Robust calibration of numerical models based on relative regret

Victor Trappler, Élise Arnaud, Arthur Vidard, Laurent Debreu

Robust calibration of numerical models based on relative regret

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

[a] *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP[1], 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 family of calibration objectives based on the notion of relative-regret with respect to the best attainable performance given the uncertainties and compare it with the minimum in the mean sense, and the minimum of variance.

Keywords: Calibration, Robust optimisation, Relative-regret, Shallow-water equations

1. Introduction

Numerical models are widely used to study or forecast natural phenomena and improve industrial processes. However, by essence models only partially represent 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] proposes 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 phenomenon underlying the behaviour of the system under study. The latter can

*Corresponding author
Email address: victor.trappler@univ-grenoble-alpes.fr (Victor Trappler)
[1] Institute of Engineering Univ. Grenoble Alpes

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 very poor in other configurations, phenomenon coined as *localized optimisation* in [4]. Taking into account aleatoric uncertainties in optimisation problems takes several names, including *robust optimisation*, *robust design* in [5], or *optimisation 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 for every realisation $\mathbf{u} \in \mathbb{U}$ of $\mathbf{U}$, finding $\mathbf{k_u} \in \mathbb{K}$ that minimises the cost function $\mathbf{k} \mapsto J(\mathbf{k}, \mathbf{U} = \mathbf{u})$ is a well-posed problem, and that the optimum is unique. Additionally, the following assumptions 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 2. Section 3 introduces a new class of estimators, by relaxing the constraint of optimality and defining regions of acceptability, similarly as [15] in discrete combinatorial problems, or [16, 17] in operation research. 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 described methods are given on a numerical exemple in Section 4.

2. 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 robust estimators, either by using the moments of the cost function; or by exploiting the distribution of its minimisers.

2.1. Optimisation of the moments

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 $\mathbf{k}_{\mathbb{E}}$ defined by:

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

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

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

A lot of different methods are readily available to solve these minimisation problems. For instance, stochastic Sample Approximation [18, 19] 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:

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

and

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

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 [20]. Even though $\mathbf{k}_{\mathbb{E}}$ is a reasonable choice, there is no guarantee that $J(\mathbf{k}_{\mathbb{E}}, \mathbf{U} = \mathbf{u})$ will not reach catastrophic level for some $\mathbf{u}$. On the other hand, using $\mathbf{k}_{\mathbb{V}}$ will ensure stability of the cost function, but without any control of its performance. Ideally, one would want to have a small expectation and a

small variance at the same time. 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 [21].

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 [22] and in [23], while some others are based on the minimisation of one of the objectives under constraints on the others, as performed in [24]. 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.

2.2. Most probable estimate

Let us consider the minimal cost attainable in each configuration brought by $\mathbf{u}$. The resulting conditional minimum is denoted 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 the 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. In the rest of the paper, the hat notation will indicate an estimation using numerical values of the underlying theoretical quantity. 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) [25], and EM (Expectation-Maximisation) [26].

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 $f_{1\mathrm{D}}$, and a popular choice is the Gaussian kernel $f_{1\mathrm{D}}(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$.

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 $\{\pi_i\}_{i=1\dots\nu}$ are the mixing coefficients.

In practice, 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.

2.3. Numerical illustration

Before going further in the explanation of our approach, let us illustrate the nature of previously detailed estimators, $\hat{\mathbf{k}}_{\mathbb{E}}, \hat{\mathbf{k}}_{\mathbb{V}}$ 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 in the estimations.

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 the top of Figure 1.

The left, (resp. right), column stands for J_{BH} (resp. J_{BHswap}). Functions $\mu(\mathbf{k})$ and $\sigma(\mathbf{k})$ are drawn on the bottom row, respectively in purple and green. The corresponding minimisers $\hat{\mathbf{k}}_{\mathbb{E}}$ and $\hat{\mathbf{k}}_{\mathbb{V}}$ are also plotted. On this figure, we can observe that $\hat{\mathbf{k}}_{\mathbb{E}}$ and $\hat{\mathbf{k}}_{\mathbb{V}}$ are close for J_{BH}, while being significantly different for J_{BHswap}.

Similarly, estimations of $\hat{\mathbf{k}}_{\mathrm{MPE}}$ are depicted on Figure 2. The top row shows the contours 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 bottom row presents three approximations of the density of minimisers: the histogram in grey (bin size selected using Freedman-Diaconis from [27]), the result of a kernel density estimation (KDE) with Gaussian kernels in red (using Scott's rule from [28] for bandwidth selection), and the estimation by a Gaussian mixture, calculated with the EM algorithm. The number of Gaussians has been fixed to 3, a guess

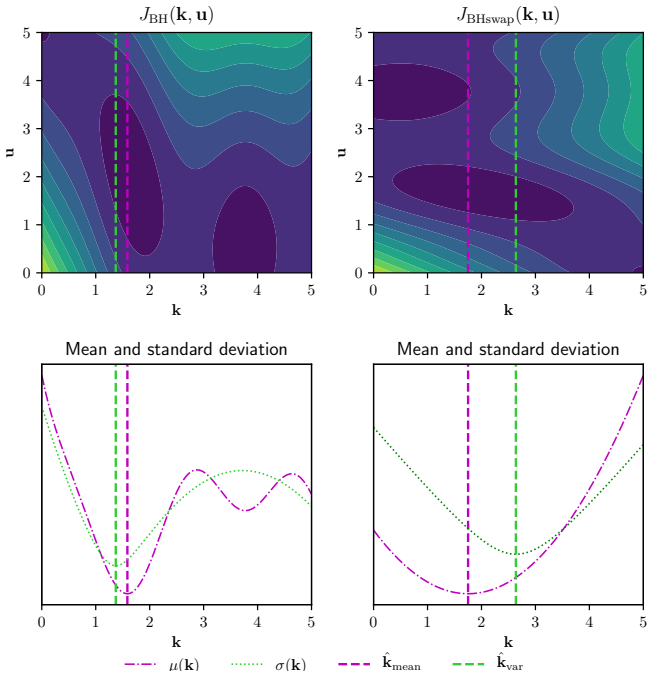

Figure 1: The left column concerns J_{BH}, while the right one deals with J_{BHswap}. Contours of both functions are plotted on the top, and curves of $\mu(\mathbf{k})$ and $\sigma(\mathbf{k})$ are shown on the bottom (respective scales are not displayed). Estimates $\hat{\mathbf{k}}_{\mathbb{E}}$ and $\hat{\mathbf{k}}_{\mathbb{V}}$ are plotted with the dashed line.

based on the general shape of the histogram. Respective estimations of $\hat{\mathbf{k}}_{\mathrm{MPE}}$ are also depicted using dashed lines.

For J_{BHswap}, we can observe that those three methods give consistent 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 J_{BH}: 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 an issue with $\mathbf{k}_{\mathrm{MPE}}$, as its estimation is possibly sensitive to the density approximation procedure.

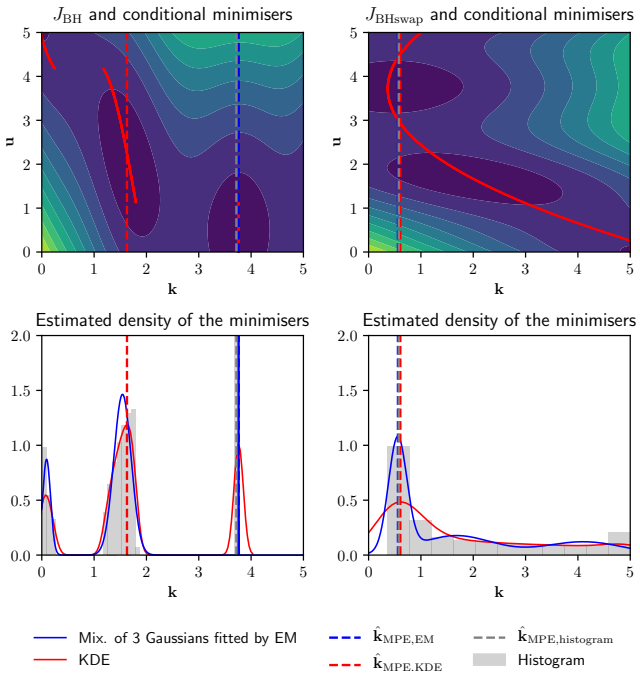

Figure 2: Top Left: J_{BH} along with conditional minimisers in red. Bottom left: Estimated densities using KDE and EM algorithm and the histogram. The dashed lines indicate the MPE found using those methods. Right: Same quantities with J_{BHswap}.

Instead of just considering the optimal minimisers, we introduce a bit of leeway, and look for "acceptably not optimal" parameters. This slackness takes the form of a relaxation coefficient and its choice defines a new family of robust estimators, where each one of its member carries information on its robustness through this coefficient.

3. Relative regret-based family of estimators

3.1. Relaxing the optimality constraint

The density of minimisers has been estimated by optimising $\mathbf{k} \mapsto J(\mathbf{k}, \mathbf{U} = \mathbf{u})$ over $\mathbb{K}$ for different realisations of $\mathbf{u}$. Instead of focusing on optimal values, we propose to consider their *acceptable* neighbourhood in terms of performance of the cost function as well. To do so, for each $\mathbf{u}$, $\mathbf{k}$ is deemed acceptable when $J(\mathbf{k}, \mathbf{U} = \mathbf{u}) \leq \alpha J^*(\mathbf{u})$, with $\alpha > 1$. In this context, for a given $\mathbf{k}$, the set $R_\alpha(\mathbf{k}) \subseteq \mathbb{U}$ is defined as the set of $\mathbf{u}$, for

which $\mathbf{k}$ is acceptable:

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

Figure 3 details the successive steps for the construction of the set R_α. First, the conditional minimisers are computed, as shown on the top plots. Afterwards, for a given level $\alpha = 1.5$, the set of acceptable $\mathbf{k}$ can be identified for each $\mathbf{u} \in \mathbb{U}$, as shown on the bottom left plot. Finally, the region $R_\alpha(\mathbf{k})$ is the subset of $\mathbb{U}$ for which $\mathbf{k}$ is acceptable, as represented with a vertical slice on the bottom right plot.

Introducing the random nature of $\mathbf{U}$, one can define $\Gamma_\alpha(\mathbf{k})$ 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 the probability that $J(\mathbf{k}, \mathbf{U})$ is between $J^*(\mathbf{U})$ and $\alpha J^*(\mathbf{U})$.

Noting that without relaxation, *i.e.* when α is set to 1, Γ_1 is non-zero if the set $\{\mathbf{u} \in \mathbb{U} \mid J(\mathbf{k}, \mathbf{U} = \mathbf{u}) = J^*(\mathbf{u})\}$ has non-zero measure with respect to $\mathbb{P}_\mathbf{U}$. It happens when the distribution of $\mathbf{K}^*$ presents atoms.

This can be linked to the definition of the distribution of the minimisers $\mathbf{K}^*$. For instance, if $\mathbb{K}$ is a discrete set, we can rewrite Γ_1 as $\Gamma_1(\mathbf{k}) = \mathbb{P}_\mathbf{U}\left[J(\mathbf{k}, \mathbf{U} = \mathbf{u}) = J^*(\mathbf{U})\right] = \mathbb{P}_\mathbf{U}\left[\mathbf{k} = \mathbf{k}^*(\mathbf{U})\right]$ which is the probability mass function of $\mathbf{K}^*$.

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 $\mathbf{k} \mapsto J(\mathbf{k}, \mathbf{U} = \mathbf{u})$ 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 on the values of $\mathbf{k}^*(\mathbf{u})$ when $J^*(\mathbf{u})$ is small.

The choice of the relaxation constant α can be made to ensure the existence of a parameter that is "acceptable" with a certain probability. For instance, 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

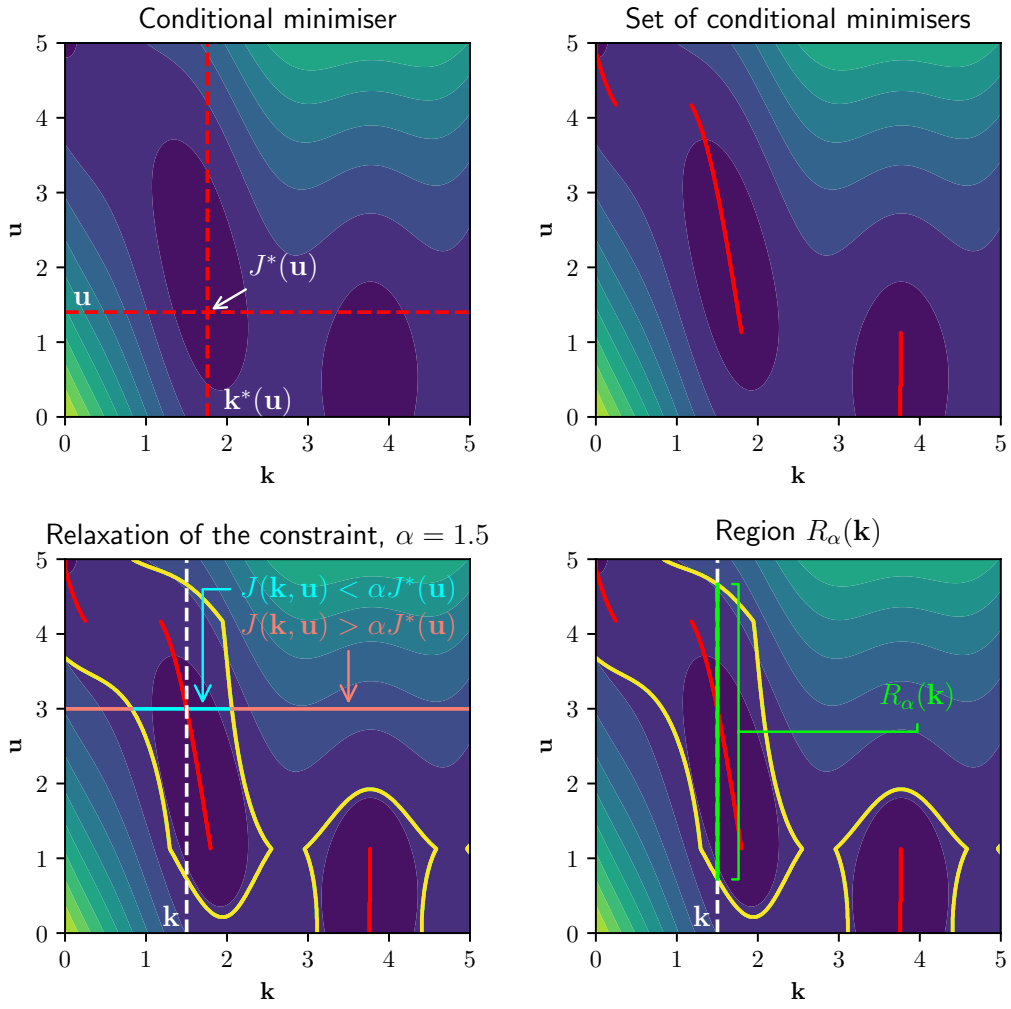

Figure 3: Principle of the relaxation of the constraint on J_{BH}, and illustration of $R_\alpha(\mathbf{k})$. Top plots: Computation of the conditional minimisers $\mathbf{k}^*(\mathbf{u})$. Top right plot: the set $(\mathbf{k}^*(\mathbf{u}), \mathbf{u})$ of conditional minimisers is represented in red. On the bottom left plot, for a relaxation $\alpha = 1.5$, and $\mathbf{u} = 3$, the acceptable $\mathbf{k}$ are in cyan, while the non-acceptable ones are in orange, while the frontier $\{(\mathbf{k}^*(\mathbf{u}), \mathbf{u}) \mid J^*(\mathbf{u}) = J(\mathbf{k}, \mathbf{U} = \mathbf{u})\}$ is in yellow. On the bottom right plot, the set $R_\alpha(\mathbf{k})$ for $\alpha = 1.5$ and $\mathbf{k} = 1.5$ is in green.

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. $\{\mathbf{k}_p \text{ for } p \in [0;1]\}$ defines the relative-regret family of robust estimators (RRE). This formulation can be linked to the quantiles and Value-at-Risk of the random variable $\max_{\mathbf{k}}\{J(\mathbf{k}, \mathbf{U})/J^*(\mathbf{U})\}$, that is a measure of risk usually applied in the financial sector (see [29]). For different levels p, and thus different α_p, Figure 4 shows examples of Γ_{α_p} for J_{BHswap}. 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.

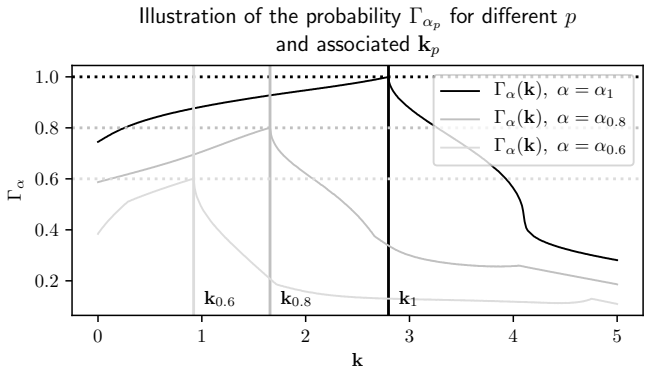

Figure 4: Illustration of the influence of different levels p on Γ_{α_p} and on $\mathbf{k}_p$ for J_{BHswap}

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

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

This alternative definition of α_1 is useful to estimate it quickly 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. From Eq. (20), choosing a level of confidence equal to 1 is

then equivalent to looking for the worst-case scenario of the ratio $J(\mathbf{k}, \mathbf{U} = \mathbf{u})/J^*(\mathbf{u})$. Therefore it may suffer from the same pitfall as the worst-case approach. As mentioned in the introduction, this is not suited for random variable with unbounded support as it may return over-conservative solution, provided that a solution exists in the first place.

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

3.2. Choosing the relaxation coefficient

In order to choose the quantities introduced above, there are three choices: either fixing p, either fixing the maximal threshold α, or looking for a compromise between p and α. The most trivial way is to set α, to find the couples of points $(\mathbf{k}, \mathbf{u})$ verifying $J(\mathbf{k}, \mathbf{U} = \mathbf{u}) \leq \alpha J^*(\mathbf{u})$, and then to estimate the probability $\Gamma_\alpha(\mathbf{k})$ defined in Eq. (16). For instance, when N sampled values of $\mathbf{U}$: $\{\mathbf{u}^i\}_{1 \leq i \leq N}$ are available, a possible estimator of Γ_α is

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

However, this may be a risky approach. Indeed if α is chosen too small, the resulting $\hat{p} = \max_{\mathbf{k}} \hat{\Gamma}_\alpha(\mathbf{k})$ will be also small, meaning that the cost function will have non acceptable values with high probability.

Similarly, if p is fixed, the corresponding $\hat{\alpha}_p$ is computed by searching for the smallest α satisfying $\max_{\mathbf{k}} \hat{\Gamma}_\alpha(\mathbf{k}) = p$. Once again, if $\hat{\alpha}_p$ 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.

Also, choosing $p = 1$ is possible only if the sample space of $\mathbf{U}$ is bounded. An

illustration of this flaw appears in the following simple problem:

$$\mathbb{K} = \mathbb{U} = [0, +\infty[$$
$$J(\mathbf{k}, \mathbf{U} = \mathbf{u}) = \mathbf{u}(\mathbf{k} - \mathbf{u})^2 + 1$$
$$\mathbf{k}^*(\mathbf{u}) = \mathbf{u}$$
$$J^*(\mathbf{u}) = 1$$

In this case, for each $\mathbf{k} \in \mathbb{K}$, $\lim_{\mathbf{u}\to+\infty} J(\mathbf{k}, \mathbf{U} = \mathbf{u})/J^*(\mathbf{u}) = +\infty$, so there is no $\mathbf{k}$ such that the ratio $J/J*$ can be bounded, hence α_1 does not exist in this case.

Looking for a compromise between p and α would be preferable. This could be achieved 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 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_{ratio} is then defined as the maximiser of $p \mapsto p/\alpha_p$.

3.3. Numerical Illustration

In this Section, we will compare the different estimators introduced previously and summarised in Table 1 on J_{BH} and J_{BHswap}.

Definition	Related quantities	Interpretation
$\arg\min_{\mathbf{k}\in\mathbb{K}} \mathbb{E}_{\mathbf{U}}[J(\mathbf{k}, \mathbf{U})]$	$\mathbf{k}_{\mathbb{E}}$	Long run performances
$\arg\min_{\mathbf{k}\in\mathbb{K}} \mathbb{V}\text{ar}_{\mathbf{U}}[J(\mathbf{k}, \mathbf{U})]$	$\mathbf{k}_{\mathbb{V}}$	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
$p_{\text{ratio}} = \arg\max p/\alpha_p$	$(p_{\text{ratio}}, \mathbf{k}_{\text{ratio}}, \alpha_{\text{ratio}})$	Maximal ratio of p and α_p

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

As stated before, we chose to model the uncertainties as a random variable uniformly distributed on $\mathbb{U}$. The bounded nature of $\mathbb{U}$ allows us to consider members of the RRE

up to a level of confidence $p = 1$. From now on, $\hat{\mathbf{k}}_{\mathrm{MPE}}$ is estimated using KDE with Gaussian kernels.

The smallest estimated relaxation $\hat{\alpha}_1$ and the corresponding $\hat{\mathbf{k}}_1$ has been computed for J_{BH} and J_{BHswap}, 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 $\{J(\mathbf{k}, \mathbf{U} = \mathbf{u}) = \alpha J^*(\mathbf{u})\}$ 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 changes. 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 histograms of the minimisers are represented.

One can notice that the relaxation allows us to avoid the issue brought by 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.

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 Section 3.2.

The plot of $p \mapsto \hat{\alpha}_p$ for J_{BH} 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 at $\hat{p}_{\mathrm{ratio}} = 0.654$ corresponds to the global maximum of the ratio. The RRE will then be evaluated for both of these values, as well as $p = 1$ for reference.

For J_{BH}, the numerical values of the robust estimators can be found in Table 2. For this particular problem, the different estimates are close to each other.

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

Estimate	Value
$\hat{\mathbf{k}}_{\mathbb{V}}$	1.371
$\hat{\mathbf{k}}_p,\ p = 1$	1.557
$\hat{\mathbf{k}}_{\mathbb{E}}$	1.587
$\hat{\mathbf{k}}_{\mathrm{MPE}}$	1.628
$\hat{\mathbf{k}}_{\mathrm{ratio}},\ \hat{p}_{\mathrm{ratio}} = 0.654$	1.637
$\hat{\mathbf{k}}_p,\ p = 0.90$	1.797

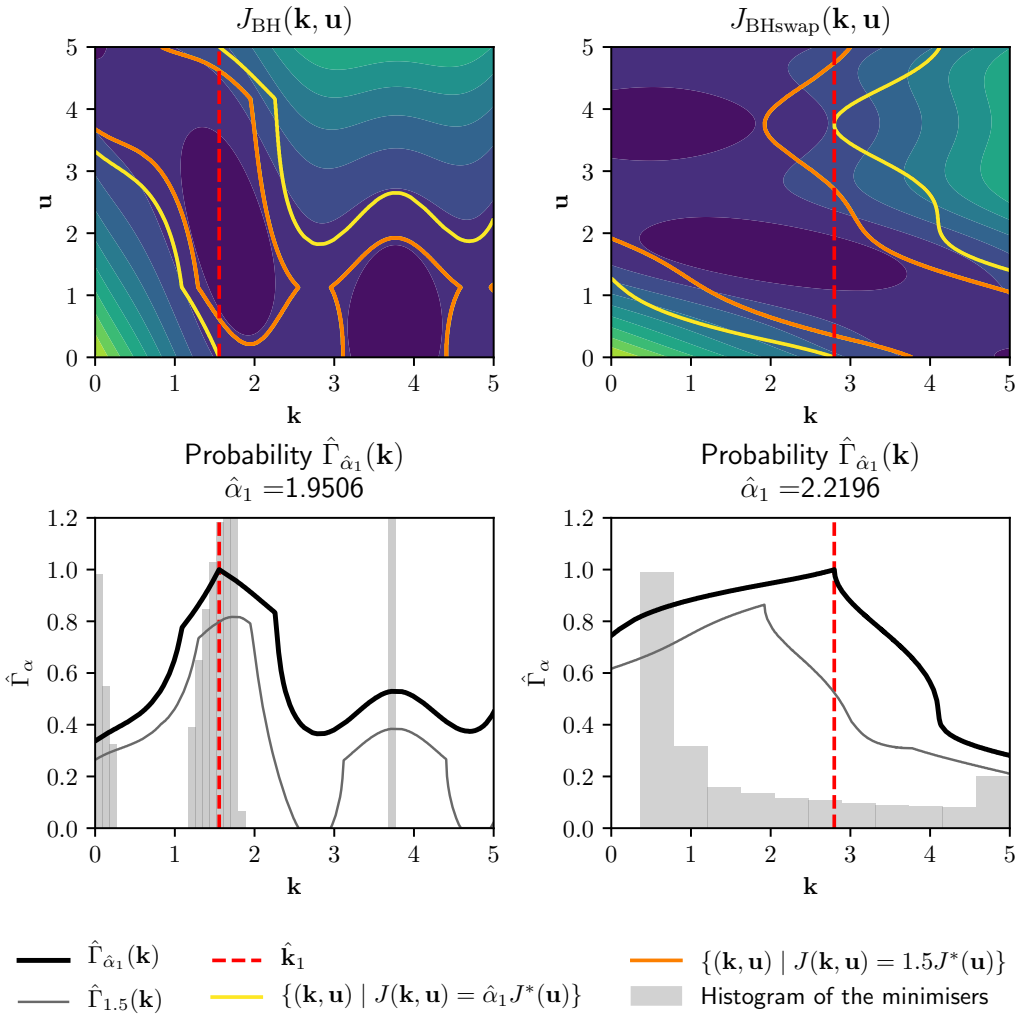

Figure 5: Top: J_{BH} and J_{BHswap} contours. The thick yellow lines are the boundaries of the acceptable region defined for $\hat{\alpha}_1$, the thick orange are for $\alpha = 1.5$. The red dashed line is the estimation $\hat{\mathbf{k}}_1$. Bottom: $\hat{\Gamma}_\alpha$ for $\alpha = 1.5$ and $\alpha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

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 $\mathbf{u} \mapsto J(\hat{\mathbf{k}}, \mathbf{U} = \mathbf{u})$, that we will call "profiles 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.

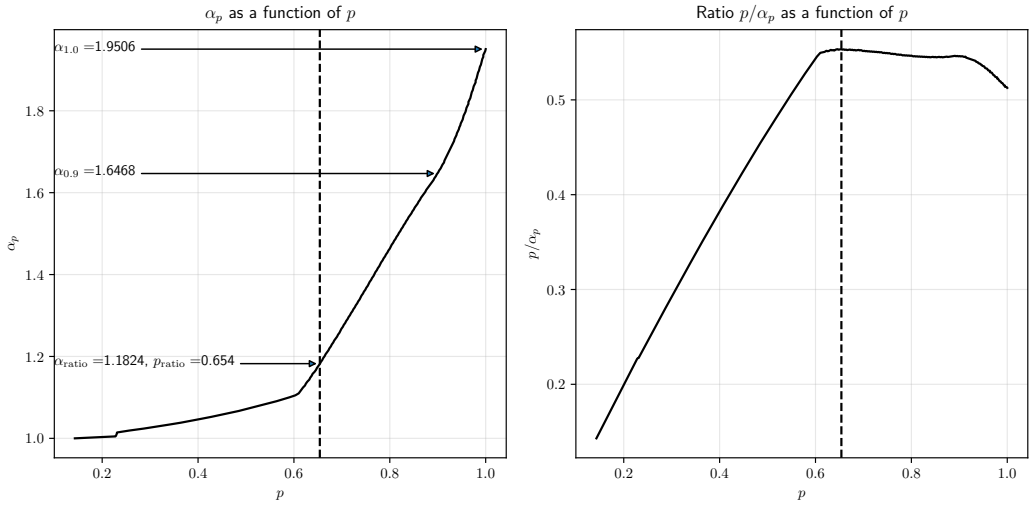

Figure 6: Evolution of the couples (p, α_p) and corresponding ratio p/α_p for J_{BH}. The dashed line indicates the level p associated with the highest ratio

For J_{BH}, the curves are plotted in Figure 7.

By construction, 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 profile of $\hat{\mathbf{k}}_{\mathbb{E}}$ in contrast, exceeds $\hat{\alpha}_1 J^*(\mathbf{u})$ for $\mathbf{u}$ close to 5, while the profile of $\hat{\mathbf{k}}_{\mathbb{V}}$ does it for $\mathbf{u}$ close to 0. Except for $\hat{\mathbf{k}}_{\mathbb{V}}$, the different estimators give somewhat comparable results.

By contrast, J_{BHswap} will show a different behaviour as Figure 8 provides the plots of $p \mapsto \hat{\alpha}_p$ and $p \mapsto p/\hat{\alpha}_p$.

Compared to the similar plots for J_{BH} in Figure 6, $\hat{\alpha}_p$ exhibits a smoother behaviour for J_{BHswap} as no abrupt change of slope is easily discernable and the ratio presents a unique maximum for $\hat{p}_{\text{ratio}} = 0.766$. The numerical values of the estimations $\hat{\mathbf{k}}$ presented in Table 3 show that, contrary to J_{BH}, the calibrated values are more spread over $\mathbb{K}$.

Profiles of the different estimates of J_{BHswap} are shown in Figure 9.

In this case, $\hat{\mathbf{k}}_{\text{MPE}}$, $\hat{\mathbf{k}}_{\mathbb{E}}$ and $\hat{\mathbf{k}}_{\text{ratio}}$ 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 $\hat{\mathbf{k}}_{\mathbb{V}}$ for this function, but it performs worse than $\hat{\mathbf{k}}_{\mathbb{E}}$ and $\hat{\mathbf{k}}_{\mathbb{V}}$ for $\mathbf{u} > 2$, even though its range is designed to stay within the interval $[J^*(\mathbf{u}); \hat{\alpha}_1 J^*(\mathbf{u})]$.

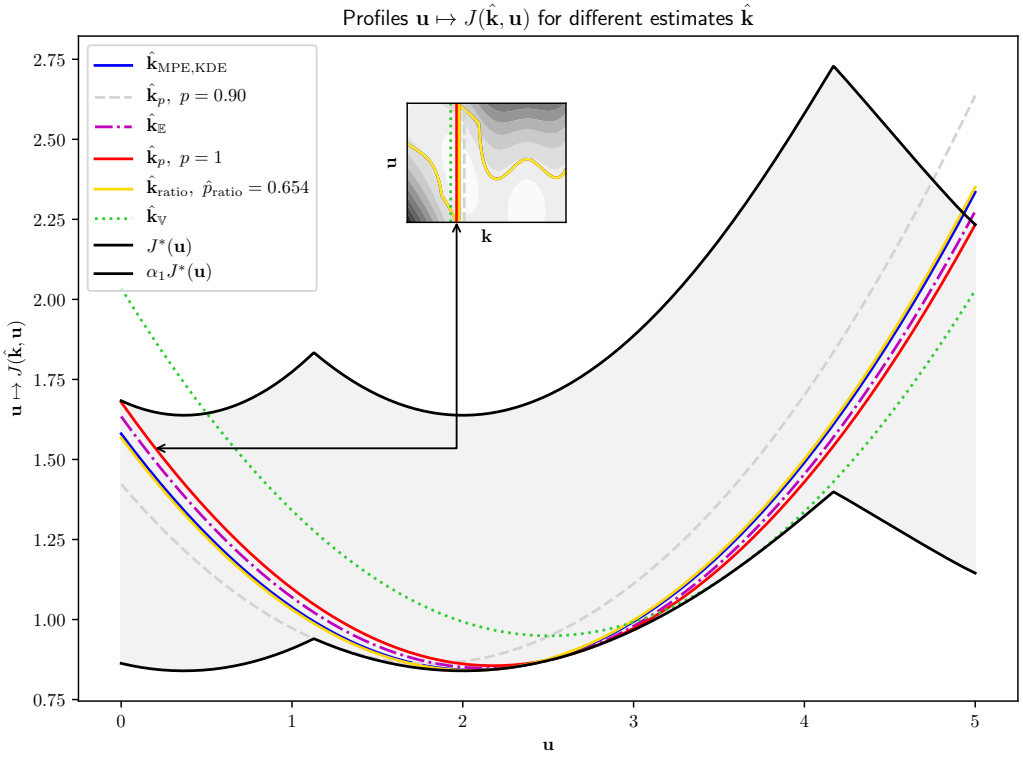

Figure 7: Profiles of the different estimates for J_{BH}, corresponding to the vertical cross sections of the contour. The shaded region corresponds to the interval $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$. The profiles of $\hat{\mathbf{k}}_{\mathrm{MPE}}$ and $\hat{\mathbf{k}}_{\mathrm{ratio}}$ coincide in this case

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

Estimate	Value
$\hat{\mathbf{k}}_{\mathrm{MPE}}$	0.606
$\hat{\mathbf{k}}_{\mathrm{ratio}},\ \hat{p}_{\mathrm{ratio}} = 0.766$	1.537
$\hat{\mathbf{k}}_{\mathbb{E}}$	1.752
$\hat{\mathbf{k}}_{\mathbb{V}}$	2.638
$\hat{\mathbf{k}}_1,\ p = 1$	2.798

We have seen how some classical robust estimators and the RRE behave on two different analytical problems. In addition to the usual levels of confidence such as 90%

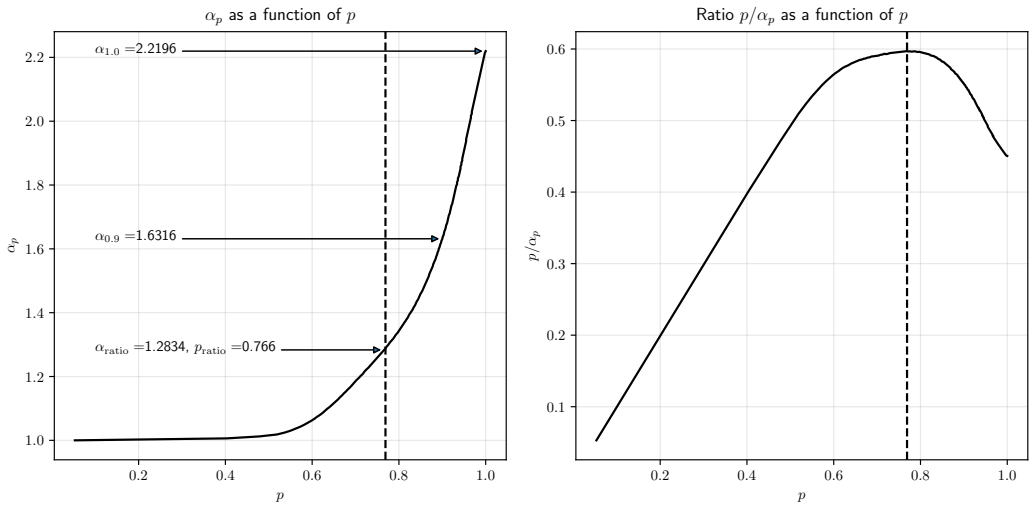

Figure 8: Evolution of the couples (p, α_p) and corresponding ratio p/α_p for J_{BHswap}. The dashed line indicates the level p associated with the highest ratio

and 95%, one can also settle for an ad-hoc compromise, where p maximizes the ratio p/α_p. When the sample space of $\mathbf{U}$ is bounded, a conservative solution is to set $p = 1$. We are now going to see how can robust minimisation is applied on the calibration of a numerical model.

4. Robust calibration of a numerical model

4.1. Calibration of a toy numerical model

We will follow the approach described in [30] in order to establish the function $\mathcal{G}$ described in the first section in Eq. (1), and the resulting cost function J.

The calibration of a numerical model is usually based on the comparison between the numerical model and some observations, during a fixed time interval $[0, T]$ called assimilation window. The modelled physical system can be seen as a map from $\mathbb{U}$ to $\mathbb{Y}$, the space of observations, denoted as $\mathcal{M}^o : \mathbf{u} \mapsto \mathcal{M}^o(\mathbf{u})$, where $\mathbf{u} \in \mathbb{U}$ is an input representing some environmental conditions. The observation mentioned above is the output of the physical system during the time-window, and is denoted by $\mathcal{M}^o(\mathbf{u}^{\text{true}}) \in \mathbb{Y}$, where $\mathbf{u}^{\text{true}} \in \mathbb{U}$ is unknown.

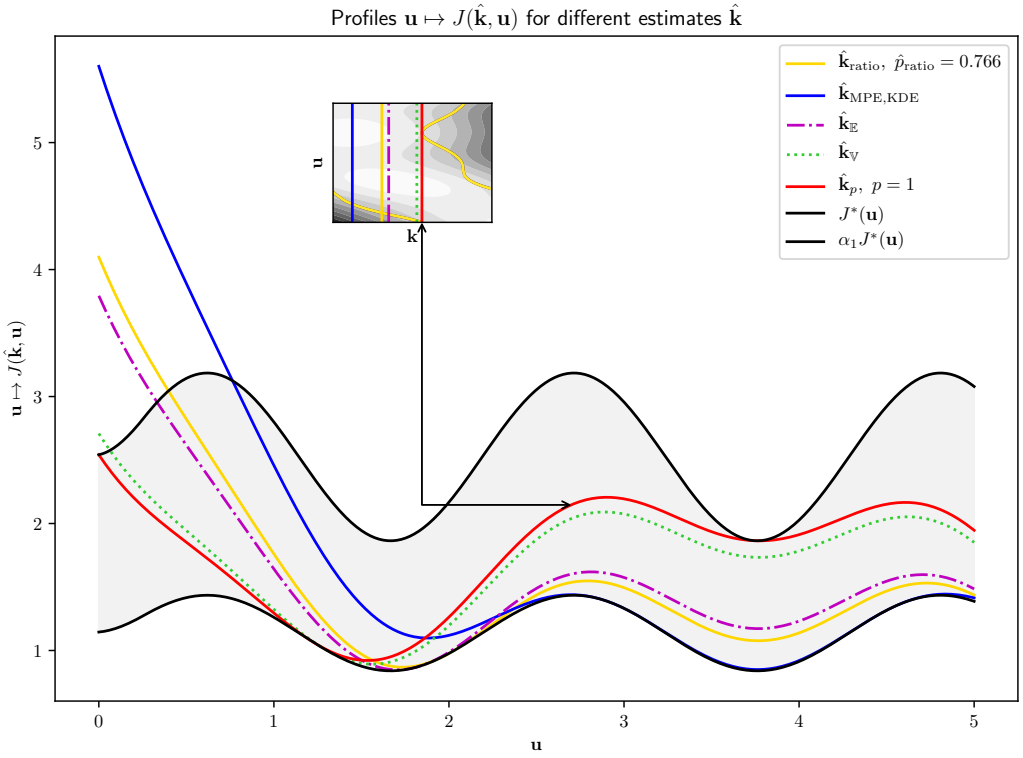

Figure 9: Profiles of the different estimates for J_{BHswap}. Those profiles are the vertical cross sections of the contours above. The shaded region corresponds to the interval $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$

In addition of $\mathbf{u}$, the numerical model $\mathcal{M}$ depends on some other input $\mathbf{k} \in \mathbb{K}$. This additional parametrization comes usually from the successive simplifications needed to implement a numerical model of the observed physical system. $\mathbf{k}$ needs 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 misfit $\mathcal{G}$ is defined as the difference between the numerical model and the observation. Choosing a squared norm, the cost function J defined in Eq. (1) is

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

4.2. The Shallow Water equations

The model to calibrate is an implementation of the 1D Shallow Water equations, described in Eq. (23), where h is the height of the water column, q the discharge, and z 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 $\mathbb{K} = [0.0, 1.3]$ and $\mathbb{U} = [0.5, 0.7]$.

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

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 random variable $\mathbf{U}$ is uniformly distributed on $\mathbb{U}$.

To generate the observation, we set $\mathbf{u}^{\text{true}} = 2/3$, 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 systematic errors are the one-dimensionality of $\mathbb{K}$, and ω_0. Given this setting, there exists no couple $(\mathbf{k}, \mathbf{u}) \in \mathbb{K} \times \mathbb{U}$ reproducing exactly the observations, thus the cost function will always be strictly positive.

4.3. Computation of the robust estimates

As the numerical model is expensive to run, it has first been evaluated on a relatively small regular grid covering $\mathbb{K} \times \mathbb{U}$, and a metamodel based on Gaussian processes is

constructed using those initial evaluations. In order to better capture the locus of the conditional minimisers $\{(\mathbf{k}^*(\mathbf{u}), \mathbf{u}) \mid \mathbf{u} \in \mathbb{U}\}$, we select points maximising the PEI criterion [31, 32] in order to add points to the design that improve the metamodel accuracy close to the conditional minimisers. Afterwards, a bigger regular grid is evaluated by the metamodel once the design space has been sufficiently explored, and these computations are used to calibrate the model.

The different steps of the estimation are illustrated Figure 10, where the top plot shows the contour plot of J, with the conditional minimisers and the RRE of level $p = 1$. On the middle plot are shown the conditional moments, $\hat{\mathbf{k}}_{\mathbb{E}}$ and $\hat{\mathbf{k}}_{\mathbb{V}}$ and the estimated density of the minimisers. At the bottom of the Figure, $\hat{\Gamma}_{\alpha_p}$ are represented for different levels p. As $\mathbb{U}$ is bounded, $p = 1$ is attainable, and $\hat{\mathbf{k}}_1$ is evaluated.

We can see that the different estimations seem less problematic for this problem than for the analytical examples shown in Section 3.3, as everything appears to be very smooth and unimodal.

Looking at $p \mapsto \alpha_p$ and $p \mapsto p/\alpha_p$ on Figure 11, we can see that α_p evolves almost linearly for $p > 0.5$, and that the ratio p/α_p is monotonically increasing, so that the maximal ratio is found for $p = 1$.

The different estimates take a wide range of values, as seen on Table 4, from 0.249 to almost 1.0, while $k_i^{\text{true}} \in [0, 0.4]$ for $1 \leq i \leq N_{\text{vol}}$. It can be noted that the calibration may lead to values outside the range given by the true ones. This can be interpreted as the fact that the calibration look to compensate for errors on ω_0. As a basis for comparison, the global minimiser of J over $\mathbb{K} \times \mathbb{U}$ has also been computed, and $\hat{\mathbf{k}}_{\text{global}}$ is then obtained by discarding the $\mathbf{u}$ value.

Similarly as for J_{BH} and J_{BHswap}, the profiles are depicted Figure 12.

We can see that the performances of $\hat{\mathbf{k}}_1$ and $\hat{\mathbf{k}}_{\mathbb{E}}$ are very similar, but $\hat{\mathbf{k}}_1$ has better performances when $\mathbf{u} > 0.6$. When comparing with the global optimiser $\hat{\mathbf{k}}_{\text{global}}$, $\hat{\mathbf{k}}_1$ performs better when $\mathbf{u} < 0.625$, so more than half of the time. We are now going to compare how well some of those calibrated values compare in a forecast context.

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

For the calibration, the model has been integrated on a time-period $[0, T]$, called assimilation window, and have been compared with the observation of the sea water

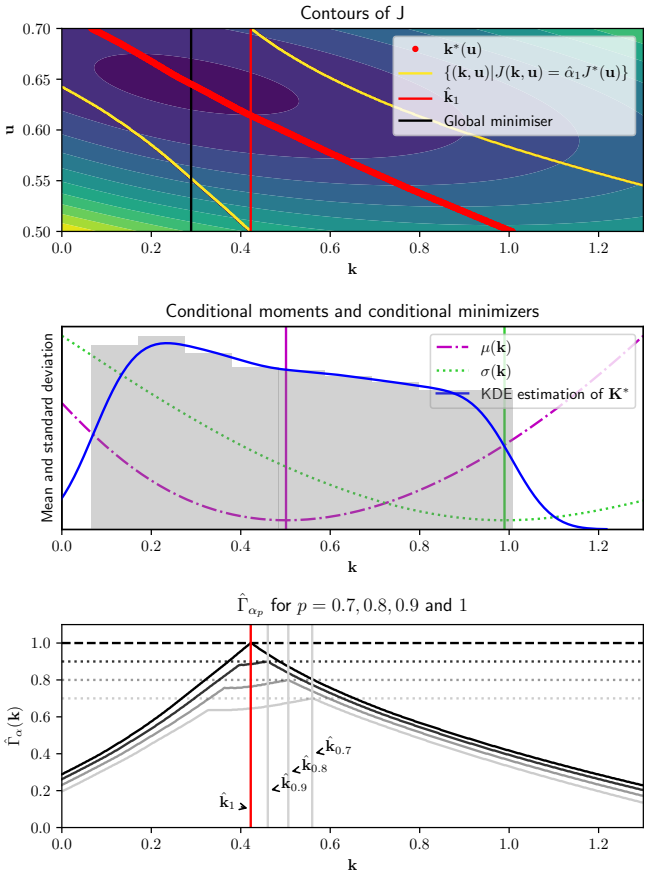

Figure 10: Procedure of robust calibration for the shallow water problem. Top: contours of J, conditional minimisers and $\{(\mathbf{k}, \mathbf{U} = \mathbf{u})|J(\mathbf{k},\mathbf{u}) = \hat{\alpha}_1 J^*(\mathbf{u})\}$. Middle: Conditional moments and histogram and KDE of the conditional minimisers.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

height on the same time-period. We now want to compare the quality of the different forecasts originating from different calibrated bottom frictions. Those forecasts result from the integration of the numerical model between the time T and a time T_{pred}: the forecast window.

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 $\mathcal{M}_{\text{pred}}(\mathbf{k}, \mathbf{U})$, that are the observation, and the numerical model on the forecast window for a calibrated $\mathbf{k}$.

Two metrics will be computed: the squared forecast error, and the Continuous Ranked

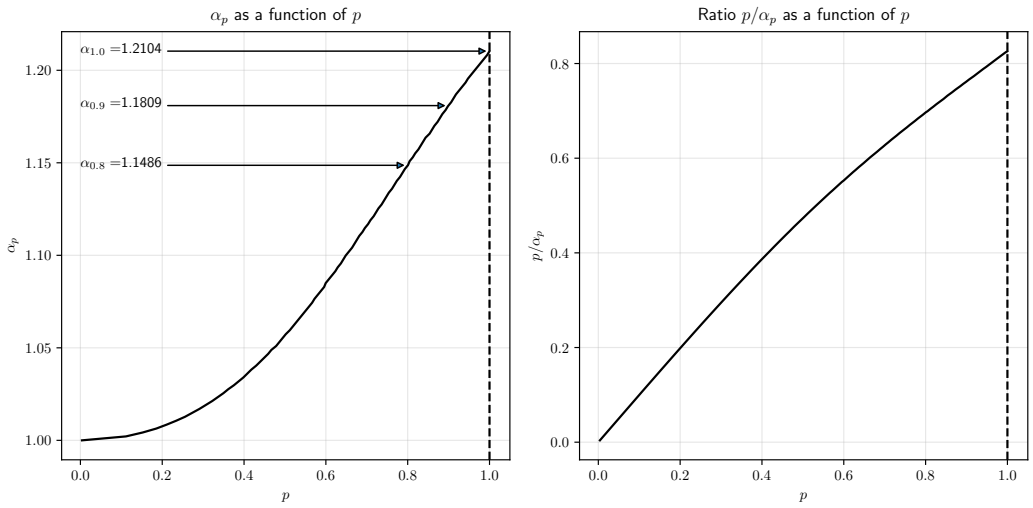

Figure 11: Evolution of α_p for different levels p, and ratio, for the shallow water problem

Table 4: Calibrated values of $\mathbf{k}$ according to different criteria, for the shallow water problem

Estimate	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
$\hat{\mathbf{k}}_{\mathbb{E}}$	0.501
$\hat{\mathbf{k}}_p,\ p = 0.80$	0.505
$\hat{\mathbf{k}}_p,\ p = 0.70$	0.560
$\hat{\mathbf{k}}_{\mathbb{V}}$	0.990

Probability Score (CRPS) [33]. The parameters satisfying a robust criterion that are compared are the minimiser of the expectation $\hat{\mathbf{k}}_{\mathbb{E}}$, the minimiser of the variance $\hat{\mathbf{k}}_{\mathbb{V}}$ and the relative regret of level $p = 1$, $\hat{\mathbf{k}}_1$ as described in the previous section. We will also feature the global minimiser $\hat{\mathbf{k}}_{\text{global}}$, and the conditional minimisers $\hat{\mathbf{k}}^*(\mathbf{u}^i)$ where the chosen environmental variables are $\{\mathbf{u}^i\}_{1\leq i\leq 4} = \{0.5, 0.55, 0.65, 0.7\}$. Those values, even though they do not meet a robustness criterion, are introduced in order to have a more

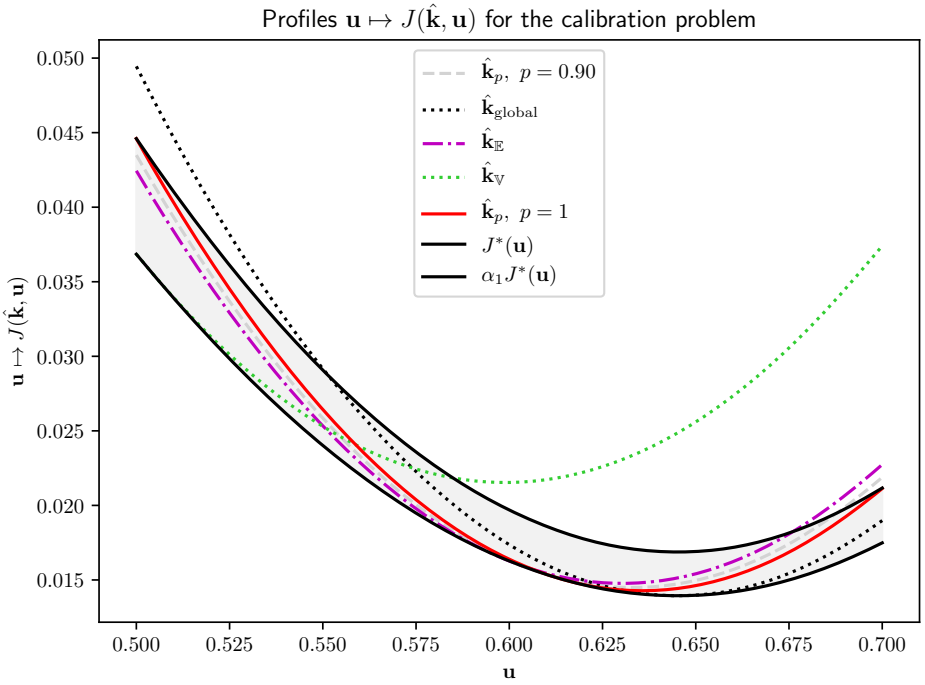

Figure 12: Profiles for the calibration problem, for the shallow water case

precise idea on the possible performances of a deterministic version of the calibration problem. The conditional minimiser $\hat{\mathbf{k}}^*(0.6)$ has been omitted as it results in a value very similar to the minimum of the expectation.

4.4.1. Squared forecast error

A first simple approach to measure the forecast quality is to take the squared difference between the numerical model and the observation for two samples of $\mathbf{U}$. Given two environmental conditions $\mathbf{u}$ and $\mathbf{u}' \in \mathbb{U}$, the former used to run the computer simulation and the latter to generate the observations, the squared forecast error for the parameter $\mathbf{k}$ is

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

Averaging over both $\mathbf{u}$ and $\mathbf{u}'$ defines the mean squared forecast error, defined on

every point of the spatial domain, and at every time-steps. This can be done using a Monte-Carlo approximation. As $\mathbf{u}$ and $\mathbf{u}'$ are i.i.d., assuming that we have a set of samples $\{\mathbf{u}^i\}_{1\leq i\leq N_\mathbf{u}}$, the squared forecast error can then be approximated:

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

The mean squared forecast error averaged over the whole space and over all the time-steps gives an indication on the overall prediction quality of the prediction given this metric, and are represented on the right of Figure 13. We can see that $\hat{\mathbf{k}}_{\text{global}}$ performs slightly better than $\hat{\mathbf{k}}_1$, itself performing slightly better than $\hat{\mathbf{k}}_{\mathbb{E}}$. Averaging $S(\hat{\mathbf{k}})$ over the time steps between T and T_{pred}, we have an indication on the quality of the forecast in the squared sense depending on the spatial coordinate x, as seen on the left of Figure 13.

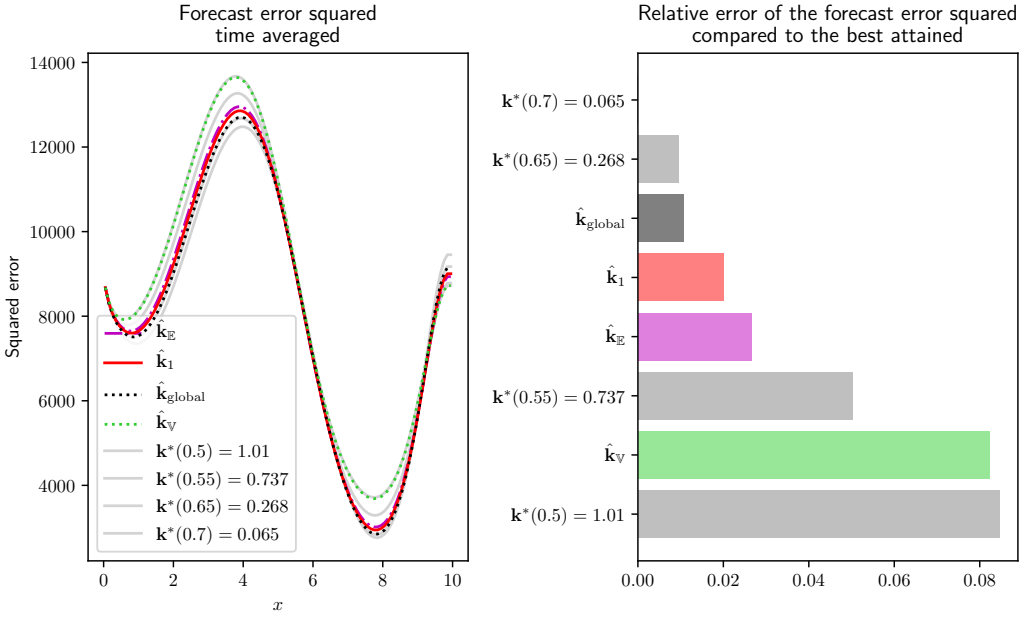

Figure 13: Squared forecast error for the shallow water case, depending on the calibrated parameter. The left figure shows the squared error time-averaged as a function of the spatial coordinate x, while the right barplot shows the relative change of the mean forecast error, averaged over time and space, taken w.r.t. to the best attained performance for $\mathbf{k}^*(0.7)$

We can see that the mean forecast error squared, on the right side, averaged over

time and space is the smallest for the conditional minimisers $\hat{\mathbf{k}}^*(0.7)$ and $\hat{\mathbf{k}}^*(0.65)$, then $\hat{\mathbf{k}}_{\text{global}}$, while $\hat{\mathbf{k}}_{1.0}$ performs slightly better than $\hat{\mathbf{k}}_{\mathbb{E}}$.

It may seem surprising that some parameters calibrated without a robustness aspect perform better than $\hat{\mathbf{k}}_1$ and $\hat{\mathbf{k}}_{\mathbb{E}}$, but their performances are largely dependent on the choice of $\mathbf{u}$ and their associated conditional minimisers. Good forecasts can be achieved as well as bad ones, as shown for $\mathbf{u} = 0.5$, that leads to bad forecasts.

One issue with the squared forecast error is that it will penalize strongly the forecasts that present high variability with respect to the environmental conditions. A way to deal with this is to use another metric, that takes the probabilistic nature of the forecasts into account.

4.4.2. 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}.

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

The left plot of Figure 14 shows the CRPS averaged over time, where x denotes the spatial coordinate and the right plot shows the value of the CRPS averaged over time and space. The difference between the squared forecast error and the CRPS is apparent when comparing the general trends shared by the different calibrated parameters. According to the squared error, the sea water height of the physical region $x = 4$ is not well predicted, while around $x = 8$, the predictions are better. On the other hand, according to the CRPS, the region around $x = 8$ provide worse forecasts than when $x = 4$ and $x = 7$. Given the properties of the two metrics, we can conclude that the region around $x = 4$ presents a lot of variability with respect to $\mathbf{u}$, for both the true model and the numerical one. However, for $x \approx 8$, there is a lot less variability, as the low squared error indicates, but probably a higher bias, due to the systematic errors between the truth and the numerical model.

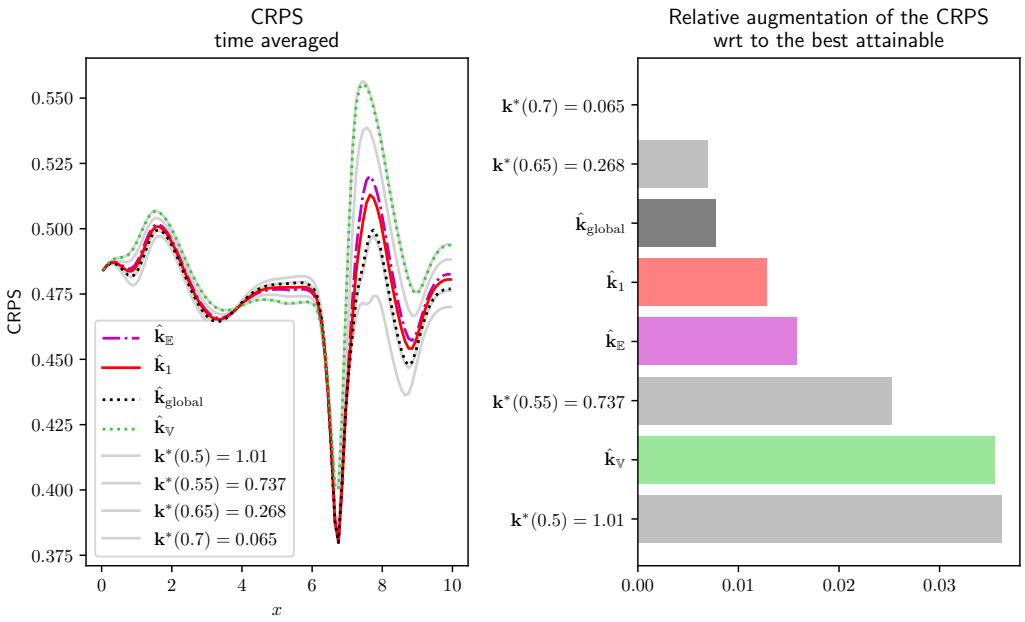

Figure 14: CRPS computed for different calibrated parameters for the shallow water case. The left figure shows the CRPS time-averaged as a function of the spatial coordinate x, and the right figure shows the relative change of the CRPS averaged over time and space, taken with respect to the best attained CRPS for $\mathbf{k}^*(0.7)$

The numerical evaluations of the CRPS for different parameters show the same order of performances observed for the squared error: the calibrated parameters that present the best performances for forecasts according to those two metrics are $\hat{\mathbf{k}}^*(0.7)$, $\hat{\mathbf{k}}^*(0.65)$ which is very similar to $\hat{\mathbf{k}}_{\text{global}}$, and then $\hat{\mathbf{k}}_1$, and $\hat{\mathbf{k}}_{\mathbb{E}}$. This presupposes to know which $\mathbf{u}$ to choose for the conditional minimisation, thus having a strong insight on the value of the parameters in the first place.

Conclusion

In this paper, we dealt with a problem of robust calibration, or in other terms, the robust minimisation of an objective function. To adress this issue, we proposed a new set of robust estimators: the family of the RRE (Relative-regret estimators), and compared it in a forecast context to some other robust estimators, for the calibration of the bottom

friction of a shallow water model.

Our approach is based on the will to be as close as possible to the conditional minimisers with high probability. At a level p, in addition to the estimation, we have the relaxation coefficient α_p that bounds with probability p the relative error of the cost function. This helps us to assess the robustness properties of the proposed estimation.

Except for simple analytical examples, the exact evaluation of a member of the RRE is very expensive computer-wise, not to say impossible. Some bottlenecks appear: the computations of the conditional minimisers $\mathbf{k}^*(\mathbf{u})$ and the minimum $J^*(\mathbf{u})$, that require very local exploration of $\mathbb{K}$ for every $\mathbf{u}$; and the computations of the probability of overshooting a given bound that depends on $J^*(\mathbf{u})$ (analogous to a probability of failure), task that requires an efficient exploration of the whole sample space $\mathbb{U}$.

For the calibration of the shallow water model presented in Section 4, we chose to construct a response surface based on Gaussian processes, which is then used for the extensive computations. This construction is first based on an initial design, that is enriched to better capture the locus of the conditional minimisers using the PEI criterion introduced in [31].

Comparing the performances of the different estimates in a forecast context, one can see that for the two metrics introduced in Section 4, $\hat{\mathbf{k}}_1$ performs better than $\hat{\mathbf{k}}_{\mathbb{E}}$, but is sometimes outperformed by "non-robust" solutions. Those require some additional knowledge about the environmental conditions, and did not present some robust properties during the calibration phase.

Practically speaking, the estimation of those quantities is very crude, so the computational cost may be very expensive. As perspective, specific strategies have to be created, based for instance on sequential design of computer experiments using Gaussian processes, to diminish the amount of runs needed of the numerical model. Ideally, one would want to develop an acquisition criterion, that makes the balance between the minimisation of a well-chosen probability of failure, and the exploration around the conditional minimisers. Finally, the models upon which this calibration procedure have been applied are very simplistic. We plan in the future to apply this to the robust calibration of the bottom friction of a realistic model of the ocean.

References

[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 (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 (1991) 15187.

[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 (2001).

[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 (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] (2014).

[8] L. W. Cook, J. P. Jarrett, Horsetail matching: A flexible approach to optimization under uncertainty, Engineering Optimization 50 (2018) 549–567.

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

[10] V. Rao, A. Sandu, M. Ng, E. Nino-Ruiz, Robust data assimilation using L_1 and Huber norms, SciRate (2015).

[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 (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 (2013) 707–727.

[13] J. Villemonteix, E. Vazquez, E. Walter, An informational approach to the global optimization of expensive-to-evaluate functions, arXiv:cs/0611143 (2006).

[14] P. Hennig, C. J. Schuler, Entropy Search for Information-Efficient Global Optimization (2011).

[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, 2013, p. 505. doi:10.1145/2422436.2422491.

[16] P. Kouvelis, A. A. Kurawarwala, G. J. Gutiérrez, Algorithms for robust single and multiple period layout planning for manufacturing systems, European Journal of Operational Research 63 (1992) 287–303.

[17] L. Snyder, M. Daskin, Stochastic p-robust location problems, Iie Transactions 38 (2004).

[18] A. Juditsky, A. S. Nemirovski, G. Lan, A. Shapiro, Stochastic Approximation Approach to Stochas-

tic Programming, in: ISMP 2009 - 20th International Symposium of Mathematical Programming, 2009.

[19] S. Kim, R. Pasupathy, S. Henderson, A Guide to Sample Average Approximation, volume 216, 2015, pp. 207–243. doi:10.1007/978-1-4939-1384-8-8.

[20] J. Janusevskis, R. Le Riche, Simultaneous Kriging-Based Sampling for Optimization and Uncertainty Propagation, Technical Report, 2010.

[21] V. Baudoui, Optimisation Robuste Multiobjectifs Par Modèles de Substitution, Ph.D. thesis, Toulouse, ISAE, 2012.

[22] O. Grodzevich, O. Romanko, Normalization and other topics in multi-objective optimization (2006).

[23] R. T. Marler, J. S. Arora, The weighted sum method for multi-objective optimization: New insights, Structural and Multidisciplinary Optimization 41 (2010) 853–862.

[24] J. S. Lehman, T. J. Santner, W. I. Notz, Designing computer experiments to determine robust control variables, Statistica Sinica (2004) 571–590.

[25] B. W. Silverman, Density estimation for statistics and data analysis, Routledge, 2018.

[26] 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.

[27] D. Freedman, P. Diaconis, On the histogram as a density estimator:L2 theory, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 57 (1981) 453–476.

[28] D. W. Scott, On Optimal and Data-Based Histograms, Biometrika 66 (1979) 605.

[29] 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.

[30] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (2001) 425–464.

[31] 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 (2014) 490–510.

[32] 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, 2015, pp. 1319–1326. doi:10.1145/2739480.2754673.

[33] T. Gneiting, M. Katzfuss, Probabilistic Forecasting, Annual Review of Statistics and Its Application 1 (2014) 125–151.