FIGURE 2. Using the logistic Gaussian process model in density estimation.

Models relying on other mappings than the *logistic density transform* have been introduced and studied. In the particular case where the sigmoid (*logistic function*) is used instead of an exponential, Murray et al. [31] introduced an exact sampling approach relying on the boundedness of the sigmoid. More recently, Donner and Opper [11] proposed an augmented model that allows for efficient inference by Gibbs sampling and an approximate variational mean field approach. However, our first experiments hinted that choosing a bounded function such as the sigmoid often leads to a lack of expressiveness, hence we decided to rely on LGP settings.

2.2. A spatial extension of the logistic Gaussian process for density field estimation

When modelling and optimising functions depending on both decision and uncontrolled variables, it can be a relevant option to use augmented models incorporating the uncontrolled variables as part of their inputs. In Picheny et al. [33], a GP indexed by the joint space of design parameters and computational time is considered to model partially converged simulation data. In Janusevskis and Le Riche [20], a GP model over the joint spaces of controllable and uncontrollable variables is used for simulation-based robust optimisation.

Here we consider a spatial extension of the logistic Gaussian process. We work with a GP indexed by the product space $D \times \mathcal{I}$ and apply a logistic density transform to each $\mathbf{x}$-slice of the process to produce a probability density field. A similar spatial LGP was introduced in Tokdar et al. [48] in the context of conditional density modelling (i.e., with random $\mathbf{x}$). To our knowledge, this family of models has not yet been used in the present context of sequential design of experiments, be it for stochastic optimisation or towards other goals.

Here we assume that our probability measures $\mu_{\mathbf{x}}$ of interest possess probability densities $p_{\mathbf{x}}$ (with respect to some prescribed reference measure(s)) and we appeal to the said generalization of LGP to model the resulting density field. The resulting Spatial Logistic Gaussian process (SLGP) is defined by

$$\mathfrak{p}_{\mathbf{x}}(t) = \frac{e^{W_{\mathbf{x},t}}}{\int_{\mathcal{I}} e^{W_{\mathbf{x},u}} \, du} \qquad (\mathbf{x} \in D, t \in \mathcal{I}), \tag{1}$$

where $W_{\mathbf{x},t} \sim \mathcal{GP}(m,k)$ is a measurable GP with given mean function $m : D \times \mathcal{I} \to \mathbb{R}$ and covariance kernel $k : (D \times \mathcal{I}) \times (D \times \mathcal{I}) \to \mathbb{R}$ such that for any $\mathbf{x} \in D$, $\int_{\mathcal{I}} e^{W_{\mathbf{x},u}} \, du < \infty$ almost surely.

Remark 2.1. When the index space D is a singleton, SLGP coincides with the LGP mentioned in Section 2.1.

Remark 2.2. Throughout the paper, the Lebesgue measure is assumed as dominating one, $\mathcal{I}$ is $[0,1]$, and the considered GPs possess smoothness properties amply guaranteeing the a.s. existence of the considered integrals.

2.3. **Conditioning on data and practicalities**

In the SLGP model, the denominator involves values of W over the whole domain. This infinite dimensional objects makes likelihood-based computations cumbersome. In Tokdar [46], this problem is reduced to the one of a finite dimensional integral by relying on a moderate number of inducing points, introduced to reduce the dimensionality of the LGP. These inducing points are selected according to a life and death process.

Here, we opt for another way to reduce the dimensionality that is more suitable for sequential optimisation. We consider finite-rank GP with the following parametrisation

$$W_{\mathbf{x},t} = m(\mathbf{x},t) + \sum_{j=1}^{p} \sqrt{\lambda_j} e_j(\mathbf{x},t) \varepsilon_j \ (\mathbf{x} \in D, \ t \in \mathcal{I}) \tag{2}$$

where $p \in \mathbb{N}$, the e_i's are functions, the λ_i's are scalars satisfying $\lambda_1 \geq \ldots \geq \lambda_p > 0$ and the ε_i's are i.i.d. $\mathcal{N}(0,1)$. Truncated Karhunen-Loève expansions constitute an important class falling under this framework. This simple form of the GP allows us to pre-calculate the values of e_j's at the grid points and then swiftly obtain evaluations of W and its integral (through a quadrature) for varying realisations (or tentative values) ϵ_i's of the ε_i's.

Now, consider that our data consist in n couples of locations and observations $\{(\mathbf{x}_i, t_i)\}_{1 \leq i \leq n} \in (D \times \mathcal{I})^n$. Moreover, assume the t_i's are realisations of some independent random variables T_i, and let $\mu_{\mathbf{x}_i}$ be the (unknown) distribution of T_i. We will note $\mathbf{T}_n = (T_i)_{1 \leq i \leq n}$ and $\mathbf{t}_n = (t_i)_{1 \leq i \leq n}$. The GP W defined by the Equation 2, as well as the associated SLGP inherit their randomness from the Gaussian random vector $\boldsymbol{\varepsilon} = (\varepsilon_i)_{1 \leq i \leq p}$. To highlight this dependency we shall also denote the SLGP $\mathfrak{p}_{\mathbf{x}}(\cdot)$ by $\mathfrak{p}_{\mathbf{x}}(\cdot; \boldsymbol{\varepsilon})$ and use the notation $\mathfrak{p}_{\mathbf{x}}(\cdot; \boldsymbol{\epsilon})$ when instantiating the SLGP with a fixed value $\boldsymbol{\epsilon}$ of $\boldsymbol{\varepsilon}$. We perform a Bayesian estimation of the true unknown density field $p_{\mathbf{x}_i}$ in the following way:

- We use the prior $\boldsymbol{\varepsilon} = (\varepsilon_i)_{1 \leq i \leq p} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, with density $\phi(\cdot)$ (with respect to Lebesgue measure in $\mathbb{R}^p$).
- The finite-rank SLGP $\mathfrak{p}_{\mathbf{x}}(\cdot; \boldsymbol{\varepsilon}) := \dfrac{e^{m(\mathbf{x},\cdot) + \sum_{j=1}^{p} \sqrt{\lambda_j} e_j(\mathbf{x},\cdot) \varepsilon_j}}{\displaystyle\int_{\mathcal{I}} e^{m(\mathbf{x},u) + \sum_{j=1}^{p} \sqrt{\lambda_j} e_j(\mathbf{x},u) \varepsilon_j} \, du}$ is used to model $p_{\mathbf{x}}(\cdot)$.

Hence, for $T_1, ..., T_n$ independent responses at locations $\mathbf{x}_1, ..., \mathbf{x}_n$, one obtains the likelihood

$$L(\cdot; \mathbf{t}_n) : \boldsymbol{\epsilon} \in \mathbb{R} \to L(\boldsymbol{\epsilon}; \mathbf{t}_n) = \prod_{i=1}^{n} \mathfrak{p}_{\mathbf{x}_i}(t_i; \boldsymbol{\epsilon}) \tag{3}$$

- The posterior density of $\boldsymbol{\varepsilon}$, given by Bayes formula is $\pi(\boldsymbol{\epsilon}|\mathbf{T}_n = \mathbf{t}_n) \propto \phi(\boldsymbol{\epsilon}) \prod_{i=1}^{n} \mathfrak{p}_{\mathbf{x}_i}(t_i; \boldsymbol{\epsilon})$

We leverage the fact that the posterior is known (up to a constant) to implement our density field estimation via a preconditioned Crank Nicholson algorithm (Cotter et al. [10]). This approach delivers a probabilistic prediction of $(p_{\mathbf{x}})_{\mathbf{x} \in D}$, and allows us to approximately sample from the SLGP posterior distribution and recommend new sampling locations.

3. SOME FIRST CONTRIBUTION IN STOCHASTIC OPTIMISATION

One of the main challenges in optimisation is to define suitable sequential design strategies. Choosing where to add observations is indeed crucial and a good design strategy must achieve a trade-off between exploration and exploitation of the input space so as to discover regions of interest while avoiding getting trapped in the vicinity of local minimisers or of artefactual basins of attraction.

Deriving such strategies requires anticipating (probabilistically) the effect of adding new observations. To this end, meta-models are commonly used. We now present the considered optimisation problem, the criterion used to derive a sequential design strategy, and a stochastic simulation-based approach to estimate this criterion.

3.1. **Optimisation problem considered**

We recall that we are interested in minimising $g(\mathbf{x}) = \rho(\mu_{\mathbf{x}})$, while we do not know the field $\mu_{\mathbf{x}}$ (or equivalently its density $p_{\mathbf{x}}$). We consider $G_{\mathbf{x}}$, the random function obtained by applying ρ to the density valued field delivered by the SLGP model. $G_{\mathbf{x}}$ is the model for $g(\mathbf{x})$ induced by the SLGP $\mathfrak{p}_{\mathbf{x}}(\cdot)$.

In the rest of the paper we will consider the particular case where ρ is the median, but the presented approach is not restricted to this choice and can be applied to arbitrary (measurable) mappings, potentially also mappings depending on $\mathbf{x}$.

Remark 3.1. $G_{\mathbf{x}}$ remains uncertain knowing $\mathbf{T_n} = \mathbf{t_n}$ because of the conditional variability of $\boldsymbol{\varepsilon}$.

In the spirit of robust optimisation, we will consider the problem of minimising an α-quantile of the random function $G_{\mathbf{x}}$. Here the value of α is arbitrarily set to 0.9 and stays fixed through the optimisation procedure. We note $Q_n(\mathbf{x})$ the α-quantile of $G_{\mathbf{x}}$. $Q_n(\mathbf{x})$ is a random function as the observations $\mathbf{T}_n$ are left in random form. We shall also consider $Q_{n+K}(\mathbf{x}; \mathbf{x}_{\text{new}})$, the α-quantile of $G_{\mathbf{x}}$ conditioned on past observations T_n and on a batch of K i.i.d. observations to be made at $\mathbf{x}_{\text{new}} \in D$. Denoting $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot | \mathbf{T_n} = \mathbf{t_n}]$, we consider:

$$\text{EQI}_n(\mathbf{x}_{\text{new}}, K) = \mathbb{E}_n \left[\left(\min_{\mathbf{x} \in D} Q_n(\mathbf{x}) - \min_{\mathbf{x} \in D} Q_{n+K}(\mathbf{x}; \mathbf{x}_{\text{new}}) \right)^+ \right]. \tag{4}$$

Remark 3.2. This criterion was inspired by the Expected Quantile Improvement presented in Picheny et al. [32], which would write here as $\mathbb{E}_n[(\min_{i \leq n} Q_n(\mathbf{x}_i) - Q_{n+1}(\mathbf{x}_{\text{new}}; \mathbf{x}_{\text{new}}))^+]$, but is modified in the spirit of knowledge gradient approaches from Frazier et al. [15] to account for improvements on the whole domain.

3.2. **Simulation-based computation of criteria**

Classically, in Sequential Uncertainty Reduction (SUR) (Bect et al. [2]) approaches, it is assumed that the function of interest is a realisation of a GP. Under these assumptions, several criteria enjoy (semi-)analytical forms, favouring criterion optimisation and the implementation of design strategies.

However, in our situation, it does not appear feasible to obtain a closed-form formula for the considered EQI criterion and we therefore estimate it via stochastic simulation.

In order to quantify the effect of adding an observation T_{n+1} at a given location $\mathbf{x}_{\text{new}}$, one needs to study the probability density of a new observation T at $\mathbf{x}$ conditioned on $\mathbf{T_n} = \mathbf{t_n}$. In favour of the law of total probability, and by considering Equation 3, one finds that it is given by:

$$\pi(t | \mathbf{T_n} = \mathbf{t_n}) \propto \int \pi(t | \boldsymbol{\epsilon}) \pi(\boldsymbol{\epsilon} | \mathbf{t_n}) \, d\boldsymbol{\epsilon} \propto \int \mathfrak{p}_{\mathbf{x}}(t; \boldsymbol{\epsilon}) \pi(\boldsymbol{\epsilon} | \mathbf{t_n}) \, d\boldsymbol{\epsilon}. \tag{5}$$

This motivates the following approach, which can be considered as a basic application of Sequential Monte Carlo (Doucet et al. [12]), where we use a simple simulation-based particle filter to approximate an unknown future quantity:

(1) The generative model given by the SLGP model is implemented within a MCMC framework and used to obtain N approximate realisations of $\boldsymbol{\varepsilon}|\mathbf{T}_n = \mathbf{t}_n$, noted $(\boldsymbol{\epsilon}_{(j)})_{1 \leq j \leq N}$.

The density of a new observation at $\mathbf{x}$ (See Equation 5) is approximated by the mixture $N^{-1} \sum_{j=1}^{N} \mathfrak{p}_{\mathbf{x}}(\cdot; \boldsymbol{\epsilon}_{(j)})$.

(2) The impact of adding K observations at a given location $\mathbf{x}_{\text{new}}$ is estimated by doing M simulations: For the i-th simulation, K realisations of the random variable $\widetilde{T}_{\text{new}}$ are independently drawn from the density $N^{-1} \sum_{j=1}^{N} \mathfrak{p}_{\mathbf{x}_{\text{new}}}(\cdot; \boldsymbol{\epsilon}_{(j)})$. The corresponding batch of response values is noted $\widetilde{\mathbf{t}}_{\text{new}}^{(i)} = (\widetilde{t}_{\text{new}}^{(i),1}, ..., \widetilde{t}_{\text{new}}^{(i),K})$. In light of Equation 5, the response density at $\mathbf{x}$ conditional on past data and on the simulated batch is given by $\pi(t|\mathbf{T_n} = \mathbf{t_n}, \widetilde{\mathbf{T}}_{\text{new}}^{(i)} = \widetilde{\mathbf{t}}_{\text{new}}^{(i)}) \propto \int \mathfrak{p}_{\mathbf{x}}(t; \boldsymbol{\epsilon}) \prod_{\ell=1}^{K} \mathfrak{p}_{\mathbf{x}_{\text{new}}}(\widetilde{t}_{\text{new}}^{(i),\ell}; \boldsymbol{\epsilon}) \pi(\boldsymbol{\epsilon}|\mathbf{t_n}) \, d\boldsymbol{\epsilon}$, leveraging the fact that the $\boldsymbol{\epsilon}_j$'s are drawn from the posterior, we can approximate the integral by the Monte Carlo sum $\sum_{j=1}^{N} \mathfrak{p}_{\mathbf{x}}(\cdot; \boldsymbol{\epsilon}_{(j)}) \omega_{i,j}$, with weights $\omega_{i,j}$ proportional to $\prod_{\ell=1}^{K} \mathfrak{p}_{\mathbf{x}_{\text{new}}}(\widetilde{t}_{\text{new}}^{(i),\ell}; \boldsymbol{\epsilon}_{(j)})$ and summing to one.

(3) The density field distribution after adding K observations at $\mathbf{x}_{\text{new}}$ is predicted by $M^{-1} \sum_{i=1}^{M} \sum_{j=1}^{K} \mathfrak{p}_{\mathbf{x}}(\cdot; \boldsymbol{\epsilon}_{(j)}) \omega_{i,j}$.

4. APPLICATIONS

For all the coming applications, we consider a zero mean GP $W_{\mathbf{x},t} = \sum_{j=1}^{p} \sqrt{\lambda_j} e_j(\mathbf{x}, t) \varepsilon_j$, $p \in \mathbb{N}$, with t and $\mathbf{x}$ being uni-variate. To ensure consistency with the rest of the document, we will keep the bold notation for $\mathbf{x}$. To define the e_j's, we start from bi-variate Fourier functions of order $q > 0$: sine and cosine of $2\pi(\omega_1 t + \omega_2 \mathbf{x})$, where ω_1 and ω_2 are integers satisfying $-q \leq \omega_1, \omega_2 \leq q$. Then, we remove redundant terms as well as those irrelevant in the SLGP setting (i.e. functions independent of t, that would be simplified with the normalisation of the process) and set the corresponding λ_j^2 to be $1/(1 + \omega_1 + \omega_2)$.

4.1. Some analytical applications

In the analytical applications, we have $D = \mathcal{I} = [0, 1]$ and consider four known density fields. The median functions of these fields appear in Figure 3 and are defined as $f_1(\mathbf{x}) = 0.25 \sin(16\mathbf{x}+9)+0.25 \sin(4.8\mathbf{x}+2.7)+0.625$ (minimum at $\mathbf{x}^* \approx 0.5095$) and $f_2(\mathbf{x}) = 0.15 + \frac{7}{72} \frac{1.1(10\mathbf{x}-5)^2 - 5(10\mathbf{x}-5)+6.1}{(10\mathbf{x}-5)^2+1}$ (minimum at $\mathbf{x}^* \approx 0.7414$).
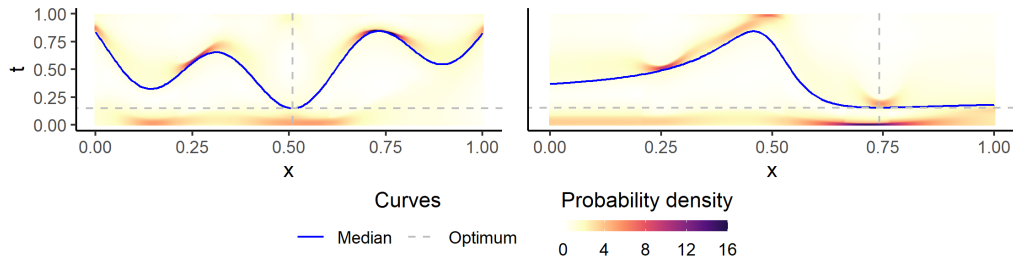

FIGURE 3. The two multi-modal reference density fields, their median and its global optimum.

Our first class of probability density fields, that we will refer to as "truncated Gaussian fields", writes as $h_{\mathbf{x}}\left(\frac{t - f_i(\mathbf{x})}{0.05}\right)$, $i \in \{1, 2\}$, with $h_{\mathbf{x}}$ a symmetrically truncated standard Gaussian with thresholds $\pm \min(f_i(\mathbf{x}), 1 - f_i(\mathbf{x}))$.

The truncation ensures that the distribution remains symmetrical around its median and mean f_i. The second case - that we will refer to as "multi-modal field"- is of the same form yet with $h_\mathbf{x}$ median-0 but multi-modal such as represented in Figure 3.

We perform density field estimation as presented in Section 2.3 for such reference fields for different sample sizes and order of the GP. In this section, we are not yet in an adaptive setting: each time a new observation is added to the model, its location is determined randomly, with uniform distribution over the index set. Figure 4 displays the posterior mean field with two sample sizes for the truncated Gaussian field with median f_1.
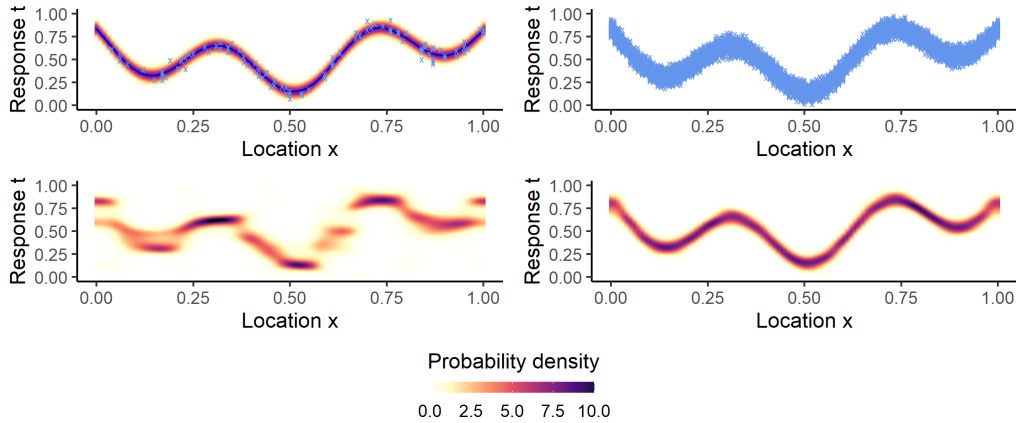

FIGURE 4. Results for the truncated Gaussian field with median f_1, using 121 basis functions. True field and samples used (top), posterior mean field (bottom) for a respective sample size of 100 (left) and 10000 (right).

We observe in Figure 4 that a higher sample size seems to yield a better estimation. In order to quantify the prediction error for different sample sizes and GP's order, we define an integrated squared Hellinger distance to measure dissimilarity between two probability density valued fields $p^i = (p^i_\mathbf{x})_{\mathbf{x} \in D}$ $(i = 1, 2)$:

$$d^2_{\text{ISH}}(p^1, p^2) = \int_D \int_\mathcal{I} \left(\sqrt{p^1_\mathbf{v}(u)} - \sqrt{p^2_\mathbf{v}(u)} \right)^2 du \, d\mathbf{v} \tag{6}$$

In Figure 5, we display the distribution of d^2_{ISH} between true and estimated fields for various sample sizes and SLGP orders. As both functions yielded close results, we show only the results for f_1. We see that the errors are comparable for small sample sizes. The order becomes limiting when more observations are available as those of the considered SLGPs relying on the smallest numbers of basis functions appear to struggle capturing small scale variations.
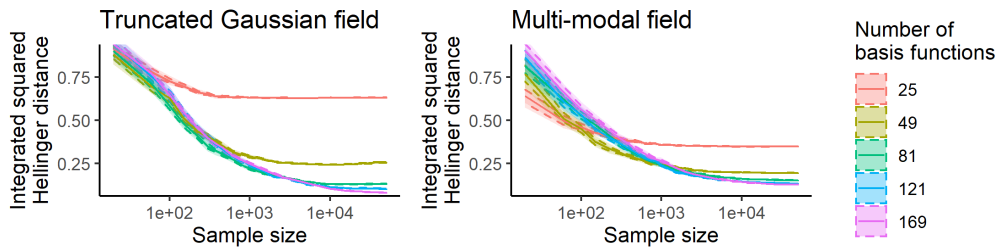

FIGURE 5. Integrated squared Hellinger distance distribution for different sample sizes and process orders, when the reference field has f_1 as its median.

4.2. **Test case: contaminant source localization under uncertain geology**

We now consider a hydro-geological test case in which, given breakthrough curves at several monitoring depths, we want to localize the source depth of a contaminant injected in a saturated aquifer whose geological properties are highly uncertain. We use the term *breakthrough map* to speak of the stacked breakthrough curves at prescribed monitoring depths. Also, denote by *observed* the given breakthrough map, even if throughout the section these observed maps stem from simulations with prescribed source depth and geology. This inference is performed by measuring the dissimilarity (*misfit*) between observed and simulated breakthrough maps, and our goal here is to uncover the depth that yields simulations with minimal median misfit. Here the index $\mathbf{x}$ stands for a candidate source depth while the response t represents the dissimilarity between simulated and observed breakthrough maps. The stochasticity of the response is due to our imperfect knowledge of the geological properties of the aquifer, as sketched on Figure 6 and further detailed below. Figure 7 shows the misfit values obtained by comparing 10000 simulations (200 replications at 50 different source depth values) to two different observed breakthrough maps, as well as the posterior mean field obtained with a SLGP of order 121.

To perform a flow simulation with a source at depth $\mathbf{x}$ while accounting for the uncertainty regarding the geological properties of the domain, we rely on plausible geological realisations. To keep the problem simple, the zone of interest of the aquifer is modelled as a 2D vertical section (10 meter deep and 5 meter wide) aligned with the main flow direction. For each simulation performed, the plausible geological realisation used is a multiple-point statistics realisation generated with the Deesse algorithm (Mariethoz et al. [28]) that reproduce the complex features of braided-river aquifer models (Pirot et al. [35]). Then, the contaminant flow and transport is simulated under steady-state flow and fixed boundary conditions (constant hydraulic gradient) using the Maflot Matlab code (Künze and Lunati [23]).
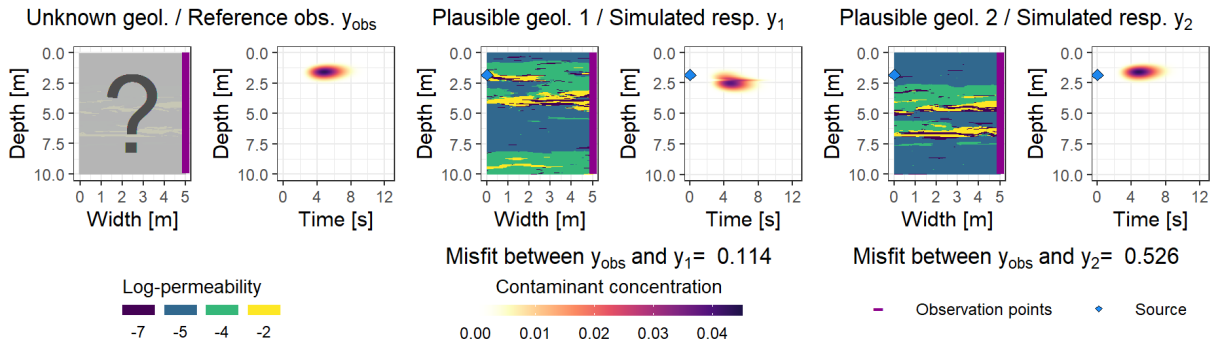

FIGURE 6. Geological medium versus breakthrough map in different situations. From left to right: the first pair of plots depicts the starting situation our inverse problem of recovering the source location given an observed breakthrough map (right) yet under unknown geology (left). The second and third pairs of plots represent simulated breakthrough maps under randomly generated geologies, and the associated misfit values with respect to the observed breakthrough map. For simplicity, the same source depth (1.84m) is assumed in all three cases.

After running such a simulation for a contaminant released at time 0, width 0 and depth $\mathbf{x}$, we obtain a concentration breakthrough map (such as depicted in Figure 6 under three different scenarios). Here 100 candidate observation depths, regularly spaced between 0m and 10m, are considered. Each breakthrough curve, i.e., section of a breakthrough map at a a prescribed depth, describes the contaminant concentration observed at width 5m for varying times (discretised over 101 points regularly spaced between 0s and 12.5s). Misfits between observed and simulated curves are defined here in terms of a re-scaled 2-norm distance.

Throughout the section and as illustrated in Figure 7, we use two reference fields of misfit distributions that are obtained by fitting SLGP models to misfit data produced under the two latent geologies of Figure 6.
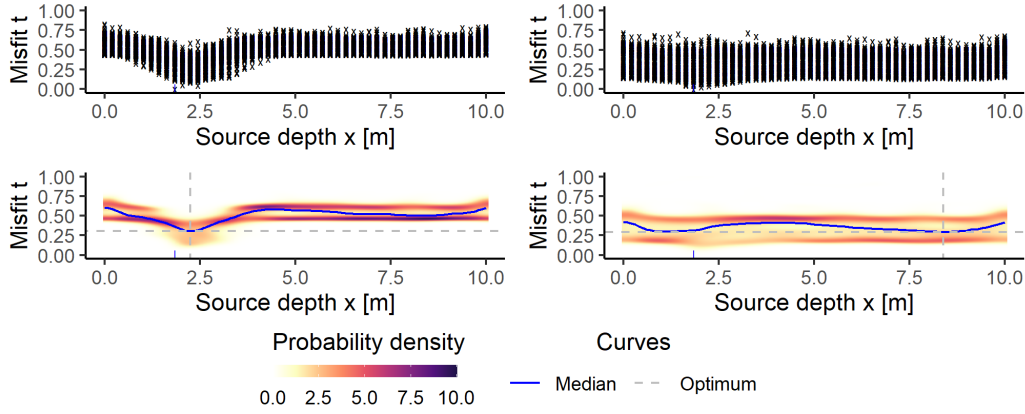
FIGURE 7. Misfit data (top) and posterior mean field (bottom) for two latent geological structures.

Using these two reference density fields to draw new samples, we follow the methodology introduced in part 3 and represent in Figure 8 the posterior mean field and its estimated median before and after running 25 steps of the algorithm on the first geological structure.
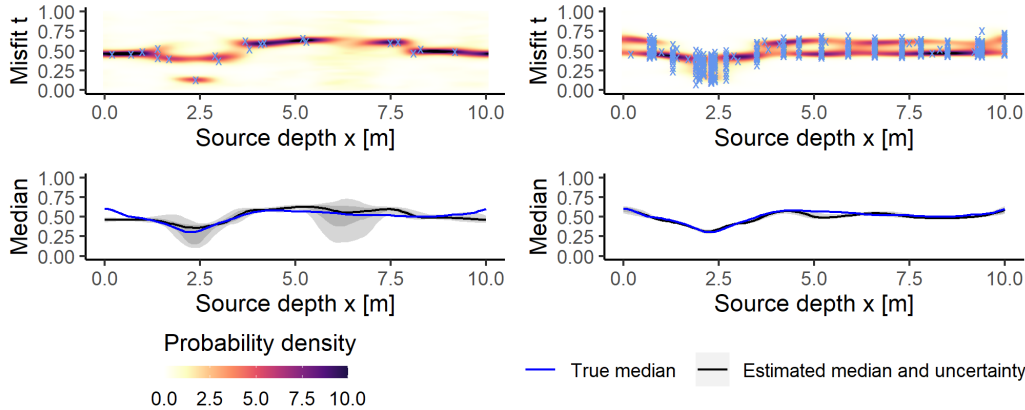

FIGURE 8. Results at the beginning of the algorithm [left] and after the 25th step [right]. Mean field estimated and samples available [top]; Estimated VS reference median [bottom].

In these settings, the global minimum (at 2,24m) is easily found and the algorithm focuses on improving its estimation of the median. This corresponds to an exploitation-oriented approach. Although not displayed here, we found out, as reflected in Figure 9 that with the other geological structure, our approach was also able to locate the minimum, this time by focusing on exploring by adding observations at new locations of interest.

4.3. Optimisation benchmark

For our small benchmark, the starting design consists in $n = 20$ data points $(\mathbf{x}_i, t_i)$ from the reference fields heterogeneously scattered across space. Their location are independent uniformly distributed across parameter space. At each step, observations are added in batches of 20 at the same location. We repeat 24 independent instances of the optimisation process for each strategy and each application and compare the performances in term of optimality gap (difference between real and estimated optimal medians). This approach is favoured due to the relatively high cost of one evaluation of the EQI criterion for SLGP, but we expect it to be detrimental to

our GP-based competitors, as GPs would benefit more from having scattered observations rather than batches scattered over different points.

We compare different strategies for modelling the field and choosing the next sampling location. The value of the minimiser is inferred by modelling the function of interest with one of three models: the first one, that will be called homoscedastic GP consists in a GP regression where the observation noise level is assumed to be uniform throughout the domain. The second one, a GP regression with input dependent noise rates as in Kersting et al. [22], Binois et al. [5] will be called heteroscedastic GP. The last one is the SLGP model.
For each of these three models, we compare a non-adaptive approach (at each step, new observations are added at a location chosen uniformly at random) to an adaptive approach. The criteria used are: Approximated Knowledge Gradient (AKG) as implemented in the R package DiceOptim version 2.0.1 [34] for the homoscedastic GP, the Expected improvement, as implemented in the R package hetGP version 1.2.1 [4] for the heteroscedastic GP, and the criterion 4 of part 3 for the SLGP. The results of the benchmark are shown in Figure 9.
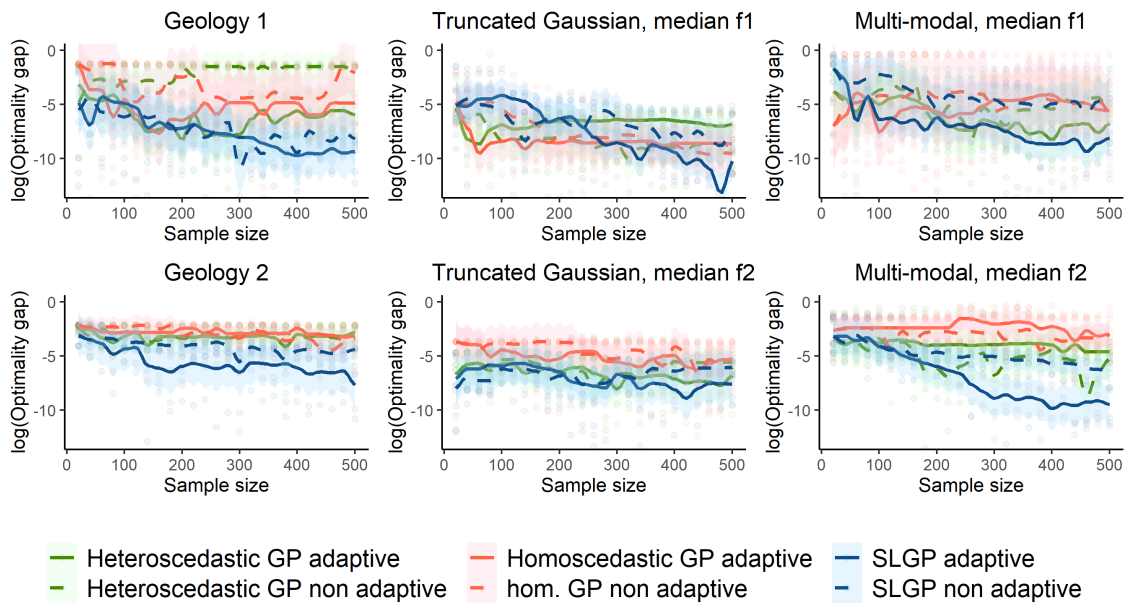

FIGURE 9. Median of the log optimality gap for the 6×6 considered strategies and test cases.

Hyper-parameters of the two GP's are estimated by maximum likelihood. For the adaptive SLGP, we decided to display the results obtained when minimising a 90% quantile of the random median, as our first experiments showed no strong sensitivity to the chosen level α. On a personal laptop with 16Go RAM and a simple R implementation, performing the SLGP density estimation took between 13 and 24 minutes depending on the sample size, and inferring the EQI with 150 simulations for 101 locations took 5 more minutes.
One notices that in the most complex situations, the sampling scheme based on GP modelling of the functional performs worse than the approaches based on SLGP modelling. We found out that GP based approaches tends to get trapped in local optima. This might be explained by the credit allocation.

5. CONCLUSION AND PERSPECTIVES

We demonstrated how a spatial generalization of the Logistic Gaussian Process model can be used for sample-based modelling of distribution fields, and how the resulting probabilistic predictions of distribution

fields can be leveraged for moderate-dimensional stochastic optimization problems under unknown heterogeneous noise distributions. In particular, we introduced a simple sequential Monte Carlo-based approach to quantify the impact of sampling at a location and showed, on a modest benchmark, that it enables implementing sequential design strategies trading off between exploration and exploitation.

The conducted experiments yielded promising results. However, further work is needed to more extensively compare algorithms with respect to the employed meta-models and criteria, respectively, and identify which of these degrees of freedom are most influential on performances. Tuning model settings and strategies appropriately could ultimately lead to substantial efficiency gains, be it towards stochastic optimization or broader uncertainty reduction goals. In particular, SLGP would benefit from more systematic model selection and parameter estimation approaches.

Upcoming research directions encompass making the SLGP model more suitable for higher dimensions and reducing associated computational costs, assessing the performance of the resulting probabilistic predictions, and deriving further uncertainty reduction strategies. It is also of interest to investigate theoretical properties of SLGP models and considered sequential strategies, notably in view of underlying mean and covariance functions.

Acknowledgements

References

[1] J. Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society, Series B: Methodological*, 44(2):139–177, 1982.

[2] J. Bect, F. Bachoc, and D. Ginsbourger. A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli*, 25(4A):2883–2919, 2019.

[3] F. Bellini and E. Di Bernardino. Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506, 2017.

[4] M. Binois and R. B. Gramacy. *hetGP: Heteroskedastic Gaussian Process Modeling and Design under Replication*, 2019. URL `https://CRAN.R-project.org/package=hetGP`. R package version 1.1.2.

[5] M. Binois, R. B. Gramacy, and M. Ludkovski. Practical heteroscedastic gaussian process modeling for large simulation experiments. *Journal of Computational and Graphical Statistics*, 27(4):808–821, 2018.

[6] M. Binois, J. Huang, R. B. Gramacy, and M. Ludkovski. Replication or exploration? Sequential design for stochastic simulation experiments. *Technometrics*, 61(1):7–23, 2019.

[7] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, 2011.

[8] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

[9] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[10] S. Cotter, G. Roberts, A. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446, 2013.

[11] C. Donner and M. Opper. Efficient Bayesian inference for a Gaussian process density model. In A. Globerson and R. Silva, editors, *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 53–62. AUAI Press, 2018.

[12] A. Doucet, N. De Freitas, and N. Gordon. *Sequential Monte Carlo methods in practice*. Springer, New York, NY, 2001.

[13] S. Efromovich. Dimension reduction and adaptation in conditional density estimation. *Journal of the American Statistical Association*, 105(490):761–774, 2010.

[14] J. J. Egozcue, J. Díaz–Barrero, and V. Pawlowsky-Glahn. Hilbert space of probability density functions based on Aitchison geometry. *Acta Mathematica Sinica, English Series*, 22(4):1175–1182, 2006.

[15] P. Frazier, W. Powell, and S. Dayanik. The knowledge-gradient policy for correlated normal beliefs. *INFORMS journal on Computing*, 21(4):599–613, 2009.

[16] P. I. Frazier. A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

[17] P. Hall, J. Racine, and Q. Li. Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(468):1015–1026, 2004.

[18] J. M. Hernández-Lobato, M. W. Hoffman, and Z. Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. 27:918–926, 2014.

[19] H. Jalali, I. Nieuwenhuyse, and V. Picheny. Comparison of Kriging-based algorithms for simulation optimization with hetero-geneous noise. *European Journal of Operational Research*, 261(1):279–301, 2017.

[20] J. Janusevskis and R. Le Riche. Simultaneous kriging-based estimation and optimization of mean response. *Journal of Global Optimization*, 55(2):313–336, 2013.

[21] D. Jones, M. Schonlau, and W. Welch. Efficient Global Optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[22] K. Kersting, C. Plagemann, P. Pfaff, and W. Burgard. Most likely heteroscedastic gaussian process regression. In *Proceedings of the 24th International Conference on Machine Learning*, ICML'07, pages 393–400, 2007.

[23] R. Künze and I. Lunati. A matlab toolbox to simulate flow through porous media. Technical report, University of Lausanne, Switzerland, 2011.

[24] P. J. Lenk. The logistic normal distribution for Bayesian, nonparametric, predictive densities. *Journal of the American Statistical Association*, 83(402):509–516, 1988.

[25] P. J. Lenk. Towards a practicable Bayesian nonparametric density estimator. *Biometrika*, 78(3):531–543, 1991.

[26] T. Leonard. Density estimation, stochastic processes and prior information. *Journal of the Royal Statistical Society, Series B: Methodological*, 40(2):113–132, 1978.

[27] S. Mandt, M. D. Hoffman, and D. M. Blei. Stochastic gradient descent as approximate bayesian inference. *The Journal of Machine Learning Research*, 18(1):1–35, 2017.

[28] G. Mariethoz, P. Renard, and J. Straubhaar. The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46(11):W11536, 2010.

[29] J. Močkus, V. Tiesis, and A. Žilinskas. The application of Bayesian methods for seeking the extremum. *Towards global optimization*, 2:117–129, 1978.

[30] V. Moutoussamy, S. Nanty, and B. Pauwels. Emulators for stochastic simulation codes. *ESAIM: Proceedings and Surveys*, 48:116–155, 2015.

[31] I. Murray, D. MacKay, and R. P. Adams. The gaussian process density sampler. In *Advances in Neural Information Processing Systems*, volume 21, pages 9–16, 2009.

[32] V. Picheny, D. Ginsbourger, Y. Richet, and G. Caplin. Quantile-based optimization of noisy computer experiments with tunable precision. *Technometrics*, 55(1):2–13, 2013.

[33] V. Picheny, T. Wagner, and D. Ginsbourger. A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626, 2013.

[34] V. Picheny, D. Ginsbourger, and O. Roustant. *DiceOptim: Kriging-Based Optimization for Computer Experiments*, 2020. URL https://CRAN.R-project.org/package=DiceOptim. R package version 2.0.1.

[35] G. Pirot, J. Straubhaar, and P. Renard. A pseudo genetic model of coarse braided-river deposits. *Water Resources Research*, 51(12):9595–9611, 2015.

[36] A. Prékopa. *Stochastic programming*. Springer Science & Business Media, 2013.

[37] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, 2006.

[38] J. Risk and M. Ludkovski. Sequential design and spatial modeling for portfolio tail risk measurement. *SIAM Journal on Financial Mathematics*, 9(4):1137–1174, 2018.

[39] H. Robbins and S. Monro. A stochastic approximation method. *The annals of mathematical statistics*, 22(3):400–407, 1951.

[40] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2(3):21–42, 2000.

[41] M. Rostek. Quantile maximization in decision theory. *The Review of Economic Studies*, 77(1):339–371, 2010.

[42] A. Ruszczyński and A. Shapiro. Stochastic programming models. volume 10, pages 1–64. Elsevier, 2003.

[43] N. Srinivas, A. Krause, S. Kakade, and M. Seeger. Gaussian Process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning*, ICML'10, page 1015–1022, 2010.

[44] R. Talská, A. Menafoglio, J. Machalová, K. Hron, and E. Fišerová. Compositional regression with functional response. *Computational Statistics & Data Analysis*, 123:66–85, 2018.

[45] W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

[46] S. T. Tokdar. Towards a faster implementation of density estimation with Logistic Gaussian Process priors. *Journal of Computational and Graphical Statistics*, 16(3):633–655, 2007.

[47] S. T. Tokdar and J. K. Ghosh. Posterior consistency of logistic Gaussian process priors in density estimation. *Journal of Statistical Planning and Inference*, 137(1):34–42, 2007.

[48] S. T. Tokdar, Y. M. Zhu, and J. K. Ghosh. Bayesian density regression with logistic gaussian process and subspace projection. *Bayesian Analysis*, 5(2):319–344, 2010.

[49] L. Torossian, V. Picheny, R. Faivre, and A. Garivier. A review on quantile regression for stochastic computer experiments. *Reliability Engineering & System Safety*, 201(C):106858, 2020.

[50] X. Zhu and B. Sudret. Replication-based emulation of the response distribution of stochastic simulators using generalized lambda distributions. *International Journal for Uncertainty Quantification*, 10(3):249–275, 2020.