Robust calibration of numerical models based on relative regret

Victor Trappler[a], Élise Arnaud[1], Arthur Vidard[1], Laurent Debreu[1]

[a] *Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP*, LJK, 38000 Grenoble, France*

Abstract

Classical methods of parameter estimation usually imply the minimisation of an objective function, that measures the error between some observations and the results obtained by a numerical model. In the presence of random inputs, the objective function becomes a random variable, and notions of robustness have to be introduced.

In this paper, we are going to present how to take into account those uncertainties by defining a notion of robustness based on the distribution of the conditional minimisers and compare it with the minimum in the mean sense, and the minimum of variance.

Keywords:

Introduction

Numerical models are widely used to study or forecast natural phenomenon and improve industrial processes. However, by essence models only partially represents reality and sources of uncertainties are ubiquitous (discretisation errors, missing physical processes, poorly known boundary conditions). Moreover, such uncertainties may be of different nature. [1] propose two to consider two categories of uncertainties. On the one hand aleatoric uncertainties, coming from the inherent variability of a phenomenon, e.g. intrinsic randomness of some environmental variables. On the other hand, epistemic uncertainties coming from a lack of knowledge about the properties and conditions of the phenomena underlying the behaviour of the system under study. The latter can be accounted for through the introduction of ad-hoc correcting terms in the numerical model, that need to be properly estimated. Thus, reducing the epistemic uncertainty can be done through parameters estimation approaches. This is usually done using optimal control techniques, leading to an optimisation of a well chosen cost function which is typically built as a comparison with reference observations. An application of such an approach, in the context of ocean circulation modeling, is the estimation of ocean bottom friction parameters in [2] and [3].

If parameters to be estimated are not the only source of uncertainties, their optimal control is doomed to overfit the data, *e.g* to artificially introduce errors in the controlled parameter to compensate for other sources. If such uncertainties are of aleatoric nature, then the parameter estimation is only optimal for the observed situation, and may be

Email address: victor.trappler@univ-grenoble-alpes.fr (Victor Trappler)

very poor in other configurations, phenomenon coined as *localized optimization* in [4]. Taking into account aleatoric uncertainties in optimisation problems takes several names, including *robust optimisation*, *robust design* in [5], or *optimization under uncertainties* [6, 7, 8].

Let's denote $\mathbf{k} \in \mathbb{K}$ the parameter set to be estimated, to reduce epistemic errors. The aleatoric uncertainties are modelled as a random vector $\mathbf{U}$ whose sample space is $\mathbb{U}$. The probability measure of $\mathbf{U}$ is $\mathbb{P}_{\mathbf{U}}$, and its density, if it exists, is $p_{\mathbf{U}}$. The cost function $J(\mathbf{k}, \mathbf{U})$ is a random variable in this context. It is most often defined as the squared norm of a given function $\mathcal{G}(\mathbf{k}, \mathbf{U})$

$$J(\mathbf{k}, \mathbf{U}) = \frac{1}{2}\|\mathcal{G}(\mathbf{k}, \mathbf{U})\|^2 \tag{1}$$

For instance, in data assimilation, J describes a distance between the output of the numerical model and given observed data, plus generally some regularization terms. For a practical purpose, we assume that $\forall\ \mathbf{u}$ a realisation of $\mathbf{U}$, finding $\mathbf{k}_{\mathbf{u}}$ that minimises the cost function $J(\mathbf{k}, \mathbf{U} = \mathbf{u})$ is a well-posed problem, and that the optimum is unique. Additionally, the following assumption are made: the cost function is strictly positive, and $\forall\ \mathbf{k} \in \mathbb{K}$, the random variable $J(\mathbf{k}, \mathbf{U})$ has finite first- and second-order moments.

In this paper, we aim at finding $\hat{\mathbf{k}}$ a *robust* estimator of $\mathbf{k}$. The definition of robustness differs depending on the context in which it is used. Indeed, one definition of the robustness of an estimate is a measure of the sensibility of said estimate to outliers [9]. This lead to the introduction of robust norms in data assimilation [10]. In a Bayesian framework, robustness may refer to the sensitivity to a wrong specification of the priors [11]. Throughout this paper, robust has to be understood as satisfactory for a broad range of $\mathbf{u}$, and/or as insensitive as possible to uncertainties encompassed in $\mathbf{U}$.

The usual practice consists in neglecting the variability of $\mathbf{U}$ by setting it to an *a priori* value $\mathbf{u}^b$. In this case, $\hat{\mathbf{k}}$ is set to the optimum $\mathbf{k}_{\mathbf{u}^b}$ of $J(\mathbf{k}, \mathbf{U} = \mathbf{u}^b)$. There is no guarantee on the performance of $\hat{\mathbf{k}}$ if the calibrated model is used for predictions, as the estimated value will compensate the error made by a possibly wrong specification of $\mathbf{u}^b$. In a data assimilation context, this situation appears if $\mathbf{u}^b$ does not properly represent the conditions on which the observations have been obtained. Another strategy, that consists in minimising J over the joint space $\mathbb{K} \times \mathbb{U}$, is not always possible or relevant. The complexity of the optimisation is increased, and the computed estimation of $\hat{\mathbf{k}}$ has no reason to be robust in the end: this kind of method does not take into consideration the variability of the uncertain variable. The worst-case approach [12] is another popular method, and is based on the minimisation with respect to $\mathbf{k}$ of the maximum of the cost function for $\mathbf{u} \in \mathbb{U}$: $\min_{\mathbf{k}} \max_{\mathbf{u}} J(\mathbf{k}, \mathbf{U} = \mathbf{u})$. This approach may yield over-conservative solutions, and does not take into account the random nature of $\mathbf{U}$.

Accounting for the probabilistic nature of $\mathbf{U}$ leads to study the distribution of the random variable $J(\mathbf{k}, \mathbf{U})$, or the distribution of its minimisers $\mathbf{k}_{\mathbf{U}}$. The latter is referred as the distribution of the conditional minimisers, notion that appeared notably in [13] and in [14] for a global optimisation purpose. Both approaches and related robust estimates are described in Section 1. Section 2 introduces a new class of estimators, by relaxing the constraint of optimality and defining regions of acceptability, similarly as [15] in discrete combinatorial problems. The intention is that a robust estimate provides values of the cost function close enough to the attainable minimum for each configuration induced by

$\mathbf{u} \in \mathbb{U}$. Illustration of the various discribed methods are given on a numerical exemple in Section 3.

1. Classical robust estimators

As mentioned before, robustness can be understood as satisfactory for a broad range of $\mathbf{u}$, and/or as insensitive as possible to uncertainties encompassed in $\mathbf{U}$. Under this definition, one may design classical robust estimators, either by optimising the moments of the cost function; or based on the distribution of its minimisers.

1.1. Optimisation of the moments

kmean ou kexpectation ?

Let us define $\mu(\mathbf{k})$ and $\sigma^2(\mathbf{k})$, the expected value and the variance of the cost variable for a given $\mathbf{k}$ as

$$\mu(\mathbf{k}) = \mathbb{E}_{\mathbf{U}}\left[J(\mathbf{k}, \mathbf{U})\right] = \int_{\mathbb{U}} J(\mathbf{k}, \mathbf{u}) p_{\mathbf{U}}(\mathbf{u})\, \mathrm{d}\mathbf{u} \tag{2}$$

$$\sigma^2(\mathbf{k}) = \mathbb{V}\mathrm{ar}_{\mathbf{U}}\left[J(\mathbf{k}, \mathbf{U})\right] = \int_{\mathbb{U}} \left(J(\mathbf{k}, \mathbf{u}) - \mu(\mathbf{k})\right)^2 p_{\mathbf{U}}(\mathbf{u})\, \mathrm{d}\mathbf{u} \tag{3}$$

Minimising the expectation leads to the estimate $\mathbf{k}_{\mathrm{mean}}$ defined by:

$$\mathbf{k}_{\mathrm{mean}} = \underset{\mathbf{k}\in\mathbb{K}}{\arg\min}\ \mu(\mathbf{k}) \tag{4}$$

In order to take into account the spread around the mean value, one can choose to minimise the variance, leading to $\mathbf{k}_{\mathrm{var}}$:

$$\mathbf{k}_{\mathrm{var}} = \underset{\mathbf{k}\in\mathbb{K}}{\arg\min}\ \sigma^2(\mathbf{k}) \tag{5}$$

A lot of different methods are readily available to solve these minimisation problems. For instance, stochastic Sample Approximation [16, 17] is based on a finite and fixed set of samples $\{\mathbf{u}^i\}_{i=1\ldots N}$ of $\mathbf{U}$. The estimations at a given $\mathbf{k}$ are computed using standard Monte Carlo, resulting in the following optimisation problems:

$$\hat{\mathbf{k}}_{\mathrm{mean}} = \underset{\mathbf{k}\in\mathbb{K}}{\arg\min} \sum_{i=1}^{N} J(\mathbf{k}, \mathbf{u}^i) \tag{6}$$

and

$$\hat{\mathbf{k}}_{\mathrm{var}} = \underset{\mathbf{k}\in\mathbb{K}}{\arg\min} \sum_{i=1}^{N} \left(J(\mathbf{k}, \mathbf{u}^i) - \sum_{j=1}^{N} J(\mathbf{k}, \mathbf{u}^j) \right)^2 \tag{7}$$

If computationally affordable, one can perform these estimations on a regular grid on $\mathbb{K} \times \mathbb{U}$. In case of expensive computer code, one can build a meta model to ease the minimisation, such as Gaussian processes [18]. Even though $\mathbf{k}_{\mathrm{mean}}$ is a reasonable choice, there is no guarantee that $J(\mathbf{k}_{\mathrm{mean}}, \mathbf{U} = \mathbf{u})$ will not reach catastrophic level for some $\mathbf{u}$. On the other hand, using $\mathbf{k}_{\mathrm{var}}$ will ensure stability of the cost function, but whithout

any control of its level of quality. Ideally, one would want to have a small mean value, and a small variance as well. Multi-objective optimisation is a proper tool to deal with these simultaneous and sometimes concurrent objectives, for exemple by computing the Pareto front of $\big(\mu(\mathbf{k}), \sigma^2(\mathbf{k})\big)$ as done in [19].

As the computation of this Pareto front is usually hard and expensive, alternative strategies based on the minimisation of a scalarized version of the vector of objectives are often considered. Some are based on a weighted sum of the objectives, as presented in [20] and in [21], while some others are based on the minimisation of one of the objectives under constraints on the others, as performed in [22]. Both of these methods are based on an *delicate* choice of weights or of constraints before any computation. This choice relies heavily on a knowledge of the properties of the cost function.

To summarise, even though the notions of mean and variance are quite easily understood, getting a satisfactory estimator is not that straightforward. One could instead consider how often a particular value $\mathbf{k}$ is a minimiser of the cost function, leading to the notion of most probable estimate, as explained in the next subsection.

mean ou expectation ?

1.2. Most probable estimate

Let us consider the minimal attainable cost in each configuration brought by $\mathbf{u}$. The resulting conditional minimum is denoted as J^*:

$$J^* : \mathbf{u} \in \mathbb{U} \longmapsto J^*(\mathbf{u}) = \min_{\mathbf{k}\in\mathbb{K}} J(\mathbf{k}, \mathbf{U} = \mathbf{u}) \tag{8}$$

Similarly, the function of conditional minimisers can then defined by:

$$\mathbf{k}^* : \mathbf{u} \in \mathbb{U} \longmapsto \mathbf{k}^*(\mathbf{u}) = \mathbf{k}_\mathbf{u} = \arg\min_{\mathbf{k}\in\mathbb{K}} J(\mathbf{k}, \mathbf{U} = \mathbf{u}) \tag{9}$$

Using this function, we can define the corresponding random variable $\mathbf{K}^*$ as

$$\mathbf{K}^* = \mathbf{k}^*(\mathbf{U}), \tag{10}$$

and its associated density function $p_{\mathbf{K}^*}(\mathbf{k})$, that will be further referred as the density of minimisers. The mode of this density is called the Most Probable Estimate (MPE) and is noted $\mathbf{k}_{\text{MPE}}$:

$$\mathbf{k}_{\text{MPE}} = \arg\max_{\mathbf{k}\in\mathbb{K}} p_{\mathbf{K}^*}(\mathbf{k}) \tag{11}$$

To give some intuition on this estimate, let us imagine that the distribution of minimisers is a dirac centered on $\mathbf{k}_{\text{MPE}}$. Then it would mean that this estimate is the minimiser of the cost function whatever the realisation of the uncertain variable, therefore optimal in all conditions. If the distribution $p_{\mathbf{K}^*}$ is heavily dominated by a single value, the MPE may be a good candidate for robust control. This is not so obvious in case of a multimodal distribution. In general, an analytical form of $p_{\mathbf{K}^*}$ is usually impossible to obtain, so an estimation $\hat{p}_{\mathbf{K}^*}$ must be used, and its maximum computed to get the MPE.

Once again, a set of samples $\{\mathbf{u}^i\}_{i=1\dots N}$ can be used to compute the set $\{\mathbf{k}_{\mathbf{u}^i}\}_{i=1\dots N}$, from which one can approximate $p_{\mathbf{K}^*}$. The resulting approximation and therefore its mode, is sensitive to the density estimation method. Main methods are KDE (Kernel Density Estimation) [**?**], and EM (Expectation-Maximisation) [23].

KDE is a non-parametric estimation technique based on the use of a kernel function f. Assuming an isotropic kernel, the estimation has the following form:

$$\hat{p}_{\mathbf{K}^*}(\mathbf{k}) = \frac{1}{N h^{\dim \mathbb{K}}} \sum_{i=1}^{N} f\left(\frac{\mathbf{k} - \mathbf{k}_{\mathbf{u}^i}}{h}\right) \tag{12}$$

where h is the bandwidth. In a multidimensional setting, one usually consider a kernel based on the product of 1D kernels, applied independently to all components: $f(\mathbf{k}) = \prod_{j=1}^{\dim \mathbb{K}} f_{1\mathrm{D}}(k^{(j)})$ where $k^{(j)}$ is the j-th component of $\mathbf{k}$. There is wide choice of available $f_{1\mathrm{D}}$, such as a rectangle function, $f_{1\mathrm{D}}(x) = 1$ if $|x| < \frac{1}{2}$ and 0 elsewhere, or a Gaussian kernel $f_{1\mathrm{D}}(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. One can notice that the rectangle function as a kernel yields non-smooth densities, very similar to the histogram of the minimisers.

The EM algorithm can also be used to estimate the density, by minimising the statistical distance between the empirical distribution and a mixture of ν Gaussian densities. The estimation has then the following form:

$$\hat{p}_{\mathbf{K}^*}(\mathbf{k}) = \sum_{i=1}^{\nu} \pi_i \phi(\mathbf{k}; \mathbf{m}_i, \mathbf{\Sigma}_i) \tag{13}$$

where $\phi(\cdot; \mathbf{m}, \mathbf{\Sigma})$ is the probability density function of the normal distribution of mean $\mathbf{m}$ and covariance matrix $\mathbf{\Sigma}$, and $\{\pi_i\}_{i=1\ldots\nu}$ are the mixing coefficients.

Despite the fact that those methods are well established, using them in a plug-in approach has some flaws. One of the basic assumption of density estimation is to assume that $\mathbf{K}^*$ is a continuous random variable, hypothesis that may be violated. Worse, the notion of mode is not well defined when the distribution of the minimisers is a discrete-continuous mixture. This may result in inconsistent estimations of $\hat{\mathbf{k}}_{\mathrm{MPE}}$ when using different methods as illustrated in next subsection.

1.3. Numerical illustration

Before going further in the explanation of our approach, let us illustrate the nature of the 3 previously detailled estimator, $\hat{\mathbf{k}}_{\mathrm{mean}}, \hat{\mathbf{k}}_{\mathrm{var}}$ and $\hat{\mathbf{k}}_{\mathrm{MPE}}$ on two analytical cost functions. These functions are based on the Branin-Hoo's function, slightly modified to ensure strict positivity:

$$\mathrm{BH}(x_1, x_2) = \frac{1}{51.95}\left[\left(\bar{x}_2 - \frac{5.1\bar{x}_1^2}{4\pi^2} + \frac{5\bar{x}_1}{\pi} - 6\right)^2 + \left(10 - \frac{10}{8\pi}\right)\cos(\bar{x}_1) - 44.81\right] + 2 \tag{14}$$

$$\text{with } \bar{x}_1 = 3x_1 - 5, \quad \bar{x}_2 = 3x_2 \tag{15}$$

Using Eq. (14), we define the two cost functions on $\mathbb{K} \times \mathbb{U} = [0,5] \times [0,5]$ as:

$$J_{\mathrm{BH}} : (\mathbf{k}, \mathbf{u}) \mapsto \mathrm{BH}(\mathbf{k}, \mathbf{u})$$
$$J_{\mathrm{BHswap}} : (\mathbf{k}, \mathbf{u}) \mapsto \mathrm{BH}(\mathbf{u}, \mathbf{k})$$

Even though the functions are quite similar, the asymmetric roles of $\mathbf{k}$ and $\mathbf{u}$ cause different behaviour.

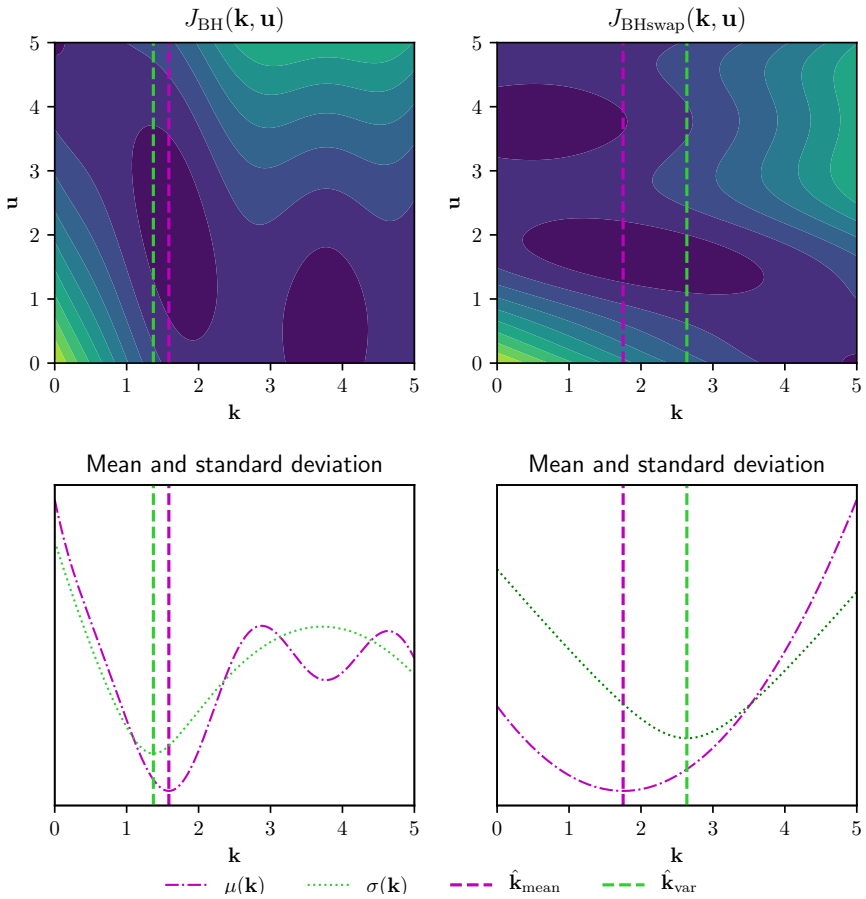

Figure 1: Left: Contour of the J_{BH} function, conditional mean and variance. The estimates $\hat{\mathbf{k}}_{\mathrm{mean}}$ and $\hat{\mathbf{k}}_{\mathrm{var}}$ are plotted with the dashed line. Right: J_{BHswap} function. The scale of the bottom plots have not been represented.

The random variable $\mathbf{U}$ is assumed to be uniformly distributed over $\mathbb{U}$. The estimations are based on a 1000×1000 regular grid over $\mathbb{K} \times \mathbb{U}$. Both cost functions are shown on Figure 1, top row.

The left, respectively right, column stands for J_{BH}, respectively J_{BHswap} function. Functions $\mu(\mathbf{k})$ and $\sigma(\mathbf{k})$ are drawn on the bottom row, respectively in purple and green. The corresponding minimizers $\hat{\mathbf{k}}_{\mathrm{mean}}$ and $\hat{\mathbf{k}}_{\mathrm{var}}$ are also plotted. On this figure, we can observe that $\hat{\mathbf{k}}_{\mathrm{mean}}$ and $\hat{\mathbf{k}}_{\mathrm{var}}$ are close for the J_{BH} function, while have significant different value for the J_{BHswap} function.

Illustrations related to $\hat{\mathbf{k}}_{\mathrm{MPE}}$ are depicted on Figure 2. As before, the set of $\mathbf{u}^i$ is taken from a regular grid. Top row shows the contour plot of both functions as well as the set of conditional minimisers $\{\mathbf{k}_{\mathbf{u}^i}\}_{1 \leq i \leq N}$ in red, as defined in Eq. (9). The left, respectively right, column stands for J_{BH}, respectively J_{BHswap} function. The bottom rows presents three approximations of the density of minimisers: the histogram in grey (bin size selected

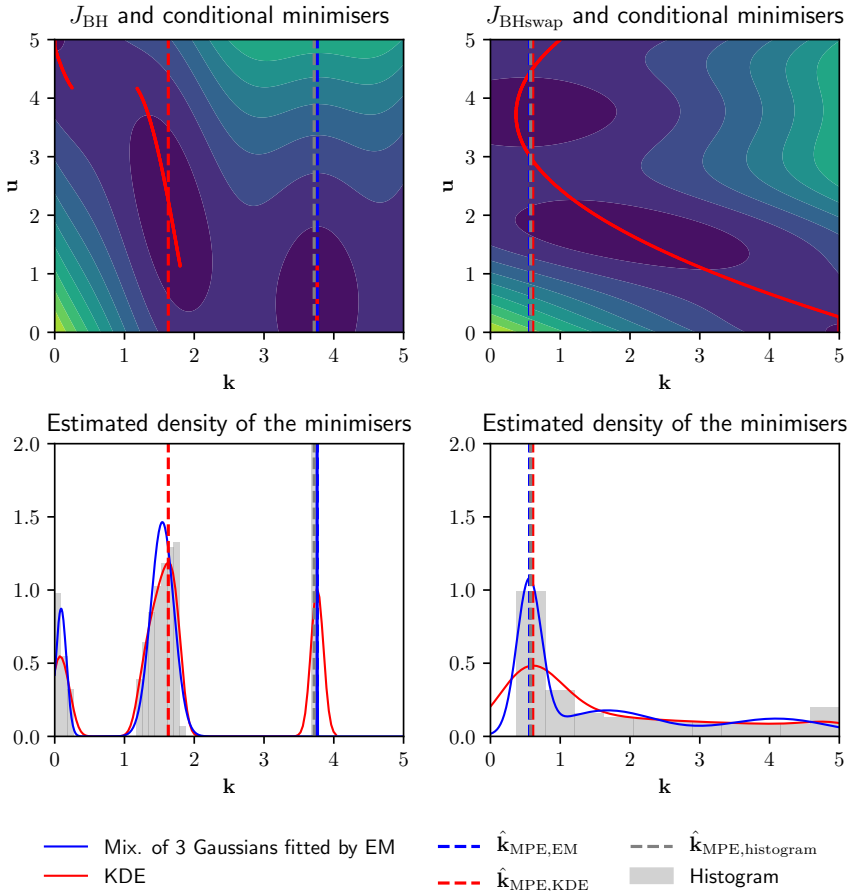

Figure 2: Top Left: J_{BH} function along with conditional minimisers in red. Bottom left: Estimated density using KDE and EM algorithm and conditional mean. The dashed lines indicate the MPE found using those methods. Right: Same quantities with J_{BHswap}. The estimation of the MPE is consistent between the methods for J_{BHswap}, but not for J_{BH}

using Freedman-Diaconis from [24]), the result of a kernel density estimation (KDE) with Gaussian kernels in red (using Scott's rule from [25] for bandwidth selection), and the estimation by a Gaussian mixture. The latter has been calculated with the EM algorithm. The number of Gaussians has been fixed to 3, a guess based on the general shape of the histogram. Respective estimations of $\hat{\mathbf{k}}_{\mathrm{MPE}}$ are also depicted using dashed lines.

For the J_{BHswap} function, we can observe that those three methods give coherent results, as $\hat{\mathbf{k}}_{\mathrm{MPE,KDE}} = \hat{\mathbf{k}}_{\mathrm{MPE,EM}} = \hat{\mathbf{k}}_{\mathrm{MPE,histogram}} \approx 0.8$. This is not the case for the J_{BH} function however. Using Kernel density estimation (Gaussian), the estimation of $\mathbf{k}_{\mathrm{MPE}}$ is $\hat{\mathbf{k}}_{\mathrm{MPE,KDE}} \approx 1.5$, while using the histogram and Gaussian mixture, $\hat{\mathbf{k}}_{\mathrm{MPE,histogram}} = \hat{\mathbf{k}}_{\mathrm{MPE,EM}} = 3.8$. This difference is explained by the accumulation of minimisers at this point: this challenges the assumption that $\mathbf{K}^*$ is continuous. As the density estimation techniques traditionally assume this continuity, the EM algorithm fits

this using a normal distribution with a very small variance, while the KDE considers a sum of Gaussian kernels of constant bandwidth, located at the same point. This particular problem highlights the issue with $\mathbf{k}_{\text{MPE}}$, as its estimation is highly sensitive to the density approximation procedure.

Alternatively, we can consider the mixing coefficients of the EM density estimation. The Gaussian centered at 1.5 has been attributed a weight of 0.608, while the one centered at 3.8 has a mixing weight of 0.226. Ultimately, an optimiser would be more likely to be close to 1.5, but we do not have information on its performance in this region around this value.

A way to avoid this issue is to introduce a bit of leeway for a value to be optimal, or at least "acceptably not optimal". This slackness takes the form of a relaxation coefficient, whose choice leads to a new family of robust estimators, along with a measure of their robustness via the coefficient.

2. Relative regret-based family of estimators

2.1. Taking the neighborhood into consideration

The density of minimisers has been estimated by optimising $\mathbf{k} \mapsto J(\mathbf{k}, \mathbf{u})$ over $\mathbb{K}$ for different realisations of $\mathbf{u}$. Instead of only considering optimal values, we propose to consider their *acceptable* neighbourhood in terms of performance of the cost function as well, as illustrated in Figure 3. To do so, for each $\mathbf{u}$, we define as acceptable values of the cost function, the values that are below $\alpha J^*(\mathbf{u})$, with $\alpha > 1$.

For each $\mathbf{u}$, we construct a region around the optimiser that produces acceptable values of the cost function, as seen on the bottom left plot. Looking at it the other way around, for a given $\mathbf{k}$, the set $R_\alpha(\mathbf{k}) \subset \mathbb{U}$ is defined as the set of $\mathbf{u}$ verifying the condition of acceptability:

$$R_\alpha(\mathbf{k}) = \{\mathbf{u} \in \mathbb{U} \mid J(\mathbf{k}, \mathbf{u}) \leq \alpha J^*(\mathbf{u})\} \tag{16}$$

Finally, for each $\mathbf{k}$, we will measure (with respect to $\mathbb{P}_{\mathbf{U}}$) the set of $\mathbf{u}$ within this acceptable neighbourhood, as illustrated in the bottom right plot.

Introducing the random nature of $\mathbf{U}$, $\Gamma_\alpha(\mathbf{k})$ is defined as the probability that $\mathbf{k}$ is acceptable given α:

$$\Gamma_\alpha(\mathbf{k}) = \mathbb{P}_{\mathbf{U}}\left[\mathbf{U} \in R_\alpha(\mathbf{k})\right] = \mathbb{P}_{\mathbf{U}}\left[J(\mathbf{k}, \mathbf{U}) \leq \alpha J^*(\mathbf{U})\right] \tag{17}$$

In other words, $\Gamma_\alpha(\mathbf{k})$ is then the probability that $J(\mathbf{k}, \mathbf{U})$ is between $J^*(\mathbf{U})$ and $\alpha J^*(\mathbf{U})$. Without relaxation, i.e. when α is set to 1, Γ_1 is non-zero if the set $\{\mathbf{u} \in \mathbb{U} \mid J(\mathbf{k}, \mathbf{u}) = J^*(\mathbf{u})\}$ has non-zero measure with respect to $\mathbb{P}_{\mathbf{U}}$, that is when the distribution of $\mathbf{K}^*$ presents atoms. This can be linked to the definition of the MPE from Eq. (11). In the more general case, the MPE can be seen as the limiting case when α tends to 1.

The motivation behind this relaxation is to take into account the local behaviour of the function around the conditional minimisers. For a given set of environmental conditions $\mathbf{u}$, if the function $\mathbf{k} \mapsto J(\mathbf{k}, \mathbf{u})$ is flat around its minimum $\mathbf{k}^*(\mathbf{u})$, then choosing $\mathbf{k}^*(\mathbf{u}) + \epsilon$ (for a small ϵ) will produce a value closer to the minimum than when the function has a high curvature. In addition to that, relaxing the constraint using a multiplicative constant puts more weight (i.e. a slower increase of the acceptable region) on the values of $\mathbf{k}^*(\mathbf{u})$ when $J^*(\mathbf{u})$ is close to zero.

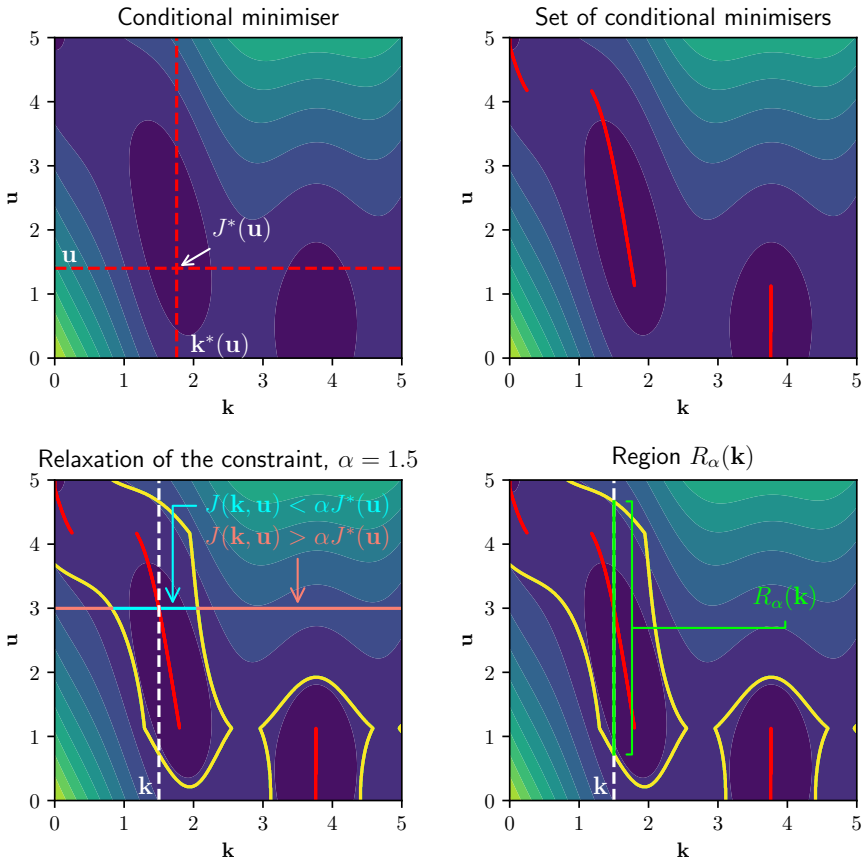

Figure 3: Principle of the relaxation of the constraint on J_{BHswap}, and illustration of $R_\alpha(\mathbf{k})$

Given that $J > 0$, $\Gamma_\alpha(\mathbf{k})$ is increasing with respect to α for any $\mathbf{k} \in \mathbb{K}$. We can then focus on the smallest value of α such that Γ_α reaches a certain level of confidence $p \in [0, 1]$. This leads to the definition of α_p

$$\begin{aligned} \alpha_p &= \inf \{\alpha \geq 1 \mid \exists \mathbf{k}_p \in \mathbb{K},\, \Gamma_\alpha(\mathbf{k}_p) \geq p\} \\ &= \inf \left\{ \alpha \geq 1 \mid \max_{\mathbf{k} \in \mathbb{K}} \Gamma_\alpha(\mathbf{k}) \geq p \right\} \end{aligned} \tag{18}$$

that is the smallest α, such that there exists a particular $\mathbf{k}_p \in \mathbb{K}$ for which $J(\mathbf{k}_p, \mathbf{U}) \leq \alpha_p J^*(\mathbf{U})$ with probability p. This formulation can be linked to the quantiles and Value-at-Risk of the random variable $\max_{\mathbf{k}} \{\frac{J(\mathbf{k},\mathbf{U})}{J^*(\mathbf{U})}\}$, that is a measure of risk usually applied in the financial sector (see [26]). For different levels p, and thus different α_p, Figure 4 shows examples of Γ_{α_p}. By definition, the associated $\mathbf{k}_p$ is then the first value for which Γ_{α_p} reaches p. We can see that changing the level p shifts the maximiser of Γ_{α_p}, and that for small α, $\Gamma_\alpha(\mathbf{k}_1)$ is also very small. This indicates that $\mathbf{k}_1$ is located quite far

from the conditional minimisers, and arise as a compromise when the relaxation is large enough.

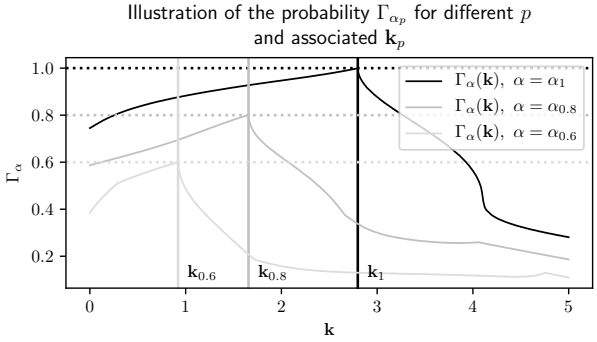

Figure 4: Illustration of the influence of different levels p on Γ_{α_p} and on $\mathbf{k}_p$ *(Pertinence ?)*

By considering the particular case where $p = 1$, a relation between $\mathbf{k}_1$ and α_1 can be derived, by using Eq. (18) and the strict positivity of the cost function:

$$\mathbb{P}_{\mathbf{U}}\left[\frac{J(\mathbf{k}_1, \mathbf{U})}{J^*(\mathbf{U})} \leq \alpha_1\right] = 1 \tag{19}$$

it follows then that

$$\alpha_1 = \sup_{\mathbf{u}\in\mathbb{U}} \frac{J(\mathbf{k}_1, \mathbf{u})}{J^*(\mathbf{u})} = \inf_{\mathbf{k}\in\mathbb{K}} \left\{\sup_{\mathbf{u}\in\mathbb{U}} \frac{J(\mathbf{k}, \mathbf{u})}{J^*(\mathbf{u})}\right\} \tag{20}$$

This alternative definition of α_1 is useful to estimate quickly $\hat{\alpha}_1$ whenever a grid over $\mathbb{K} \times \mathbb{U}$ has already been evaluated, thus avoiding multiple computations of $\Gamma_\alpha(\mathbf{k})$ until its maximum reaches 1. By Eq.(20), choosing a level of confidence equal to 1 is then equivalent to looking for the worst-case scenario of the ratio $\frac{J(\mathbf{k},\mathbf{u})}{J^*(\mathbf{u})}$, and therefore suffers from the same pitfall of the worst-case approach. As mentioned in the introduction, this returns over-conservative solution, and is not suited for random variable with unbounded support.

A high value of α_1 corresponds to a high ratio $\frac{J(\mathbf{k},\mathbf{u})}{J^*(\mathbf{u})}$ for at least one particular $\mathbf{u}$. In that sense, $J(\mathbf{k}_1, \mathbf{u})$ may be possibly very high. More generally, for a level of confidence p fixed, α_p is the slackness needed to be able to reach the probability p. So, in addition to the value of the estimate $\mathbf{k}_p$, we have an indicator of the robustness associated. We can then see that it is sensible to study jointly p, α_p and $\mathbf{k}_p$ in order to make a final decision.

2.2. Concurrence between p and α

For a practical purpose, we suppose that we have sampled N values of $\mathbf{U}$, namely $\{\mathbf{u}^i\}_{1\leq i\leq N}$. In order to estimate the quantities introduced above, there are two starting points: either fixing p or fixing the maximal threshold α. The most trivial way is to set

Definition	Related quantities	Interpretation
$\arg\min_{\mathbf{k}\in\mathbb{K}} \mathbb{E}_{\mathbf{U}}\left[J(\mathbf{k},\mathbf{U})\right]$	$\mathbf{k}_{\text{mean}}$	Long run performances
$\arg\min_{\mathbf{k}\in\mathbb{K}} \mathbb{V}\text{ar}_{\mathbf{U}}\left[J(\mathbf{k},\mathbf{U})\right]$	$\mathbf{k}_{\text{var}}$	Steady performances
$\arg\max_{\mathbf{k}\in\mathbb{K}} p_{\mathbf{K}^*}(\mathbf{k})$	$\mathbf{k}_{\text{MPE}}$	Most probable minimiser
$\inf\{\alpha \mid \exists \mathbf{k}_p \in \mathbb{K},\, \Gamma_\alpha(\mathbf{k}_p) \geq p\}$	$(p, \mathbf{k}_p, \alpha_p)$	Acceptable values with fixed probability p
$p_{\text{maxratio}} = \arg\max p/\alpha_p$	$(p_{\text{maxratio}}, \mathbf{k}_{\text{maxratio}}, \alpha_{\text{maxratio}})$	Maximal ratio of p and α_p

Table 1: Robust estimators, based on a cost function J

α, to find the couples of points $(\mathbf{k},\mathbf{u})$ verifying $J(\mathbf{k},\mathbf{u}) \leq \alpha J^*(\mathbf{u})$, and then to estimate the probability $\Gamma_\alpha(\mathbf{k})$ defined in Eq. (16) as

$$\hat{\Gamma}_\alpha(\mathbf{k}) = \frac{\#\left\{\mathbf{u}^i \mid J(\mathbf{k},\mathbf{u}^i) \leq \alpha J^*(\mathbf{u}^i)\right\}}{N} \tag{21}$$

If α is chosen too small, the resulting $\hat{p} = \max_{\mathbf{k}} \hat{\Gamma}_\alpha(\mathbf{k})$ will be also too small, meaning that the cost function will overshoot the value $\alpha J^*(\mathbf{U})$ with high probability.

Similarly, if p is fixed, the corresponding $\hat{\alpha}_p$ is computed by searching for the smallest α such that $\max_{\mathbf{k}} \hat{\Gamma}_\alpha(\mathbf{k})$ is equal to the level p fixed beforehand. In the end, if the value $\hat{\alpha}_p$ is too large, the relaxation needed to get acceptable values with probability p is very high, so the resulting estimation $\hat{\mathbf{k}}_p$ may not be relevant for the future application. *(chapeau car ici je parle de l'estimation en pratique)*

Without posing any constraints on p nor on α, a compromise between p and α would be preferable, and a rough idea of this balance can be obtained by studying $p \mapsto \alpha_p$, and particularly its slope. If this curve presents a steep increase, the multiplicative constant α_p must be increased by a large amount in order to increase the probability p by a small amount. Interesting couples (p, α_p) would then be the ones located before an abrupt increase of the slope of $p \mapsto \alpha_p$.

Another possibility is to combine directly p and α_p in order to find a compromise. We opted to model this compromise by the ratio (p/α_p), as it increases with respect to p and decreases with respect to α_p. The level of confidence p_{maxratio} is then defined as the maximiser of $p \mapsto p/\alpha_p$.

The different estimators introduced in this paper are summarised in Table 2.2.

2.3. Estimation of $\mathbf{k}_p$ and α_p for the J_{BH} and J_{BHswap} functions

As stated before, we chose to model the uncertainties as a random variable uniformly distributed on $\mathbb{U}$. The bounded nature of $\mathbb{U}$ allow us to consider the estimate introduced before up to a level of confidence $p = 1$. From now on, $\hat{\mathbf{k}}_{\text{MPE}}$ is estimated using KDE with Gaussian kernels. The smallest relaxation $\hat{\alpha}_1$ and the corresponding $\hat{\mathbf{k}}_1$ has been computed for the J_{BH} and J_{BHswap} functions, using a regular grid of 1000×1000 points on $\mathbb{K} \times \mathbb{U}$. The contour plots of those functions can be seen in the top plots of Figure 5. The frontier corresponding to the couples of points $(\mathbf{k},\mathbf{u})$ verifying $\{J(\mathbf{k},\mathbf{u}) = \alpha J^*(\mathbf{u})\}$ has been drawn on top of these contour plots, for $\alpha = \hat{\alpha}_1$ and an arbitrary $\alpha = 1.5 < \hat{\alpha}_1$, to illustrate the effect of the acceptable region when the relaxation α increases. On the bottom plots, the curves $\mathbf{k} \mapsto \hat{\Gamma}_\alpha(\mathbf{k})$ for $\alpha = \hat{\alpha}_1$ and $\alpha = 1.5$ along with the histogram

of the minimisers are represented. One can notice that the relaxation does not present an issue with the accumulation of the minimisers of J_{BH} at 3.8, as opposed to the MPE and its dependence on the estimation procedure of the distribution.

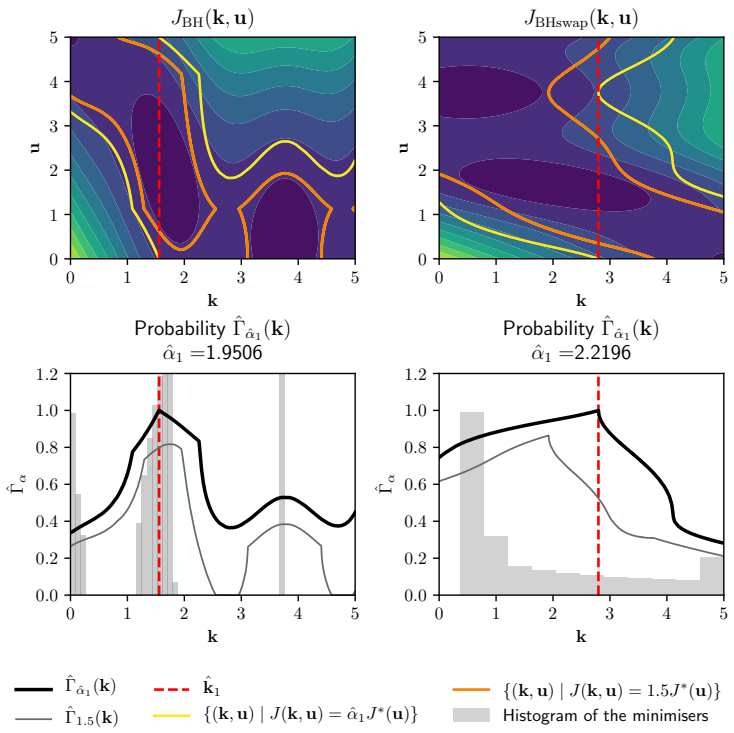

Figure 5: Top: J_{BH} and J_{BHswap} contours. The thick yellow lines are the boundaries of the acceptable region defined for $\hat{\alpha}_1$, the thick orange are for $\alpha = 1.5$. The red dashed line is the estimation $\hat{\mathbf{k}}_1$. Bottom: $\hat{\Gamma}_\alpha$ for $\alpha = 1.5$ and $\alpha = \hat{\alpha}_1$, and estimated density of the minimisers.

In order to choose a satisfying level of confidence p, we are going to study $p \mapsto \hat{\alpha}_p$ and $p \mapsto p/\hat{\alpha}_p$, as described in 2.2.

The plot of $p \mapsto \hat{\alpha}_p$ for the J_{BH} function on Figure 6 shows what seems to be a piecewise linear behaviour. The last change of slope, i.e. for $p \approx 0.9$ corresponds to a local maximum of the ratio, while the first change of slope corresponds to the global maximum of the ratio.

For the J_{BH} function, the numerical values can be found in Table 2. One can notice that the different estimates are close to each other for this problem.

Practically speaking, in order to compare the effective values taken by the objective function given an estimate $\hat{\mathbf{k}}$ we are going to consider the functions of the form $\mathbf{u} \mapsto J(\hat{\mathbf{k}}, \mathbf{u})$, that we will call "profile of $\hat{\mathbf{k}}$". Those profiles are well suited for the representation of the cost function for an estimate $\hat{\mathbf{k}}$ fixed as the uncertain variable is modelled with a 1D uniform random variable.

For the J_{BH} function, the curves are plotted in Figure 7. Based on its definition, the profile of $\hat{\mathbf{k}}_1$ is always within the shaded region, corresponding to $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$. The

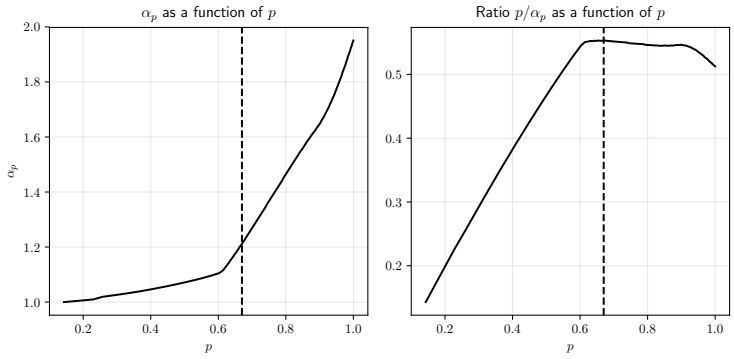

Figure 6: Evolution of the couples (p, α_p) and corresponding ratio p/α_p for the J_{BH} function. The dashed line indicates the level p associated with the highest ratio

Table 2: Estimation performed for J_{BH}, sorted by value

Estimate	Value
$\hat{\mathbf{k}}_{\text{var}}$	1.371
$\hat{\mathbf{k}}_p,\ p = 1$	1.557
$\hat{\mathbf{k}}_{\text{mean}}$	1.587
$\hat{\mathbf{k}}_{\text{MPE}}$	1.628
$\hat{\mathbf{k}}_{\text{maxratio}},\ \hat{p}_{\text{maxratio}} = 0.654$	1.637
$\hat{\mathbf{k}}_p,\ p = 0.95$	1.672

profile of $\hat{\mathbf{k}}_{\text{mean}}$ in contrast, exceeds $\hat{\alpha}_1 J^*(\mathbf{u})$ for $\mathbf{u}$ close to 5, while the profile of $\hat{\mathbf{k}}_{\text{var}}$ does for $\mathbf{u}$ close to 0. As the different $\hat{\mathbf{k}}$ are very close for every criteria introduced in this paper, the profiles are almost undistinguishable.

We are now going to shift our focus on the J_{BHswap} function, as Figure 8 provides the plots of $p \mapsto \hat{\alpha}_p$ and $p \mapsto p/\hat{\alpha}_p$.

Compared to the plots for J_{BH} in Figure 6, $\hat{\alpha}_p$ exhibits a smoother behaviour for the J_{BHswap} function as no change of slope is easily discernable. The computation of the ratio advocates for a level of confidence p_{maxratio} close to 0.75. The numerical values of the estimations $\hat{k}$ presented in Table 3 show that contrary to the J_{BH} function, the objective are concurrent.

In a similar fashion, the profiles of the different estimates for the J_{BHswap} function are shown in Figure 9.

In this case, $\hat{\mathbf{k}}_{\text{MPE}}$, $\hat{\mathbf{k}}_{\text{mean}}$ and $\hat{\mathbf{k}}_{\text{maxratio}}$ present a similar behaviour. They perform very well for $\mathbf{u} > 2$, especially for $\hat{\mathbf{k}}_{\text{MPE}}$ which is very close to the minimal value; However for $\mathbf{u} < 2$, they produce high values of the function.

The performances of $\hat{\mathbf{k}}_1$ are closer to the performances of $\hat{\mathbf{k}}_{\text{var}}$ for this function, but it performs worse than $\hat{\mathbf{k}}_{\text{mean}}$ and $\hat{\mathbf{k}}_{\text{var}}$ for $\mathbf{u} > 2$, even though its range is designed to stay within the interval $[J^*(\mathbf{u}); \hat{\alpha}_1 J^*(\mathbf{u})]$.

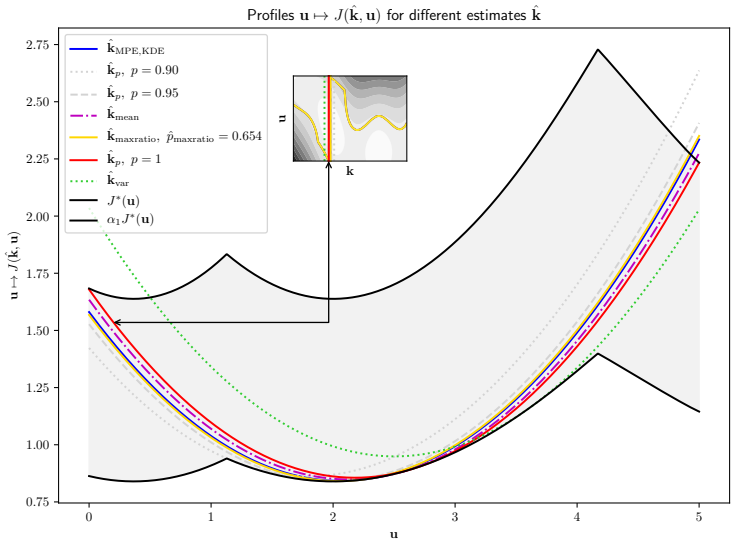

Figure 7: Profiles of the different estimates for the J_{BH} function, corresponding to the vertical cross sections of the contour. The shaded region corresponds to the interval $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$. The red dashed line, corresponding to the values taken by the cost function when the control variable is set to $\mathbf{k}_1$, is always between $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$ per design.

Table 3: Estimations performed for J_{BHswap}, sorted by value

Estimate	Value
$\hat{\mathbf{k}}_{\text{MPE}}$	0.606
$\hat{\mathbf{k}}_{\text{maxratio}},\ \hat{p}_{\text{maxratio}} = 0.766$	1.537
$\hat{\mathbf{k}}_{\text{mean}}$	1.752
$\hat{\mathbf{k}}_p,\ p = 0.90$	2.112
$\hat{\mathbf{k}}_p,\ p = 0.95$	2.457
$\hat{\mathbf{k}}_{\text{var}}$	2.638
$\hat{\mathbf{k}}_1,\ p = 1$	2.798

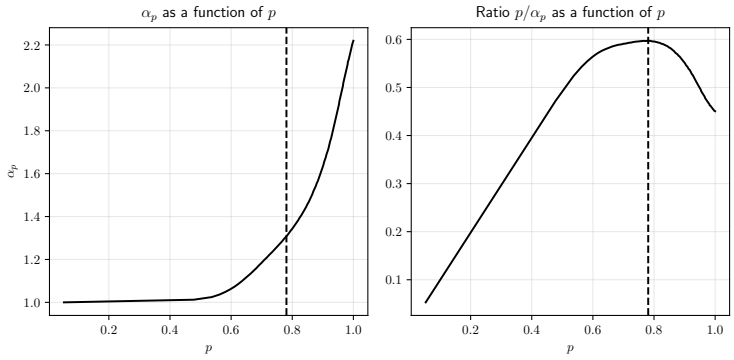

Figure 8: Evolution of the couples (p, α_p) and corresponding ratio p/α_p for the J_{BHswap} function. The dashed line indicates the level p associated with the highest ratio

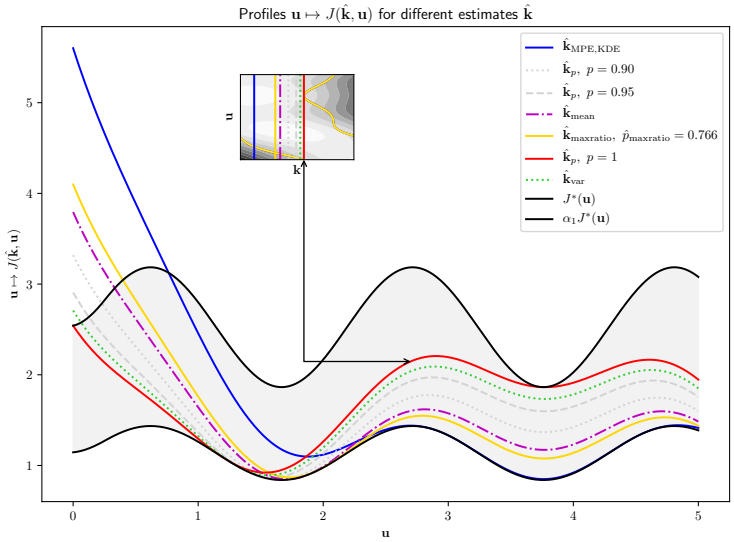

Figure 9: Profiles of the different estimates for the J_{BHswap} function. Those profiles are the vertical cross sections of the contours above. The shaded region corresponds to the interval $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$. The red dashed line, corresponding to the values taken by the cost function when the control variable is set to $\mathbf{k}_1$, is always between $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$ per design.

3. Robust calibration of a numerical model

3.1. Calibration of a toy numerical model

We will follow Kennedy and O'Hagan's approach in order to establish the function $\mathcal{G}$ described in the first section in Eq. (1), and the resulting cost function J.

Calibration of a numerical model is usually based on the comparison between the numerical model and some observations in a fixed time interval $[0, T]$ called assimilation window. The physical system at stake is based on an input $\mathbf{u} \in \mathcal{U}$, that represents some operating conditions. This system can then be seen as a map from $\mathcal{U}$ to $\mathcal{Y}$, denoted as $\mathcal{M}^o : \mathbf{u} \mapsto \mathcal{M}^o(\mathbf{u})$.

In addition to the operating conditions $\mathbf{u}$, the numerical model is also dependent on some other input $\mathbf{k} \in \mathcal{K}$. This additional parametrization comes usually from the successive simplifications needed to implement a numerical model of the physical system observed, and from some unobserved physical parameters. These parameters need to be calibrated accordingly, so that the numerical model can be used to predict the behaviour of the physical system under different operating conditions.

The true value taken by the physical system during this window is denoted by $\mathcal{M}^o(\mathbf{u}^{\text{true}}) \in \mathcal{Y}$, where $\mathbf{u}^{\text{true}} \in \mathcal{U}$ is unknown. The misfit $\mathcal{G}$ and the cost function J defined Eq.(1) can then be defined as

$$J(\mathbf{k}, \mathbf{u}) = \frac{1}{2}\|\mathcal{G}(\mathbf{k}, \mathbf{u})\|^2 = \frac{1}{2}\|\mathcal{M}(\mathbf{k}, \mathbf{u}) - \mathcal{M}^o(\mathbf{u}^{\text{true}})\|^2 \tag{22}$$

3.2. The Shallow Water equations

The model we wish to calibrate is based on the Shallow Water equations, described Eq. (23). h is the height of the water column, q is the discharge, and z is the bathymetry, while g is the usual gravitation constant. The parameter to calibrate, $\mathbf{k}$, is the quadratic friction term, proportional to the square of the inverse of Manning-Strickler coefficient. The environmental parameter $\mathbf{u}$ is the amplitude of a sine wave of period $1/\omega_0$. The domain of those two parameters are $\mathcal{K} = [0.0, 1.3]$ and $\mathcal{U} = [0.5, 0.7]$.

$$\begin{cases} \partial_t h + \partial_x q &= 0 \\ \partial_t q + \partial_x \left(\frac{q^2}{h} + \frac{g}{2}h^2\right) &= -gh\partial_x z - \mathbf{k}q|q|h^{-7/3} \\ h(0,t) &= 20.0 + 3 \cdot \sin\left(\frac{2\pi t}{2}\right) + 1.5 \cdot \mathbf{u} \cdot \sin\left(\frac{2\pi t}{\omega_0}\right) \\ \partial_x q(0,t) &= 0 \end{cases} \tag{23}$$

These equations are integrated using a finite-volume scheme on a discretized domain $[0, L]$, up to a time T. The output of the computer code is the sea surface height h, on the center of all the volumes and at all the time-steps, that will be denoted $\mathcal{M}(\mathbf{k}, \mathbf{u}; \omega_0 = 1.0)$. In this setting, the variable input is uniformly distributed on $\mathcal{U}$. To generate the observations, we choose $\mathbf{u}^{\text{true}} = 2/3$, and define $\mathcal{M}^o$ based on the computer model $\mathcal{M}$, such that $\mathcal{M}^o(\mathbf{u}^{\text{true}}) = \mathcal{M}(\mathbf{k}^{\text{true}}, \mathbf{u}^{\text{true}}; \omega_0 = 0.999)$. ω_0 represents here the uncontrollable error between the observations and the numerical model and will now be omitted systematically in the notation. The true value of the bottom friction $\mathbf{k}^{\text{true}} = (k_1^{\text{true}}, k_2^{\text{true}}, \ldots, k_{N_{\text{vol}}}^{\text{true}})$ is not constant over the whole domain, and is defined as

$$k_i^{\text{true}} = 0.2 \cdot \left(1 + \sin\left(\frac{2\pi x_i}{L}\right)\right)$$

where x_i is the center of the i-th volume. The two sources of representativity errors are ω_0, and the fact that we are looking to estimate $\mathbf{k}$ in a 1-dimensional space $\mathbb{K}$.

3.3. Estimation

Given that the numerical model is expensive to evaluate, it has first been evaluated on a relatively small regular grid. A metamodel based on Gaussian process is constructed and the PEI criterion [28, 29] allow us to add point to evaluate and to better capture the locus of the conditional minimisers $\{(\mathbf{k}^*(\mathbf{u}), \mathbf{u}) \mid \mathbf{u} \in \mathbb{U}\}$. A bigger regular grid is evaluated by the metamodel once the design space has been sufficiently explored. The different steps of the estimation are illustrated Figure 10. We can see that the estimation for this problem seems less problematic than the analytical examples shown above, as all derived quantites are unimodal. The different estimates however take a wide range of values, as seen on Table 4. As a basis for comparison, the global minimiser of J over $\mathbb{K} \times \mathbb{U}$ has been computed, and $\hat{\mathbf{k}}_{\text{global}}$ is then obtained by discarding the $\mathbf{u}$ value.

As discussed earlier, the estimation of the MPE is dependent on the method used, and then its estimation using KDE in the middle plot of Figure 10 is here only for illustrative purposes.

Table 4: Calibrated values of $\mathbf{k}$ according to different criteria

Estimator	Value
$\hat{\mathbf{k}}_{\text{MPE}}$	0.249
$\hat{\mathbf{k}}_{\text{global}}$	0.290
$\hat{\mathbf{k}}_p,\ p = 1$	0.423
$\hat{\mathbf{k}}_p,\ p = 0.90$	0.458
$\hat{\mathbf{k}}_{\text{mean}}$	0.501
$\hat{\mathbf{k}}_p,\ p = 0.80$	0.505
$\hat{\mathbf{k}}_p,\ p = 0.70$	0.560
$\hat{\mathbf{k}}_{\text{var}}$	0.990

Similarly as for J_{BH} and J_{BHswap}, the profiles are depicted Figure 11. We can see that the performances of $\hat{\mathbf{k}}_1$ and $\hat{\mathbf{k}}_{\text{mean}}$ are very similar, but $\hat{\mathbf{k}}_1$ has better performances when $\mathbf{u} > 0.6$.

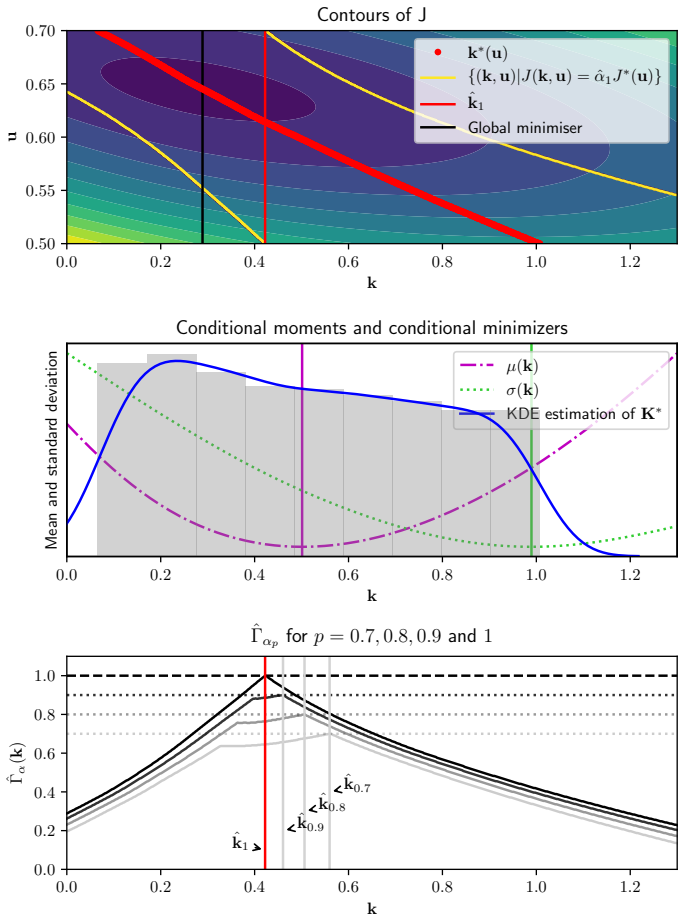

Figure 10: Procedure of robust calibration for the shallow water problem. Top: contours of J, conditional minimisers and $\{(\mathbf{k},\mathbf{u})|J(\mathbf{k},\mathbf{u}) = \hat{\alpha}_1 J^*(\mathbf{u})\}$. Middle: Conditional moments and histogram and KDE of the conditional minimisers. Bottom: $\hat{\Gamma}_{\alpha_p}$ for different levels p

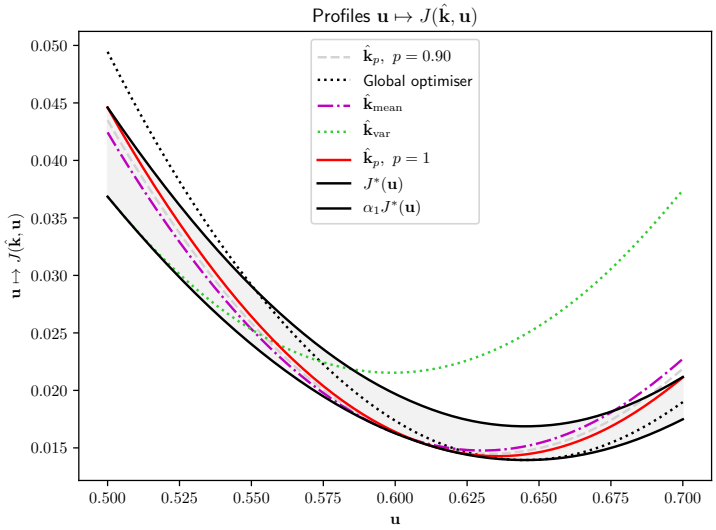

Figure 11: Profiles for the calibration problem

3.4. Assessing the quality of the forecast of the calibrated model

The calibration of the model has been performed using by integrating the model on a time-period $[0, T]$, called assimilation window. We now want to compare the quality of the different forecasts, from different calibrated bottom friction. Those forecasts result from the integration of the numerical model between the time T and a time T_{pred}.

Given the probabilistic nature of the environmental conditions $\mathbf{U}$, the forecasts will also be probabilistic. We will then compare $\mathcal{M}^o_{\text{pred}}(\mathbf{U})$ and $\mathcal{M}_{\text{pred}}(\mathbf{k}, \mathbf{U})$, for a calibrated $\mathbf{k}$.

3.4.1. Squared forecast error

Given two environmental conditions $\mathbf{u}$, used to run the computer simulation and $\mathbf{u}'$, used to generate the observations, the squared forecast error for the parameter $\mathbf{k}$ is

$$S_{\text{pred}}(\mathbf{k}, \mathbf{u}, \mathbf{u}') = \left(\mathcal{M}_{\text{pred}}(\mathbf{k}, \mathbf{u}) - \mathcal{M}^o_{\text{pred}}(\mathbf{u}')\right)^2 \tag{24}$$

By averaging over $\mathbf{u}$ and $\mathbf{u}'$ using a Monte-Carlo approximation, we can define the mean squared forecast error, defined on every point of the spatial domain, and at every time-steps.

$$S(\mathbf{k}) = \frac{1}{N_{\mathbf{u}} N_{\mathbf{u}'}} \sum_{\mathbf{u},\mathbf{u}' \sim \mathbf{U}} S_{\text{pred}}(\mathbf{k}, \mathbf{u}, \mathbf{u}') = \frac{1}{N_{\mathbf{u}} N_{\mathbf{u}'}} \sum_{\mathbf{u},\mathbf{u}' \sim \mathbf{U}} \left(\mathcal{M}_{\text{pred}}(\mathbf{k}, \mathbf{u}) - \mathcal{M}^o_{\text{pred}}(\mathbf{u}')\right)^2 \tag{25}$$

The mean squared forecast error is then averaged over all time-steps between T and T_{pred}, indicating on which point of the domain the forecast error is greater as seen on the left plot of Figure 12. We can see that the mean forecast error squared, on the right

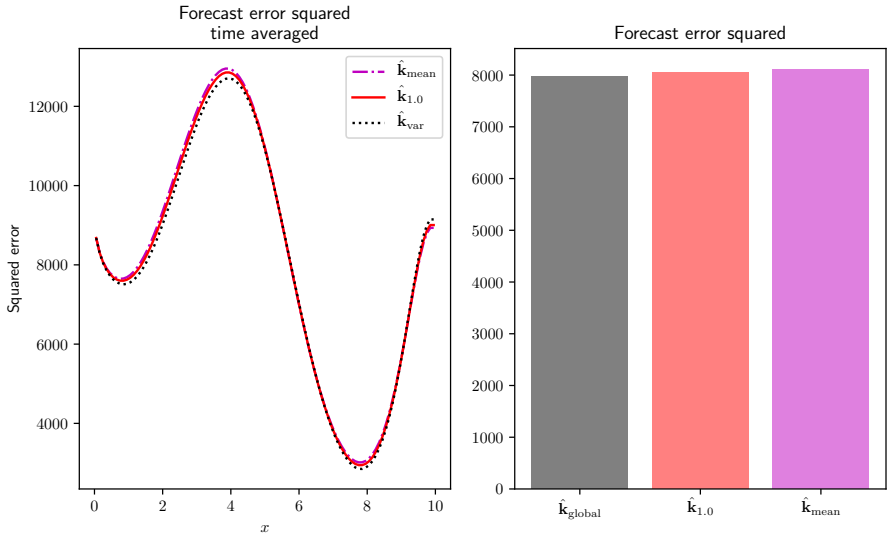

Figure 12: Quality of forecast, depending on the calibrated parameter. The left figures show the metric averaged over time, while the right figure show the mean forecast error, averaged over time and space

side, averaged over time and space is the smallest for $\hat{\mathbf{k}}_{\text{global}}$, while $\hat{\mathbf{k}}_{1.0}$ performs slightly better $\hat{\mathbf{k}}_{\text{mean}}$.

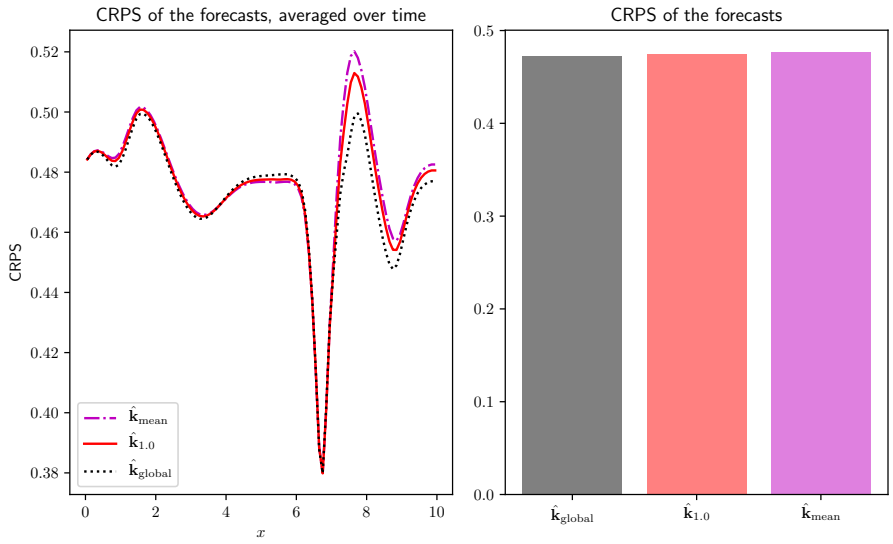

Figure 13: CRPS computed for different calibrated parameters

3.4.2. Continuous Ranked Probability Score

Given the random variables $\mathcal{M}_{\text{pred}}(\hat{\mathbf{k}}, \mathbf{U})$ and $\mathcal{M}^o_{\text{pred}}(\mathbf{U})$, representing the probabilistic forecast and the probabilistic observations, we can define the cumulative distribution functions (CDF) $F_{\text{pred}}(\cdot, \hat{\mathbf{k}})$ and $F^o_{\text{pred}}(\cdot)$ The Continuous ranked probability score (CRPS) measures the squared difference between the predicted CDF F_{pred} using a calibrated value, and the CDF of the observations F^o_{pred}.

$$\text{CRPS}(\mathbf{k}) = \int_{\mathbb{R}} (F_{\text{pred}}(x, \mathbf{k}) - F^o_{\text{pred}}(x))^2 \, \mathrm{d}x \tag{26}$$

Conclusion

This paper deals with the problem of robust calibration of a computer code in the presence of uncertain inputs.

Our contribution consists in introducing a new criterion of robustness, based on the distribution of the minimisers. More specifically, this estimate is bounding the ratio between the cost function and its conditional minimum. Experimental results show that this new estimate is able to make a compromise between the minimum of the expected value and the minimum of the variance, will also giving the range of the ratio evoked above.

From a practical perspective, this estimation is very expensive: each evaluation of $\mathbf{k}^*$ requires an optimisation procedure. The distribution of the conditional minimisers is then very complicated to compute. In addition to that, in order to relax the constraint, the cost function must also be evaluated in the vicinity of the conditional minimisers. Computing precisely this estimate is then intractable if each run of the underlying model is expensive (longer than a couple of seconds).

A solution to explore is the use of surrogate models. Based on an initial design of experiment, the (expensive) cost function is replaced by a function, that is cheap to evaluate. For instance, [28] uses Gaussian Process regression and sequential methods to explore the conditional minimisers.

Finally, the adaptibility of this method has to be studied in a higher dimensional setting. High-dimensional density estimation is a challenge in itself, but is necessary in order to tackle more realistic applications.

[1] W. E. Walker, P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. van Asselt, P. Janssen, M. P. Krayer von Krauss, Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support, Integrated assessment 4 (1) (2003) 5–17.

[2] S. K. Das, R. W. Lardner, On the Estimation of Parameters of Hydraulic Models by Assimilation of Periodic Tidal Data, Journal of Geophysical Research 96 (C8) (1991) 15187, ISSN 0148-0227, doi:10.1029/91JC01318.

[3] M. Boutet, Estimation Du Frottement Sur Le Fond Pour La Modélisation de La Marée Barotrope, Ph.D. thesis, Université d'Aix Marseille, 2015.

[4] L. Huyse, D. M. Bushnell, Free-Form Airfoil Shape Optimization under Uncertainty Using Maximum Expected Value and Second-Order Second-Moment Strategies .

[5] N. Lelièvre, P. Beaurepaire, C. Mattrand, N. Gayton, A. Otsmane, On the Consideration of Uncertainty in Design: Optimization-Reliability-Robustness, Structural and Multidisciplinary Optimization 54 (6) (2016) 1423–1437.

[6] G. Petrone, G. Iaccarino, D. Quagliarella, Robustness Criteria in Optimization under Uncertainty, Evolutionary and deterministic methods for design, optimization and control (EUROGEN 2011). CIRA, Capua (2011) 244–252.

[7] P. Seshadri, P. Constantine, G. Iaccarino, G. Parks, A Density-Matching Approach for Optimization under Uncertainty, arXiv:1409.7089 [math, stat] .

[8] L. W. Cook, J. P. Jarrett, Horsetail Matching: A Flexible Approach to Optimization under Uncertainty, Engineering Optimization 50 (4) (2018) 549–567, ISSN 0305-215X, 1029-0273, doi: 10.1080/0305215X.2017.1327581.

[9] P. J. Huber, Robust Statistics, in: International Encyclopedia of Statistical Science, Springer, 1248–1251, 2011.

[10] V. Rao, A. Sandu, M. Ng, E. Nino-Ruiz, Robust Data Assimilation Using L_1 and Huber Norms, SciRate .

[11] J. O. Berger, E. Moreno, L. R. Pericchi, M. J. Bayarri, J. M. Bernardo, J. A. Cano, J. De la Horra, J. Martín, D. Ríos-Insúa, B. Betrò, An Overview of Robust Bayesian Analysis, Test 3 (1) (1994) 5–124.

[12] J. Marzat, E. Walter, H. Piet-Lahanier, Worst-Case Global Optimization of Black-Box Functions

through Kriging and Relaxation, Journal of Global Optimization 55 (4) (2013) 707–727, ISSN 0925-5001, 1573-2916, doi:10.1007/s10898-012-9899-y.
[13] J. Villemonteix, E. Vazquez, E. Walter, An Informational Approach to the Global Optimization of Expensive-to-Evaluate Functions, arXiv:cs/0611143 .
[14] P. Hennig, C. J. Schuler, Entropy Search for Information-Efficient Global Optimization .
[15] J. M. Buhmann, M. Mihalak, R. Sramek, P. Widmayer, Robust Optimization in the Presence of Uncertainty, in: Proceedings of the 4th Conference on Innovations in Theoretical Computer Science - ITCS '13, ACM Press, Berkeley, California, USA, ISBN 978-1-4503-1859-4, 505, doi:10.1145/2422436.2422491, 2013.
[16] A. Juditsky, A. S. Nemirovski, G. Lan, A. Shapiro, Stochastic Approximation Approach to Stochastic Programming, in: ISMP 2009 - 20th International Symposium of Mathematical Programming, 2009.
[17] S. Kim, R. Pasupathy, S. Henderson, A Guide to Sample Average Approximation, vol. 216, 207–243, doi:10.1007/978-1-4939-1384-8_8, 2015.
[18] J. Janusevskis, R. Le Riche, Simultaneous Kriging-Based Sampling for Optimization and Uncertainty Propagation, Tech. Rep., 2010.
[19] V. Baudoui, Optimisation Robuste Multiobjectifs Par Modèles de Substitution, Ph.D. thesis, Toulouse, ISAE, 2012.
[20] O. Grodzevich, O. Romanko, Normalization and Other Topics in Multi-Objective Optimization .
[21] R. T. Marler, J. S. Arora, The Weighted Sum Method for Multi-Objective Optimization: New Insights, Structural and Multidisciplinary Optimization 41 (6) (2010) 853–862, ISSN 1615-147X, 1615-1488, doi:10.1007/s00158-009-0460-7.
[22] J. S. Lehman, T. J. Santner, W. I. Notz, Designing Computer Experiments to Determine Robust Control Variables, Statistica Sinica (2004) 571–590.
[23] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the royal statistical society. Series B (methodological) (1977) 1–38.
[24] D. Freedman, P. Diaconis, On the Histogram as a Density Estimator:L2 Theory, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 57 (4) (1981) 453–476, ISSN 1432-2064, doi:10.1007/BF01025868.
[25] D. W. Scott, On Optimal and Data-Based Histograms, Biometrika 66 (3) (1979) 605, ISSN 00063444, doi:10.2307/2335182.
[26] R. T. Rockafellar, S. P. Uryasev, M. Zabarankin, Deviation Measures in Risk Analysis and Optimization, SSRN Scholarly Paper ID 365640, Social Science Research Network, Rochester, NY, 2002.
[27] M. C. Kennedy, A. O'Hagan, Bayesian Calibration of Computer Models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464, ISSN 1467-9868, doi:10.1111/1467-9868.00294.
[28] D. Ginsbourger, J. Baccou, C. Chevalier, F. Perales, N. Garland, Y. Monerie, Bayesian Adaptive Reconstruction of Profile Optima and Optimizers, SIAM/ASA Journal on Uncertainty Quantification 2 (1) (2014) 490–510, ISSN 2166-2525, doi:10.1137/130949555.
[29] J. Bossek, B. Bischl, T. Wagner, G. Rudolph, Learning Feature-Parameter Mappings for Parameter Tuning via the Profile Expected Improvement, in: Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation, GECCO '15, ACM, New York, NY, USA, ISBN 978-1-4503-3472-3, 1319–1326, doi:10.1145/2739480.2754673, 2015.