



HAL
open science

Robust calibration of numerical models based on relative regret

Victor Trappier, Élise Arnaud, Arthur Vidard, Laurent Debreu

► **To cite this version:**

Victor Trappier, Élise Arnaud, Arthur Vidard, Laurent Debreu. Robust calibration of numerical models based on relative regret. 2020. hal-02464780v2

HAL Id: hal-02464780

<https://hal.science/hal-02464780v2>

Preprint submitted on 19 Oct 2020 (v2), last revised 6 Nov 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Robust calibration of numerical models based on relative regret

2 Victor Trappler^{a,*}, Élise Arnaud^a, Arthur Vidard^a, Laurent Debreu^a

3 ^a*Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP¹, LJK, 38000 Grenoble, France*

4 Abstract

5 Classical methods of **calibration** usually imply the minimisation of an objective function
6 **with respect to some control parameters**. This function measures the error between
7 some observations and the results obtained by a numerical model. In the presence of
8 **uncontrollable additional parameters that we model as random inputs**, the objective
9 function becomes a random variable, and notions of robustness have to be introduced **for**
10 **such an optimisation problem**.

11 In this paper, we are going to present how to take into account those uncertainties by
12 defining the relative-regret. This quantity allow us to compare the value of the objective
13 function to its best performance achievable given a realisation of the random additional
14 parameters. By controlling this relative-regret using a probabilistic constraint, we can
15 then define a new family of estimators, whose robustness with respect to the random
16 inputs can be adjusted.

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

18 1. Introduction

19 Numerical models are widely used to study or forecast natural phenomena and im-
20 prove industrial processes. However, by essence models only partially represent reality
21 and sources of uncertainties are ubiquitous (discretisation errors, missing physical pro-
22 cesses, poorly known boundary conditions). Moreover, such uncertainties may be of
23 different nature. [1] proposes to consider two categories of uncertainties. On the one

*Corresponding author

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

¹Institute of Engineering Univ. Grenoble Alpes

24 hand aleatoric uncertainties, coming from the inherent variability of a phenomenon, e.g.
25 intrinsic randomness of some environmental variables. On the other hand, epistemic
26 uncertainties coming from a lack of knowledge about the properties and conditions of
27 the phenomenon underlying the behaviour of the system under study. The latter can
28 be accounted for through the introduction of ad-hoc correcting terms in the numerical
29 model, that need to be properly estimated. Thus, reducing the epistemic uncertainty can
30 be done through parameters estimation approaches. This is usually done using optimal
31 control techniques, leading to an optimisation of a well chosen cost function which is
32 typically built as a comparison with reference observations. An application of such an
33 approach, in the context of ocean circulation modelling, is the estimation of ocean bottom
34 friction parameters, as done in [2, 3, 4]. Moreover, as such studies are often performed
35 at a coastal and regional scale, those models are often designed with open boundary
36 conditions, and require external forcings, such as tidal and wind forcing. Those are then
37 a source of aleatoric uncertainties, that should be taken into account as in [5].

38 If we overlook the aleatoric uncertainties by choosing a specific outcome, the opti-
39 mal control of the parameters to be estimated can lead to *localized optimisation* [6] and
40 *overcalibration*, that is choosing a value that is optimal for the given situation. This
41 value does not carry the optimality to other situations. In geophysics and especially in
42 hydrological models, this overcalibration may lead to the appearance of aberrations in
43 the predictions as those uncertainties become prevalent sources of errors. In hydrology,
44 uncertainties are the principal culprit of the existence of so called “Hydrological mon-
45 sters” [7], that are calibrated models that perform really badly because some uncertainties
46 in the modelling have been omitted, such as measurements errors. In flood modelling [8],
47 the aleatoric uncertainty come from the structure of the model, and neglecting those
48 uncertainties leads to underestimating hazard. In other domains as well, the aleatoric
49 uncertainties can represent some manufacturing errors or environmental conditions such
50 as wind direction and speed in wind turbine modelling [9] or atmospheric conditions in
51 aerospace vehicle design [10], thus they represent an important role in risk management.

52 It is then crucial to be able to take into account aleatoric uncertainties in optimi-
53 sation problems. This consideration is called *robust optimisation*, or also *robust design*
54 in [11], or *optimisation under uncertainties* [12, 13, 14]. Furthermore, the distinction is

55 sometimes made between *stochastic optimisation* and *robust optimisation*, depending on
56 the knowledge available on the aleatoric uncertainties.

57 Let us denote $\mathbf{k} \in \mathbb{K}$ the parameter to be estimated in order to reduce epistemic
58 uncertainties. We assume that the aleatoric uncertainties can be modelled as a random
59 vector \mathbf{U} whose sample space is \mathbb{U} . The probability measure associated with \mathbf{U} is $\mathbb{P}_{\mathbf{U}}$,
60 and its density, if it exists, is $p_{\mathbf{U}}$. This distribution is assumed to be known, and that it
61 is possible to sample from it. This choice is motivated by the fact that in various appli-
62 cations, the aleatoric uncertainties come from expert knowledge, empirical observations,
63 or the knowledge acquired using other models, for instance by ensemble assimilation.
64 However, if it is not the case, [15] provides a comprehensive review of optimisation un-
65 der uncertainty, especially on the modelling and sampling of the aleatoric uncertainties.
66 Since the source of the aleatoric uncertainties is considered external, the choice of \mathbf{k} does
67 not have any influence on the distribution of the random variable \mathbf{U} , and therefore the
68 aleatoric and epistemic uncertainties are assumed independent.

69 The cost function $J(\mathbf{k}, \mathbf{U})$ is a random variable in this context. It is most often
70 defined as the squared norm of a given function $\mathcal{G}(\mathbf{k}, \mathbf{U})$

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

71 For instance, in data assimilation, J describes a distance between the output of the
72 numerical model and given observed data, plus generally some regularization terms. An
73 example of such a function will be treated Section 4.

74 For a practical purpose, we assume that for every realisation $\mathbf{u} \in \mathbb{U}$ of \mathbf{U} , finding
75 $\mathbf{k}_{\mathbf{u}} \in \mathbb{K}$ that minimises the cost function $\mathbf{k} \mapsto J(\mathbf{k}, \mathbf{U} = \mathbf{u})$ is a well-posed problem, and
76 that the optimum is unique. Additionally, the following assumptions are made: the cost
77 function is strictly positive, and $\forall \mathbf{k} \in \mathbb{K}$, the random variable $J(\mathbf{k}, \mathbf{U})$ has finite first-
78 and second-order moments.

79 In this paper, we aim at finding $\hat{\mathbf{k}}$ a *robust* estimator of \mathbf{k} . The definition of robustness
80 differs depending on the context in which it is used. Indeed, one definition of the robust-
81 ness of an estimate is a measure of the sensibility of said estimate to outliers [16]. This
82 leads to the introduction of robust norms in data assimilation [17]. In a Bayesian frame-
83 work, robustness may refer to the sensitivity to a wrong specification of the priors [18].

84 As the distribution of \mathbf{U} is assumed known, we are not going to consider *distributional*
85 *robustness*, which deals with uncertainties on the distribution of \mathbf{U} [19].

86 Throughout this paper, robust has to be understood as satisfactory for a broad range
87 of \mathbf{u} , and/or as insensitive as possible to uncertainties encompassed in \mathbf{U} .

88 The usual practice consists in neglecting the variability of \mathbf{U} by setting it to an *a*
89 *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
90 no guarantee on the performance of $\hat{\mathbf{k}}$ if the calibrated model is used for predictions, as
91 the estimated value will compensate the error made by a possibly wrong specification of
92 \mathbf{u}^b : *this may be a case of overcalibration*. In a data assimilation context, this situation
93 appears if \mathbf{u}^b does not properly represent the conditions in which the observations have
94 been obtained.

95 Another strategy, that consists in minimising J over the joint space $\mathbb{K} \times \mathbb{U}$, is not
96 always possible or relevant. *Indeed, joint optimisation increases* the complexity of the
97 optimisation, and the computed estimation of $\hat{\mathbf{k}}$ has no reason to be robust in the end:
98 this kind of method does not take into consideration the *intrinsic* variability of the *en-*
99 *vironmental* variable. The worst-case approach [20] is another popular method, and is
100 based on the minimisation with respect to \mathbf{k} of the maximum of the cost function for
101 $\mathbf{u} \in \mathbb{U}$: $\min_{\mathbf{k}} \max_{\mathbf{u}} J(\mathbf{k}, \mathbf{U} = \mathbf{u})$. This approach may yield over-conservative solutions,
102 and does not take into account the random nature of \mathbf{U} .

103 Accounting for the probabilistic nature of \mathbf{U} leads to study the distribution of the
104 random variable $J(\mathbf{k}, \mathbf{U})$, or the distribution of its minimisers $\mathbf{k}_{\mathbf{U}}$. The latter is referred
105 as the distribution of the conditional minimisers, notion that appeared notably in [21] and
106 in [22] for a global optimisation purpose. Both approaches and related robust estimates
107 are described in Section 2. Section 3 introduces a new class of estimators, by relaxing the
108 constraint of optimality and defining regions of acceptability, similarly as [23] in discrete
109 combinatorial problems, or [24, 25] in operations research. *The rationale behind this*
110 *relaxation is to be able to construct an estimate $\hat{\mathbf{k}}$ which produces values of the cost*
111 *function close enough from the minimal cost attainable given the configuration induced*
112 *by $\mathbf{u} \in \mathbb{U}$, with high enough probability. This similarity will be measured using the*
113 *relative regret. By adjusting either the relaxation or the confidence level, we can then*
114 *define the RRE, the relative-regret family estimators. Illustration of the various described*

115 methods are given on a numerical [example](#) in Section 4.

116 2. Classical robust estimators

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

121 2.1. Optimisation of the moments

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

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

$$\sigma^2(\mathbf{k}) = \text{Var}_{\mathbf{U}} [J(\mathbf{k}, \mathbf{U})] = \int_{\mathbf{U}} (J(\mathbf{k}, \mathbf{u}) - \mu(\mathbf{k}))^2 p_{\mathbf{U}}(\mathbf{u}) \, d\mathbf{u} \quad (3)$$

124 Minimising the expectation leads to the estimate $\mathbf{k}_{\mathbb{E}}$ defined by:

$$\mathbf{k}_{\mathbb{E}} = \arg \min_{\mathbf{k} \in \mathbb{K}} \mu(\mathbf{k}) \quad (4)$$

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

$$\mathbf{k}_{\mathbb{V}} = \arg \min_{\mathbf{k} \in \mathbb{K}} \sigma^2(\mathbf{k}) \quad (5)$$

127 A lot of different methods are readily available to solve these minimisation problems.
 128 For instance, stochastic Sample Approximation [26, 27] is based on a finite and fixed set
 129 of samples $\{\mathbf{u}^i\}_{i=1 \dots N}$ of \mathbf{U} . The estimations at a given \mathbf{k} are computed using standard
 130 Monte Carlo, resulting in the following optimisation problems:

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

131 and

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

132 If computationally affordable, one can perform these estimations on a regular grid
 133 on $\mathbb{K} \times \mathbb{U}$. In case of expensive computer code, one can build a meta model to ease the
 134 minimisation, such as Gaussian processes [28]. Even though $\mathbf{k}_{\mathbb{E}}$ is a reasonable choice,
 135 there is no guarantee that $J(\mathbf{k}_{\mathbb{E}}, \mathbf{U} = \mathbf{u})$ will not reach catastrophic level for some \mathbf{u} .
 136 On the other hand, using $\mathbf{k}_{\mathbb{V}}$ will ensure stability of the cost function, but without any
 137 control of its performance. Ideally, one would want to have a small expectation and a
 138 small variance at the same time. Multi-objective optimisation is a proper tool to deal
 139 with these simultaneous and sometimes concurrent objectives, for exemple by computing
 140 the Pareto front of $(\mu(\mathbf{k}), \sigma^2(\mathbf{k}))$ as done in [29].

141 As the computation of this Pareto front is usually hard and expensive, alternative
 142 strategies based on the minimisation of a scalarized version of the vector of objectives
 143 are often considered. Some are based on a weighted sum of the objectives, as presented
 144 in [30] and in [31], while some others are based on the minimisation of one of the objectives
 145 under constraints on the others, as performed in [32]. Both of these methods are based
 146 on an delicate choice of weights or of constraints before any computation. This choice
 147 relies heavily on a knowledge of the properties of the cost function.

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

152 2.2. Most probable estimate

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

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

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

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

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

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

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

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

160 To give some intuition on this estimate, let us imagine that the distribution of min-
 161 imisers is a dirac centered on \mathbf{k}_{MPE} . Then it would mean that this estimate is the
 162 minimiser of the cost function whatever the realisation of the uncertain variable, there-
 163 fore optimal in all conditions. If the distribution $p_{\mathbf{K}^*}$ is heavily dominated by a single
 164 value, the MPE may be a good candidate for robust control. This is not so obvious in
 165 the case of a multimodal distribution.

166 In general, an analytical form of $p_{\mathbf{K}^*}$ is impossible to obtain, so an estimation $\hat{p}_{\mathbf{K}^*}$
 167 must be used, and its maximum computed to get the MPE. In the rest of the paper,
 168 the hat notation will indicate an estimation using numerical values of the underlying
 169 theoretical quantity. Once again, a set of samples $\{\mathbf{u}^i\}_{i=1\dots N}$ can be used to compute
 170 the set $\{\mathbf{k}_{\mathbf{u}^i}\}_{i=1\dots N}$, from which one can approximate $p_{\mathbf{K}^*}$. The resulting approximation
 171 and therefore its mode, is sensitive to the density estimation method. Main methods are
 172 KDE (Kernel Density Estimation) [33], and EM (Expectation-Maximisation) [34].

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

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

175 where h is the bandwidth. In a multidimensional setting, one usually consider a
 176 kernel based on the product of 1D kernels, applied independently to all components:
 177 $f(\mathbf{k}) = \prod_{j=1}^{\dim \mathbb{K}} f_{1\text{D}}(k^{(j)})$ where $k^{(j)}$ is the j -th component of \mathbf{k} . There is wide choice of
 178 available $f_{1\text{D}}$, and a popular choice is the Gaussian kernel $f_{1\text{D}}(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$.

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

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

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

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

190 2.3. Numerical illustration

191 Before going further in the explanation of our approach, let us illustrate the nature
 192 of previously detailed estimators, $\hat{\mathbf{k}}_{\text{E}}$, $\hat{\mathbf{k}}_{\text{V}}$ and $\hat{\mathbf{k}}_{\text{MPE}}$ on two analytical cost functions.
 193 These functions are based on the Branin-Hoo's function, slightly modified to ensure
 194 strict positivity:

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

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

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

$$J_{\text{BH}} : (\mathbf{k}, \mathbf{u}) \mapsto \text{BH}(\mathbf{k}, \mathbf{u})$$
(15)

$$J_{\text{BHswap}} : (\mathbf{k}, \mathbf{u}) \mapsto \text{BH}(\mathbf{u}, \mathbf{k})$$
(16)

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

198 The random variable \mathbf{U} is assumed to be uniformly distributed over \mathbb{U} . The estima-
 199 tions are based on a 1000×1000 regular grid over $\mathbb{K} \times \mathbb{U}$. Both cost functions are shown
 200 on the top of Figure 1.

201 The left (resp. right) column stands for J_{BH} (resp. J_{BHswap}). Functions $\mu(\mathbf{k})$ and
 202 $\sigma(\mathbf{k})$ are drawn on the bottom row, respectively in purple and green. The corresponding
 203 minimisers $\hat{\mathbf{k}}_{\text{E}}$ and $\hat{\mathbf{k}}_{\text{V}}$ are also plotted. On this figure, we can observe that $\hat{\mathbf{k}}_{\text{E}}$ and $\hat{\mathbf{k}}_{\text{V}}$
 204 are close for J_{BH} , while being significantly different for J_{BHswap} .

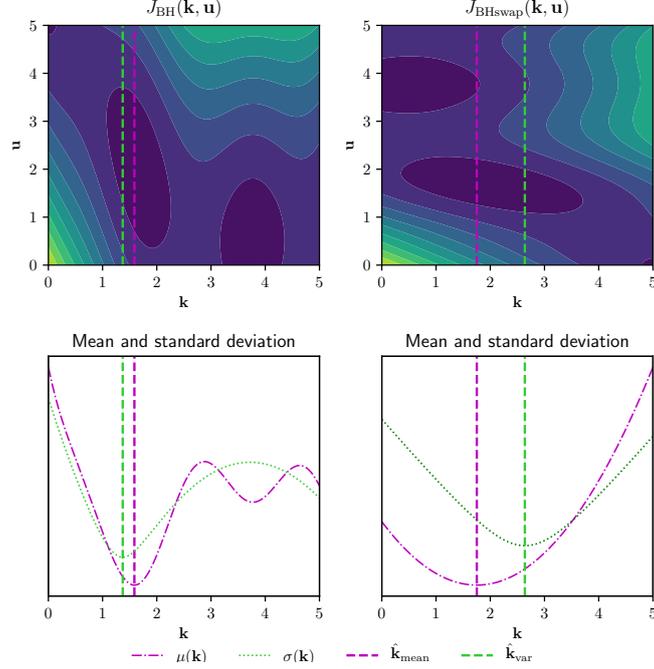


Figure 1: The left column concerns J_{BH} , while the right one deals with J_{BHswap} . Contours of both functions are plotted on the top, and curves of $\mu(\mathbf{k})$ and $\sigma(\mathbf{k})$ are shown on the bottom (respective scales are not displayed). Estimates $\hat{\mathbf{k}}_{\mu}$ and $\hat{\mathbf{k}}_{\sigma}$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

205 Similarly, estimations of $\hat{\mathbf{k}}_{\text{MPE}}$ are depicted on Figure 2. The top row shows the
 206 contours of both functions as well as the set of conditional minimisers $\{\mathbf{k}_{\mathbf{u}^i}\}_{1 \leq i \leq N}$ in red,
 207 as defined in Eq. (9). The bottom row presents three approximations of the density of
 208 minimisers: the histogram in grey (bin size selected using Freedman-Diaconis from [35]),
 209 the result of a kernel density estimation (KDE) with Gaussian kernels in red (using
 210 Scott's rule from [36] for bandwidth selection), and the estimation by a Gaussian mixture,
 211 calculated with the EM algorithm. The number of Gaussians has been fixed to 3, a guess
 212 based on the general shape of the histogram. Respective estimations of $\hat{\mathbf{k}}_{\text{MPE}}$ are also
 213 depicted using dashed lines.

214 For J_{BHswap} , we can observe that those three methods give consistent results, as
 215 $\hat{\mathbf{k}}_{\text{MPE,KDE}} = \hat{\mathbf{k}}_{\text{MPE,EM}} = \hat{\mathbf{k}}_{\text{MPE,histogram}} \approx 0.8$. This is not the case for J_{BH} : using
 216 Kernel density estimation (Gaussian), the estimation of \mathbf{k}_{MPE} is $\hat{\mathbf{k}}_{\text{MPE,KDE}} \approx 1.5$, while
 217 using the histogram and Gaussian mixture, $\hat{\mathbf{k}}_{\text{MPE,histogram}} = \hat{\mathbf{k}}_{\text{MPE,EM}} = 3.8$.

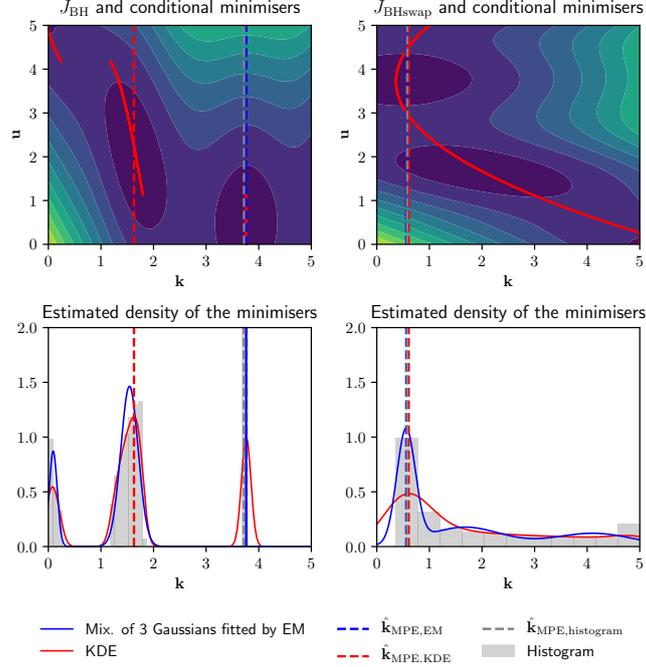


Figure 2: Top Left: J_{BH} along with conditional minimisers in red. Bottom left: Estimated densities using KDE and EM algorithm and the histogram. The dashed lines indicate the MPE found using those methods. Right: Same quantities with J_{BHswap} .

218 This difference is explained by the accumulation of minimisers at this point: this
 219 challenges the assumption that \mathbf{K}^* is continuous. As the density estimation techniques
 220 traditionally assume this continuity, the EM algorithm fits this using a normal distri-
 221 bution with a very small variance, while the KDE considers a sum of Gaussian kernels
 222 of constant bandwidth, located at the same point. This particular problem highlights
 223 an issue with \mathbf{k}_{MPE} , as its estimation is possibly sensitive to the density approximation
 224 procedure.

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

229 **3. Relative regret-based family of estimators**

230 *3.1. Relaxing the optimality constraint*

231 The density of minimisers has been estimated by optimising $\mathbf{k} \mapsto J(\mathbf{k}, \mathbf{U} = \mathbf{u})$ over
 232 \mathbb{K} for different realisations of \mathbf{u} . Instead of focusing on optimal values, we propose to
 233 construct an *acceptable* neighbourhood in terms of performances of the cost function as
 234 well. In order to do so, we are going to introduce a relaxation coefficient $\alpha \geq 1$, and
 235 given $\mathbf{u} \in \mathbb{U}$, we say that $\mathbf{k} \in \mathbb{K}$ is acceptable when $J(\mathbf{k}, \mathbf{U} = \mathbf{u}) \leq \alpha J^*(\mathbf{u})$. Using the
 236 strict positivity of the objective function, we can define the relative-regret as the ratio
 237 $J(\mathbf{k}, \mathbf{U} = \mathbf{u})/J^*(\mathbf{u})$, and the acceptability of \mathbf{k} will then be tied to the maximal value
 238 taken by this ratio.

239 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
 240 \mathbf{k} is acceptable:

$$R_\alpha(\mathbf{k}) = \{\mathbf{u} \in \mathbb{U} \mid J(\mathbf{k}, \mathbf{U} = \mathbf{u}) \leq \alpha J^*(\mathbf{u})\} \quad (17)$$

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

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

$$\Gamma_\alpha(\mathbf{k}) = \mathbb{P}_{\mathbf{U}}[\mathbf{U} \in R_\alpha(\mathbf{k})] = \mathbb{P}_{\mathbf{U}}[J(\mathbf{k}, \mathbf{U}) \leq \alpha J^*(\mathbf{U})] \quad (18)$$

248 In other words, $\Gamma_\alpha(\mathbf{k})$ is the probability that $J(\mathbf{k}, \mathbf{U})$ is between $J^*(\mathbf{U})$ and $\alpha J^*(\mathbf{U})$.

249 Noting that without relaxation, *i.e.* when α is set to 1, Γ_1 is non-zero if the set
 250 $\{\mathbf{u} \in \mathbb{U} \mid J(\mathbf{k}, \mathbf{U} = \mathbf{u}) = J^*(\mathbf{u})\}$ has non-zero measure with respect to $\mathbb{P}_{\mathbf{U}}$. It happens
 251 when the distribution of \mathbf{K}^* presents atoms.

252 This linked to the definition of the distribution of the minimisers \mathbf{K}^* . For instance,
 253 if \mathbb{K} is a discrete set, \mathbf{K}^* is a discrete random variable, and we can rewrite Γ_1 as $\Gamma_1(\mathbf{k}) =$
 254 $\mathbb{P}_{\mathbf{U}}[J(\mathbf{k}, \mathbf{U} = \mathbf{u}) = J^*(\mathbf{U})] = \mathbb{P}_{\mathbf{U}}[\mathbf{k} = \mathbf{k}^*(\mathbf{U})]$ which is the probability mass function of
 255 \mathbf{K}^* .

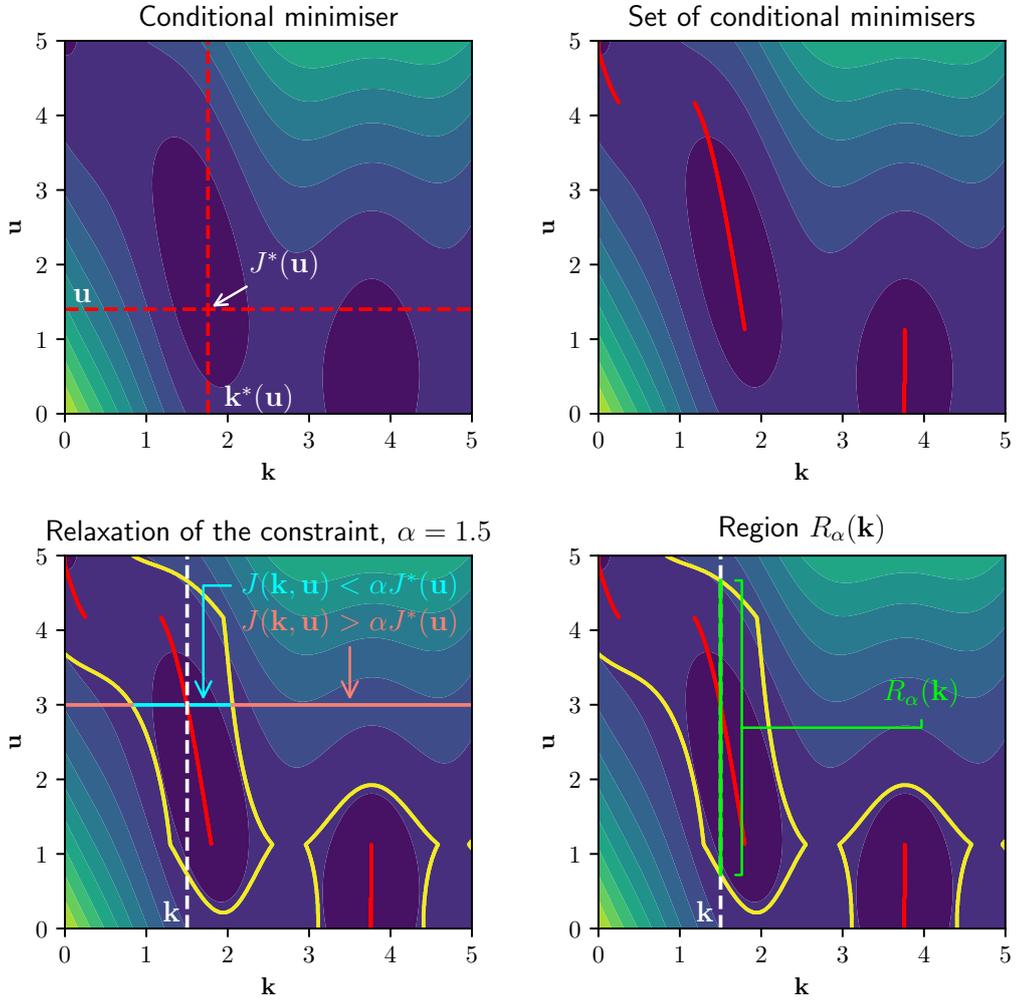


Figure 3: Principle of the relaxation of the constraint on J_{BH} , and illustration of $R_\alpha(\mathbf{k})$. Top plots: Computation of the conditional minimisers $\mathbf{k}^*(\mathbf{u})$. Top right plot: the set $(\mathbf{k}^*(\mathbf{u}), \mathbf{u})$ of conditional minimisers is represented in 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.

256 The motivation behind this relaxation is to take into account the local behaviour
 257 of the function around the conditional minimisers. For a given set of environmental
 258 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

259 choosing $\mathbf{k}^*(\mathbf{u}) + \epsilon$ (for a small ϵ) will produce a value closer to the minimum than when
 260 the function has a high curvature.

261 We consider in this work exclusively a multiplicative relaxation, instead of an additive
 262 one in the form $J(\mathbf{k}, \mathbf{u}) \leq J^*(\mathbf{u}) + \beta$ for $\beta > 0$. Both formulations of the regret: the
 263 additive $J - J^*$ and the relative J/J^* , take into account the value of the conditional
 264 minimum $J^*(\mathbf{u})$, but relative regret allow us to scale the region of acceptable points with
 265 respect to the best situation available: if $J^*(\mathbf{u})$ is close to 0, the region of acceptable
 266 points will grow slowly, so at a given α , we put emphasis on $\mathbf{k}^*(\mathbf{u})$. On the other hand,
 267 if $J^*(\mathbf{u})$ is large, it means that tuning the parameter to $\mathbf{k}^*(\mathbf{u})$ will not lead to very good
 268 performances of the cost function anyway. We can then put less weight on this value
 269 $\mathbf{k}^*(\mathbf{u})$ in the estimation, hence the region of acceptability that grows quickly. Finally, the
 270 relative regret is a normalized quantity, in contrast to the additive regret, which shares
 271 the same unit as J . This normalization, as well as the direct relation to the relative error
 272 $\frac{J-J^*}{J^*}$ allows for an interpretation of the relaxation as a fraction of error.

273 The choice of the relaxation constant α can be made to ensure the existence of a
 274 parameter that is “acceptable” with a certain probability. For instance, given that $J > 0$,
 275 $\Gamma_\alpha(\mathbf{k})$ is increasing with respect to α for any $\mathbf{k} \in \mathbb{K}$. We can then focus on the smallest
 276 value of α such that Γ_α reaches a certain level of confidence $p \in [0, 1]$. This leads to the
 277 definition of α_p

$$\begin{aligned} \alpha_p &= \inf \{ \alpha \geq 1 \mid \exists \mathbf{k}_p \in \mathbb{K}, \Gamma_\alpha(\mathbf{k}_p) \geq p \} \\ &= \inf \left\{ \alpha \geq 1 \mid \max_{\mathbf{k} \in \mathbb{K}} \Gamma_\alpha(\mathbf{k}) \geq p \right\} \end{aligned} \quad (19)$$

278 Rewriting the equation above, we can express α_p as the solution of the following
 279 chance constrained problem

$$\begin{cases} \min & q \\ \text{s.t.} & \max_{\mathbf{k}} \mathbb{P}_{\mathbf{U}} \left[\frac{J(\mathbf{k}, \mathbf{U})}{J^*(\mathbf{U})} \leq q \right] \geq p \end{cases} \quad (20)$$

280 that is the smallest α , such that there exists a particular $\mathbf{k}_p \in \mathbb{K}$ for which $J(\mathbf{k}_p, \mathbf{U}) \leq$
 281 $\alpha_p J^*(\mathbf{U})$ with probability p . As highlighted by the formulation of Eq. (20), \mathbf{k}_p and α_p are
 282 the result of the optimisation of the Value-at-Risk of the random variable $J(\mathbf{k}, \mathbf{U})/J^*(\mathbf{U})$,
 283 which is a measure of risk usually applied in the financial sector (see [37]). In this spirit,
 284 when choosing \mathbf{k}_p , the relative error of the function J will be less than α_p with probability

285 p . The maximal relative regret of the function will be α_p , except for the 100(1 - p)%
 286 least favourable cases.

287 The set of maximisers of Γ_{α_p} for different p : $\{\mathbf{k}_p \text{ for } p \in [0; 1]\}$ is what we are calling
 288 the relative-regret family of estimators RRE.

289 Figure 4 shows examples of Γ_{α_p} for J_{BHswap} , at different levels p , and the associated
 290 estimates of the RRE. We can see that changing the level p shifts the maximiser of Γ_{α_p} .
 291 Since \mathbf{k}_1 is located quite far from the conditional minimisers, it arises as a compromise
 292 when the relaxation is large enough.

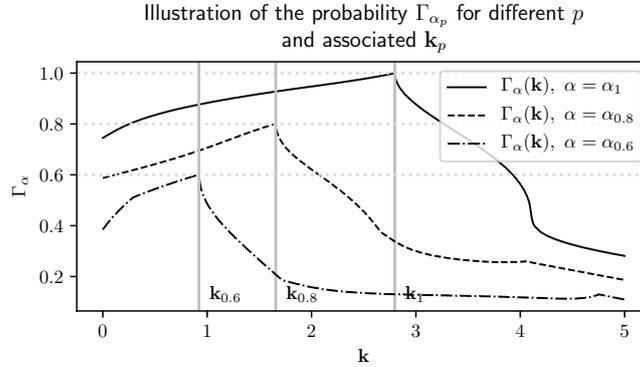


Figure 4: Illustration of the influence of different levels p on Γ_{α_p} and on \mathbf{k}_p for J_{BHswap} . When p increases, Γ_{α_p} increases as well.

293 We have then introduced the RRE, a family of estimators among which we can choose
 294 either large level of confidence p *i.e.* a large robustness as we are controlling the relative
 295 regret for a large fraction of possible configurations, or we can look for almost optimal
 296 performances albeit for a more reduced number of situations. Those quantities require
 297 the evaluation and optimisation of probabilities, so solving such a problem can be quite
 298 challenging. Moreover, the choice of a member of the RRE, that is the choice of a level
 299 of confidence p is also up to the modeller.

300 3.2. Choice and computation of the relaxation coefficient

301 3.2.1. Almost-surely bounded relative-regret

302 Let us first consider that we want to satisfy the chance constraint of Eq. (20) almost
 303 surely: this is the particular case of $p = 1$. Given the strict positivity of J , the choice

304 of such a level of confidence is possible only if there exists a \mathbf{k} such that the ratio
 305 $J(\mathbf{k}, \mathbf{U})/J^*(\mathbf{U})$ is bounded almost surely, and α_1 is the solution of the (almost-surely)
 306 chance constrained problem:

$$\begin{cases} \min & q \\ \text{s.t.} & \min_{\mathbf{k}} \frac{J(\mathbf{k}, \mathbf{U})}{J^*(\mathbf{U})} \leq q \quad \text{a.s.} \end{cases} \quad (21)$$

307

308 thus, \mathbf{k}_1 verifies

$$\mathbb{P}_{\mathbf{U}} \left[\frac{J(\mathbf{k}_1, \mathbf{U})}{J^*(\mathbf{U})} \leq \alpha_1 \right] = 1 \iff \frac{J(\mathbf{k}_1, \mathbf{U})}{J^*(\mathbf{U})} \leq \alpha_1 \quad \text{a.s.} \quad (22)$$

309 Under the assumption that $J(\mathbf{k}, \cdot)$ is continuous for all \mathbf{k} , this is equivalent to ap-
 310 proaching this problem using uncertain sets to model the uncertain nature of \mathbf{U} :

$$\frac{J(\mathbf{k}_1, \mathbf{u})}{J^*(\mathbf{u})} \leq \alpha_1 \quad \forall \mathbf{u} \in \mathbb{U} \quad (23)$$

311 it follows then that

$$\alpha_1 = \max_{\mathbf{u} \in \mathbb{U}} \frac{J(\mathbf{k}_1, \mathbf{U} = \mathbf{u})}{J^*(\mathbf{u})} = \min_{\mathbf{k} \in \mathbb{K}} \left\{ \max_{\mathbf{u} \in \mathbb{U}} \frac{J(\mathbf{k}, \mathbf{U} = \mathbf{u})}{J^*(\mathbf{u})} \right\} \quad (24)$$

312 This can then be linked to Savage's minimax approach [38], which consists in a worst-
 313 case approach for the additive regret $J - J^*$, when in Eq. (24) we look to minimise the
 314 worst-case scenario in terms of the ratio.

315 Using Sample Average Approximation (SAA), based on N i.i.d. samples of \mathbf{U} : $\{\mathbf{u}^i\}_{1 \leq i \leq N}$,
 316 we can reformulate the problem Eq. (21) as

$$\begin{cases} \min & q \\ \text{s.t.} & \min_{\mathbf{k}} \frac{J(\mathbf{k}, \mathbf{u}^i)}{J^*(\mathbf{u}^i)} \leq q \quad \text{for } 1 \leq i \leq N \end{cases} \quad (25)$$

317 For a solution $\hat{\alpha}_1$ of Eq. (25), as SAA acts as a relaxation of the initial problem
 318 Eq. (21), the solution found acts as lower bound on the true value α_1 : $\hat{\alpha}_1 \leq \alpha_1$.

319 Moreover, this estimated value $\hat{\alpha}_1$ can be used for the estimation of the relaxation
 320 constant for a level p close to 1. Using Clopper-Pearson intervals [39, 40], if $\hat{\alpha}_1$ is a
 321 solution of Eq. (25), it is also a feasible solution of Eq. (20) at a level $p = \left(\frac{\eta}{2}\right)^{1/N}$ with
 322 probability $1 - \eta$: $\hat{\alpha}_1$ is then a probabilistic upper bound on $\alpha \left(\frac{\eta}{2}\right)^{\frac{1}{N}}$.

$$\hat{\alpha}_1 \in \left[\alpha \left(\frac{\eta}{2}\right)^{1/N}, \alpha_1 \right] \quad \text{with probability } 1 - \eta \quad (26)$$

323 The estimated value $\hat{\alpha}_1$ will then be optimistic in the sense that it will underestimate
 324 the true value α_1 . We can however control this approximation using the probabilistic
 325 bound lower bound of Eq. (26)

326 Choosing a level of confidence $p = 1$ suffers from the same pitfall as the worst-case
 327 approach, as it may return over-conservative solutions, provided that a solution exists in
 328 the first place.

329 3.2.2. Estimation of RRE

330 We are now going to focus on the more general case, where we can choose: to fix
 331 $p \leq 1$ and deduce α_p , or to set a maximal threshold α and maximise the measure of the
 332 acceptable region, or finally to find a compromise p and α_p in order to keep the latter
 333 not too large.

334 A first approach is to set α , and then to estimate the function $\mathbf{k} \mapsto \Gamma_\alpha(\mathbf{k})$ as defined
 335 in Eq. (17). For instance, when N samples from \mathbf{U} are available, namely $\{\mathbf{u}^i\}_{1 \leq i \leq N}$, a
 336 possible estimate of Γ_α is

$$\hat{\Gamma}_\alpha(\mathbf{k}) = \frac{\text{Card}\{\mathbf{u}^i \mid J(\mathbf{k}, \mathbf{U} = \mathbf{u}^i) \leq \alpha J^*(\mathbf{u}^i)\}}{N} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{J(\mathbf{k}, \mathbf{U}=\mathbf{u}^i) \leq \alpha J^*(\mathbf{u}^i)\}} \quad (27)$$

337 This expression is maximised with respect to \mathbf{k} , giving $\hat{p} = \max_{\mathbf{k}} \hat{\Gamma}_\alpha(\mathbf{k})$. We can also
 338 provide a confidence interval for the true value of $\Gamma_\alpha(\mathbf{k})$ where $\mathbf{k} = \arg \max \hat{\Gamma}_\alpha$.

339 However, unless the modeller has some precise idea of desired relaxation, doing so
 340 may lead to an unsatisfactory pairing of α and p . Indeed if α is chosen too small, the
 341 resulting $\hat{p} = \max_{\mathbf{k}} \hat{\Gamma}_\alpha(\mathbf{k})$ will be also small, meaning that the cost function will have
 342 non acceptable values with high probability.

343 Similarly, if p is fixed, the corresponding $\hat{\alpha}_p$ is computed by searching for the smallest
 344 α satisfying $\max_{\mathbf{k}} \hat{\Gamma}_\alpha(\mathbf{k}) = p$, or equivalently, by minimizing the quantile of order p of the
 345 relative-regret, which is the ratio J/J^* . Once again, if $\hat{\alpha}_p$ is too large, the cost function
 346 may not be controlled enough for the contemplated application.

347 Looking for a compromise between p and α would be preferable. This could be
 348 achieved by studying $p \mapsto \alpha_p$, and particularly its slope. If this curve presents a steep
 349 increase, the multiplicative constant α_p must be increased by a large amount in order to
 350 increase the probability p by a small amount. Interesting couples (p, α_p) would then be
 351 the ones located before an abrupt increase of the slope of $p \mapsto \alpha_p$.

352 Another possibility is to model this compromise by the ratio (p/α_p) , as it increases
 353 with respect to p and decreases with respect to α_p . The level of confidence p_{ratio} is then
 354 defined as the maximiser of $p \mapsto p/\alpha_p$.

355 3.3. Numerical Illustration

356 In this Section, we will compare the different estimators introduced previously and
 357 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
$\arg \min_{\mathbf{k} \in \mathbb{K}} \text{Var}_{\mathbf{U}} [J(\mathbf{k}, \mathbf{U})]$	$\mathbf{k}_{\mathbb{V}}$	Steady performances
$\arg \max_{\mathbf{k} \in \mathbb{K}} p_{\mathbf{K}^*}(\mathbf{k})$	\mathbf{k}_{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

358 As stated before, we chose to model the uncertainties as a random variable uniformly
 359 distributed on \mathbb{U} . The bounded nature of \mathbb{U} allows us to consider members of the RRE
 360 up to a level of confidence $p = 1$. From now on, $\hat{\mathbf{k}}_{\text{MPE}}$ is estimated using KDE with
 361 Gaussian kernels.

362 The smallest estimated relaxation $\hat{\alpha}_1$ and the corresponding $\hat{\mathbf{k}}_1$ has been computed for
 363 J_{BH} and J_{BHswap} , using a regular grid of 1000×1000 points on $\mathbb{K} \times \mathbb{U}$. The contour plots
 364 of those functions can be seen in the top plots of Figure 5. The frontier corresponding
 365 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
 366 top of these contour plots, for $\alpha = \hat{\alpha}_1$ and an arbitrary $\alpha = 1.5 < \hat{\alpha}_1$ to illustrate the
 367 effect of the acceptable region when the relaxation α changes. On the bottom plots, the
 368 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
 369 are represented.

370 One can notice that the relaxation allows us to avoid the issue brought by the accu-
 371 mulation of the minimisers of J_{BH} at 3.8, as opposed to the MPE and its dependence on
 372 the estimation procedure of the distribution.

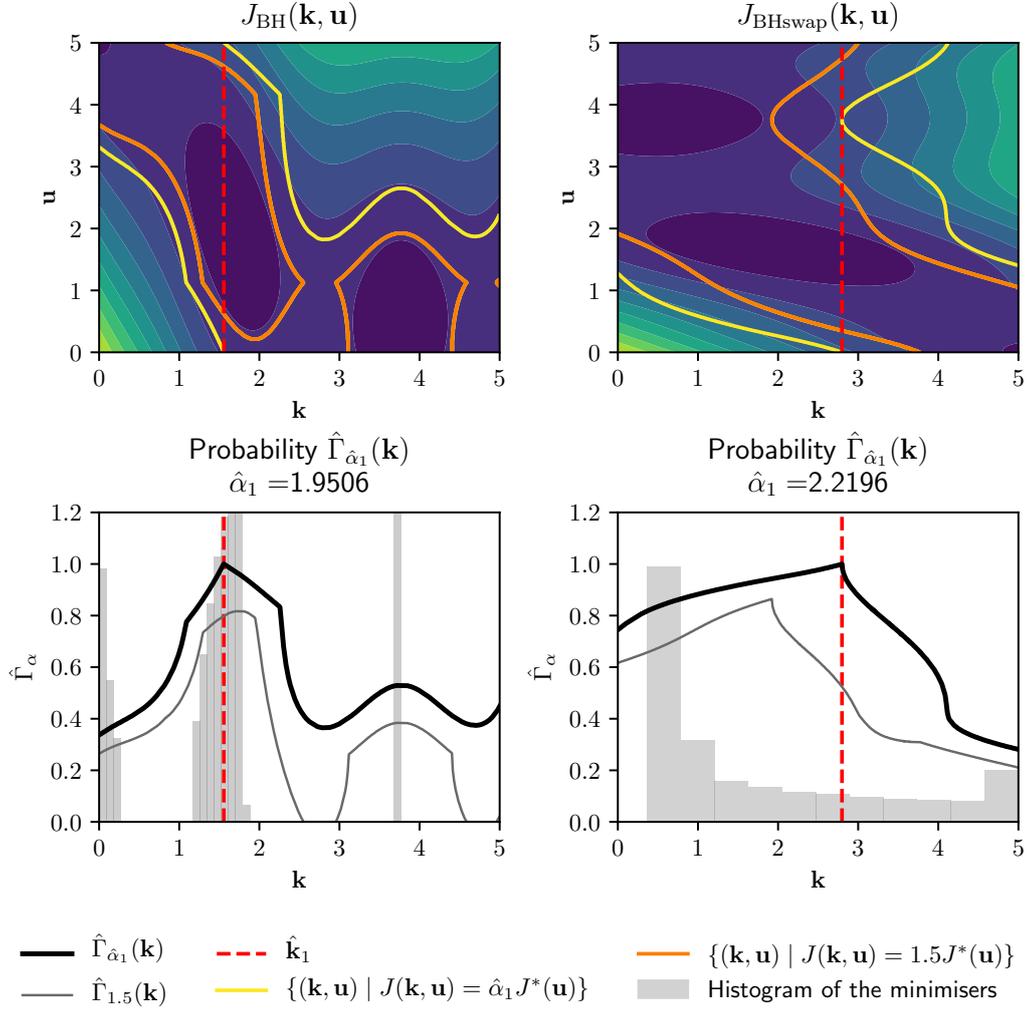


Figure 5: Top: J_{BH} and J_{BHswap} contours. The thick yellow lines are the boundaries of the acceptable region defined for $\hat{\alpha}_1$, the thick orange are for $\alpha = 1.5$. The red dashed line is the estimation $\hat{\mathbf{k}}_1$. Bottom: $\hat{\Gamma}_{\alpha}$ for $\alpha = 1.5$ and $\alpha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

373 In order to choose a satisfying level of confidence p , we are going to study $p \mapsto \hat{\alpha}_p$
 374 and $p \mapsto p/\hat{\alpha}_p$, as described in Section 3.2.

375 The plot of $p \mapsto \hat{\alpha}_p$ for J_{BH} on Figure 6 shows what seems to be a piecewise linear
 376 behaviour. The last change of slope, *i.e.* for $p \approx 0.9$ corresponds to a local maximum
 377 of the ratio, while the first change of slope at $\hat{p}_{\text{ratio}} = 0.654$ corresponds to the global

378 maximum of the ratio. The RRE will then be evaluated for both of these values, as well
 379 as $p = 1$ for reference.

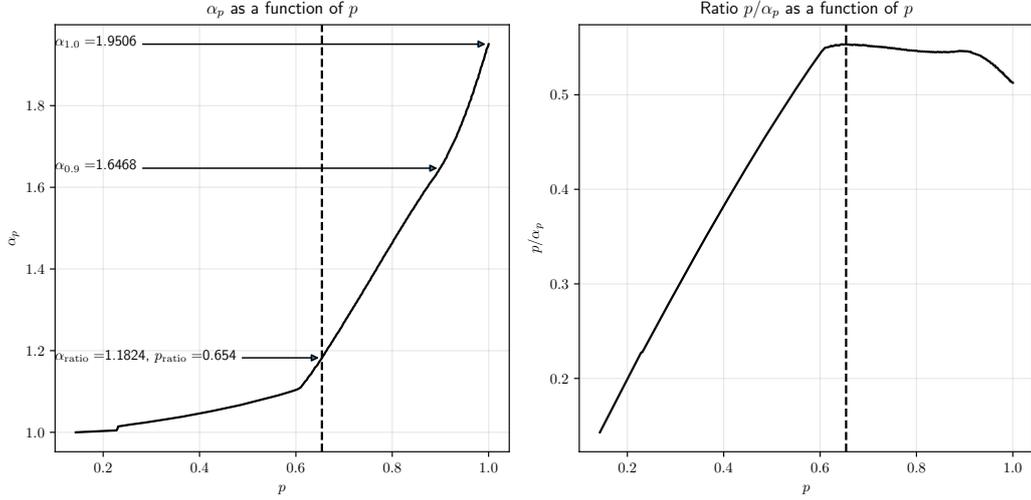


Figure 6: Evolution of the couples (p, α_p) and corresponding ratio p/α_p for J_{BH} . The dashed line indicates the level p associated with the highest ratio

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

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

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

382 Practically speaking, in order to compare the effective values taken by the objective
 383 function given an estimate $\hat{\mathbf{k}}$ we are going to consider the functions $\mathbf{u} \mapsto J(\hat{\mathbf{k}}, \mathbf{U} = \mathbf{u})$,
 384 that we will call “profiles of $\hat{\mathbf{k}}$ ”. Those profiles are well suited for the representation of

385 the cost function for an estimate $\hat{\mathbf{k}}$ fixed as the uncertain variable is modelled with a 1D
 386 uniform random variable.

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

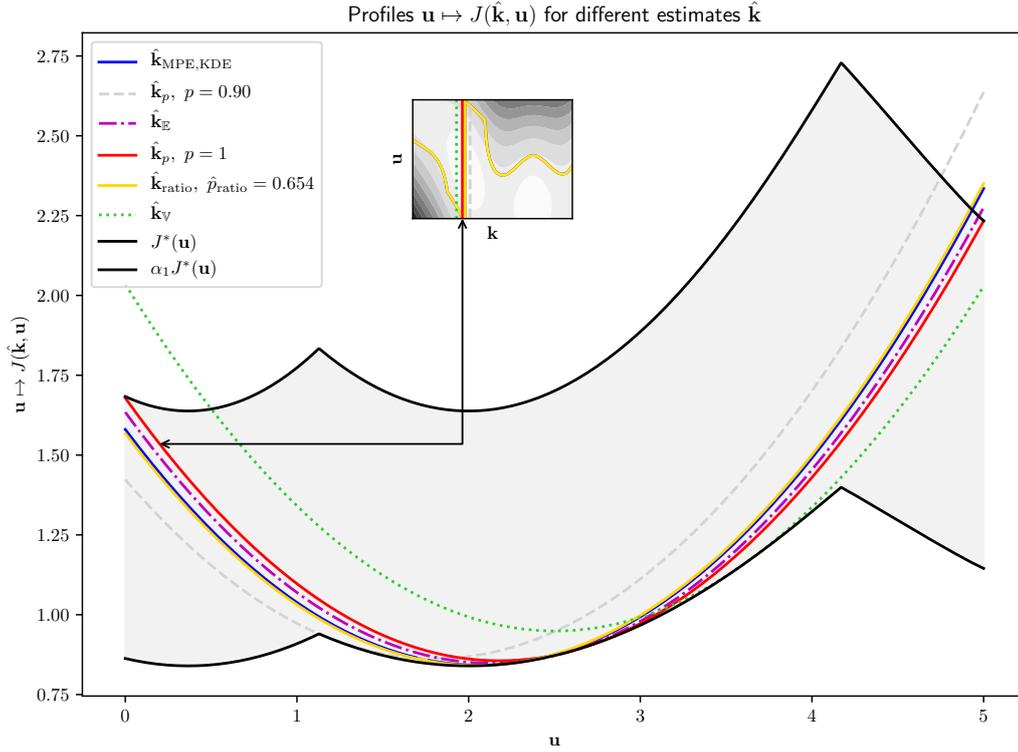


Figure 7: Profiles of the different estimates for J_{BH} , corresponding to the vertical cross sections of the contour. The shaded region corresponds to the interval $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$. The profiles of $\hat{\mathbf{k}}_{\text{MPE}}$ and $\hat{\mathbf{k}}_{\text{ratio}}$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

388 By construction, the profile of $\hat{\mathbf{k}}_1$ is always within the shaded region, corresponding
 389 to $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$. The profile of $\hat{\mathbf{k}}_{\text{E}}$ in contrast, exceeds $\hat{\alpha}_1 J^*(\mathbf{u})$ for \mathbf{u} close to 5, while
 390 the profile of $\hat{\mathbf{k}}_{\text{V}}$ does it for \mathbf{u} close to 0. Except for $\hat{\mathbf{k}}_{\text{V}}$, the different estimators give
 391 somewhat comparable results.

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

394 Compared to the similar plots for J_{BH} in Figure 6, $\hat{\alpha}_p$ exhibits a smoother behaviour

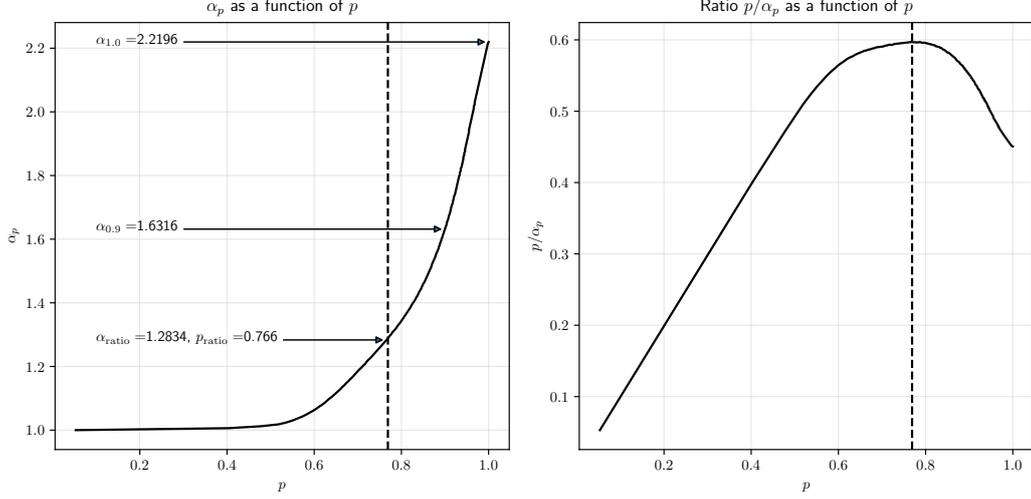


Figure 8: Evolution of the couples (p, α_p) and corresponding ratio p/α_p for J_{BHswap} . The dashed line indicates the level p associated with the highest ratio

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

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

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

398 Profiles of the different estimates of J_{BHswap} are shown in Figure 9.
 399 In this case, $\hat{\mathbf{k}}_{\text{MPE}}$, $\hat{\mathbf{k}}_{\text{E}}$ and $\hat{\mathbf{k}}_{\text{ratio}}$ present a similar behaviour. They perform very
 400 well for $\mathbf{u} > 2$, especially for $\hat{\mathbf{k}}_{\text{MPE}}$ which is very close to the minimal value; however
 401 for $\mathbf{u} < 2$, they produce high values of the function. The performances of $\hat{\mathbf{k}}_1$ are closer
 402 to the performances of $\hat{\mathbf{k}}_{\text{V}}$ for this function, but it performs worse than $\hat{\mathbf{k}}_{\text{E}}$ and $\hat{\mathbf{k}}_{\text{V}}$ for

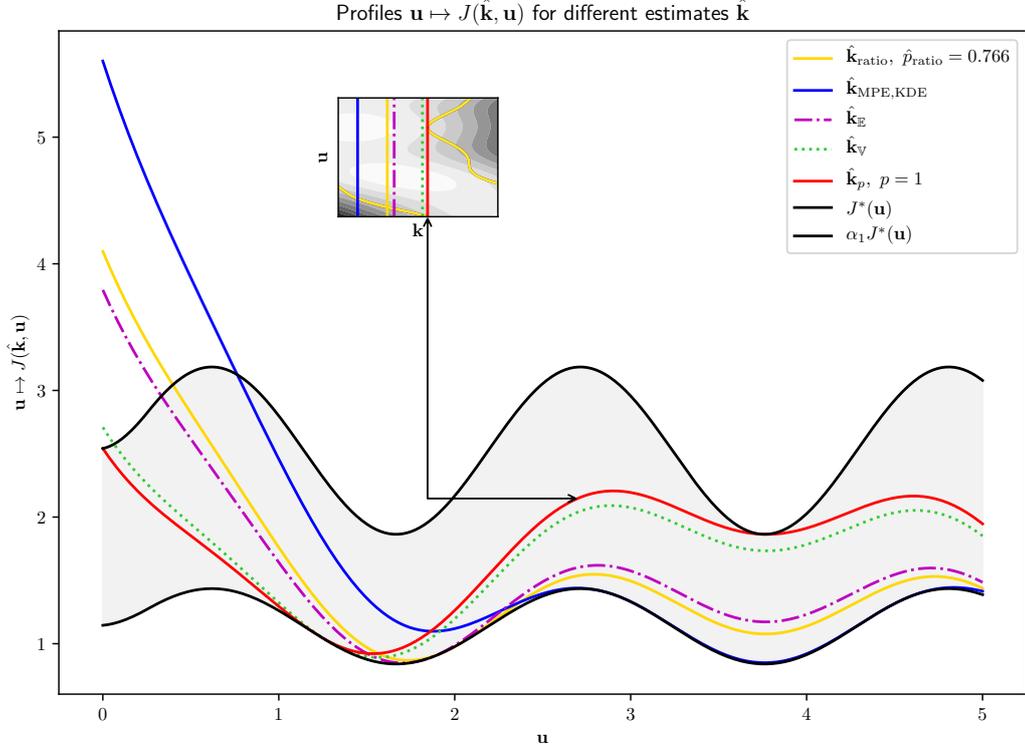


Figure 9: Profiles of the different estimates for J_{BHswap} . Those profiles are the vertical cross sections of the contours above. The shaded region corresponds to the interval $[J^*(\mathbf{u}), \hat{\alpha}_1 J^*(\mathbf{u})]$

403 $\mathbf{u} > 2$, even though its range is designed to stay within the interval $[J^*(\mathbf{u}); \hat{\alpha}_1 J^*(\mathbf{u})]$.

404 We have seen how some classical robust estimators and the RRE behave on two
 405 different analytical problems. In addition to the usual levels of confidence such as 90%
 406 and 95%, one can also settle for an ad-hoc compromise, where p maximises the ratio
 407 p/α_p . When the sample space of \mathbf{U} is bounded, a conservative solution is to set $p = 1$.
 408 We are now going to see how can robust minimisation is applied on the calibration of a
 409 numerical model.

410 4. Robust calibration of a numerical model

411 4.1. Calibration of a toy numerical model

412 We will follow the approach described in [41] in order to establish the function \mathcal{G}
413 described in the first section in Eq. (1), and the resulting cost function J .

414 The calibration of a numerical model is usually based on the comparison between
415 the numerical model and some observations, during a fixed time interval $[0, T]$ called
416 assimilation window. The modelled physical system can be seen as a map from \mathbb{U} to
417 \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
418 representing some environmental conditions. The observation mentioned above is the
419 output of the physical system during the time-window, and is denoted by $\mathcal{M}^o(\mathbf{u}^{\text{true}}) \in \mathbb{Y}$,
420 where $\mathbf{u}^{\text{true}} \in \mathbb{U}$ is unknown.

421 In addition of \mathbf{u} , the numerical model \mathcal{M} depends on some other input $\mathbf{k} \in \mathbb{K}$. This
422 additional parametrization comes usually from the successive simplifications needed to
423 implement a numerical model of the observed physical system. \mathbf{k} needs to be calibrated
424 accordingly, so that the numerical model can be used to predict the behaviour of the
425 physical system under different operating conditions.

426 The misfit \mathcal{G} is defined as the difference between the numerical model and the obser-
427 vation. Choosing a squared norm, the cost function J defined in Eq. (1) is

$$J(\mathbf{k}, \mathbf{U} = \mathbf{u}) = \frac{1}{2} \|\mathcal{G}(\mathbf{k}, \mathbf{u})\|^2 = \frac{1}{2} \|\mathcal{M}(\mathbf{k}, \mathbf{u}) - \mathcal{M}^o(\mathbf{u}^{\text{true}})\|^2 \quad (28)$$

428 4.2. The Shallow Water equations

429 The model to calibrate is an implementation of the 1D Shallow Water equations,
430 described in Eq. (29), where h is the height of the water column, q the discharge, and
431 z the bathymetry, while g is the usual gravitation constant. The parameter to calibrate
432 \mathbf{k} is the quadratic friction term, proportional to the square of the inverse of Manning-
433 Strickler coefficient. The environmental parameter \mathbf{u} is the amplitude of a sine wave of
434 period $1/\omega_0$. The domain of those two parameters are $\mathbb{K} = [0.0, 1.3]$ and $\mathbb{U} = [0.5, 0.7]$.

$$\left\{ \begin{array}{l} \partial_t h + \partial_x q = 0 \\ \partial_t q + \partial_x \left(\frac{q^2}{h} + \frac{g}{2} h^2 \right) = -gh \partial_x z - \mathbf{k}q|q|h^{-7/3} \\ h(0, t) = 20.0 + 3 \cdot \sin\left(\frac{2\pi t}{2}\right) + 1.5 \cdot \mathbf{u} \cdot \sin\left(\frac{2\pi t}{\omega_0}\right) \\ \partial_x q(0, t) = 0 \end{array} \right. \quad (29)$$

435 These equations are integrated using a finite-volume scheme on a discretized domain
 436 $[0, L]$, up to a time T . The output of the computer code is the sea surface height h , on the
 437 center of all the volumes and at all the time-steps, that will be denoted $\mathcal{M}(\mathbf{k}, \mathbf{u}; \omega_0 = 1.0)$.
 438 In this setting, the random variable \mathbf{U} is uniformly distributed on \mathbb{U} .

439 To generate the observation, we set $\mathbf{u}^{\text{true}} = 2/3$, and define \mathcal{M}^o based on the computer
 440 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
 441 uncontrollable error between the observations and the numerical model and will now be
 442 omitted systematically in the notation.

443 The true value of the bottom friction $\mathbf{k}^{\text{true}} = (k_1^{\text{true}}, k_2^{\text{true}}, \dots, k_{N_{\text{vol}}}^{\text{true}})$ is not constant
 444 over the whole domain, and is defined as

$$k_i^{\text{true}} = 0.2 \cdot \left(1 + \sin\left(\frac{2\pi x_i}{L}\right) \right) \quad (30)$$

445 where x_i is the center of the i -th volume. The two sources of systematic errors
 446 are the one-dimensionality of \mathbb{K} , and ω_0 . Given this setting, there exists no couple
 447 $(\mathbf{k}, \mathbf{u}) \in \mathbb{K} \times \mathbb{U}$ reproducing exactly the observations, thus the cost function will always
 448 be strictly positive.

449 4.3. Computation of the robust estimates

450 As the numerical model is expensive to run, it has first been evaluated on a relatively
 451 small regular grid covering $\mathbb{K} \times \mathbb{U}$, and a metamodel based on Gaussian processes is
 452 constructed using those initial evaluations. In order to better capture the locus of the
 453 conditional minimisers $\{(\mathbf{k}^*(\mathbf{u}), \mathbf{u}) \mid \mathbf{u} \in \mathbb{U}\}$, we select points maximising the PEI crite-
 454 rion [42, 43] in order to add points to the design that improve the metamodel accuracy
 455 close to the conditional minimisers. Afterwards, a bigger regular grid is evaluated by the
 456 metamodel once the design space has been sufficiently explored, and these computations
 457 are used to calibrate the model.

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

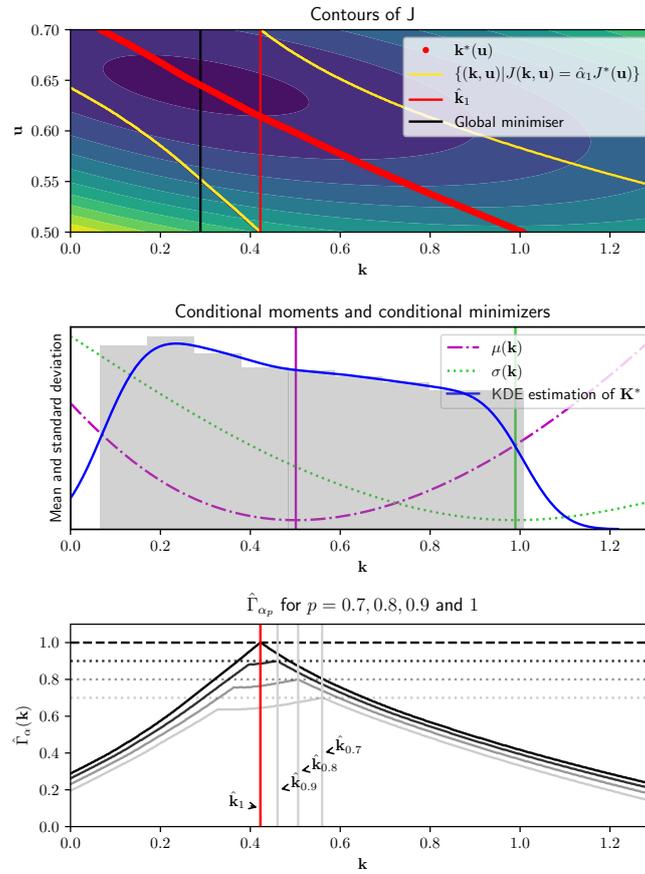


Figure 10: Procedure of robust calibration for the shallow water problem. Top: contours of J , conditional minimisers and $\{(\mathbf{k}, \mathbf{U} = \mathbf{u}) | J(\mathbf{k}, \mathbf{u}) = \hat{\alpha}_1 J^*(\mathbf{u})\}$. Middle: Conditional moments and histogram and KDE of the conditional minimisers.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

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

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

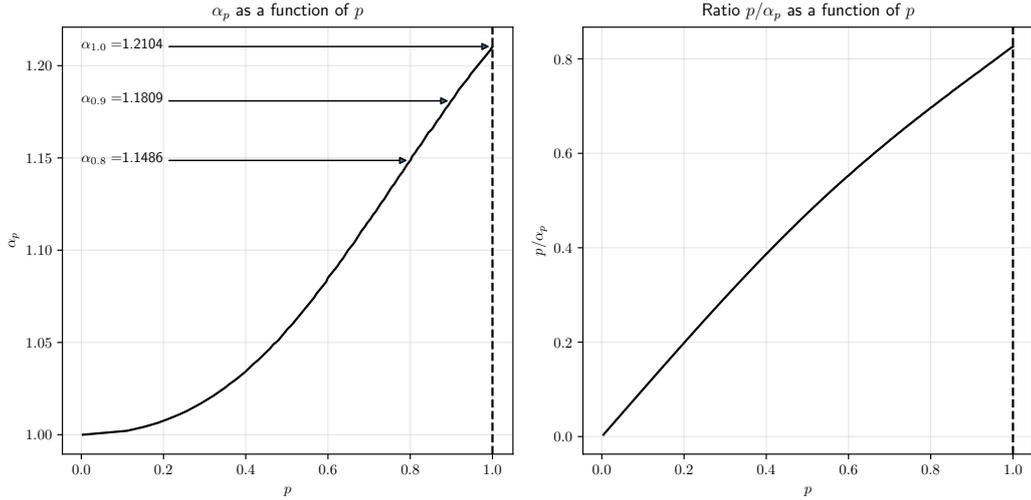


Figure 11: Evolution of α_p for different levels p , and ratio, for the shallow water problem

469 The different estimates take a wide range of values, as seen on Table 4, from 0.249
 470 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
 471 may lead to values outside the range given by the true ones. This can be interpreted
 472 as the fact that the calibration look to compensate for errors on ω_0 . As a basis for
 473 comparison, the global minimiser of J over $\mathbb{K} \times \mathbb{U}$ has also been computed, and $\hat{\mathbf{k}}_{\text{global}}$
 474 is then obtained by discarding the \mathbf{u} value.

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

476 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
 477 performances when $\mathbf{u} > 0.6$. When comparing with the global optimiser $\hat{\mathbf{k}}_{\text{global}}$, $\hat{\mathbf{k}}_1$
 478 performs better when $\mathbf{u} < 0.625$, so more than half of the time. We are now going to
 479 compare how well some of those calibrated values compare in a forecast context.

480 4.4. Assessing the quality of the forecast of the calibrated model

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

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

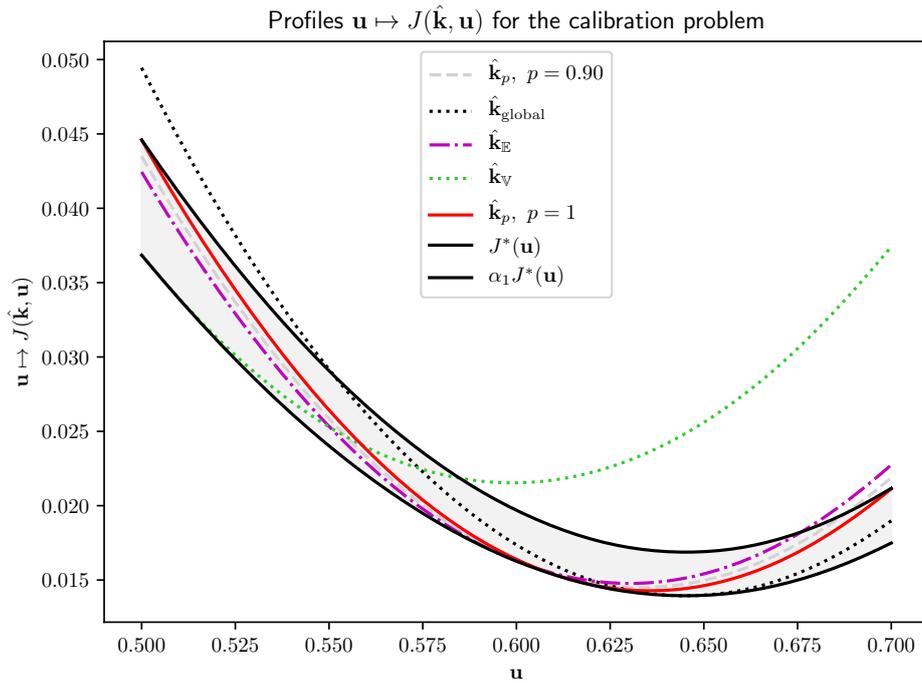


Figure 12: Profiles for the calibration problem, for the shallow water case

483 height on the same time-period. We now want to compare the quality of the different
 484 forecasts originating from different calibrated bottom frictions. Those forecasts result

485 from the integration of the numerical model between the time T and a time T_{pred} : the
 486 forecast window.

487 Given the probabilistic nature of the environmental conditions \mathbf{U} , the forecasts will
 488 also be probabilistic. We will then compare $\mathcal{M}_{\text{pred}}^o(\mathbf{U})$ and $\mathcal{M}_{\text{pred}}(\mathbf{k}, \mathbf{U})$, that are the
 489 observation, and the numerical model on the forecast window for a calibrated \mathbf{k} .

490 Two metrics will be computed: the squared forecast error, and the Continuous Ranked
 491 Probability Score (CRPS) [44]. The parameters satisfying a robust criterion that are
 492 compared are the minimiser of the expectation $\hat{\mathbf{k}}_{\mathbb{E}}$, the minimiser of the variance $\hat{\mathbf{k}}_{\mathbb{V}}$ and
 493 the relative regret of level $p = 1$, $\hat{\mathbf{k}}_1$ as described in the previous section. We will also
 494 feature the global minimiser $\hat{\mathbf{k}}_{\text{global}}$, and the conditional minimisers $\hat{\mathbf{k}}^*(\mathbf{u}^i)$ where the
 495 chosen environmental variables are $\{\mathbf{u}^i\}_{1 \leq i \leq 4} = \{0.5, 0.55, 0.65, 0.7\}$. Those values, even
 496 though they do not meet a robustness criterion, are introduced in order to have a more
 497 precise idea on the possible performances of a deterministic version of the calibration
 498 problem. The conditional minimiser $\hat{\mathbf{k}}^*(0.6)$ has been omitted as it results in a value
 499 very similar to the minimum of the expectation.

500 4.4.1. Squared forecast error

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

$$S_{\text{pred}}(\mathbf{k}, \mathbf{u}, \mathbf{u}') = (\mathcal{M}_{\text{pred}}(\mathbf{k}, \mathbf{u}) - \mathcal{M}_{\text{pred}}^o(\mathbf{u}'))^2 \quad (31)$$

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

$$S(\mathbf{k}) = \frac{1}{N_{\mathbf{u}}^2} \sum_{i=1}^{N_{\mathbf{u}}} \sum_{j=1}^{N_{\mathbf{u}}} S_{\text{pred}}(\mathbf{k}, \mathbf{u}^i, \mathbf{u}^j) = \frac{1}{N_{\mathbf{u}}^2} \sum_{i=1}^{N_{\mathbf{u}}} \sum_{j=1}^{N_{\mathbf{u}}} (\mathcal{M}_{\text{pred}}(\mathbf{k}, \mathbf{u}^i) - \mathcal{M}_{\text{pred}}^o(\mathbf{u}^j))^2 \quad (32)$$

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

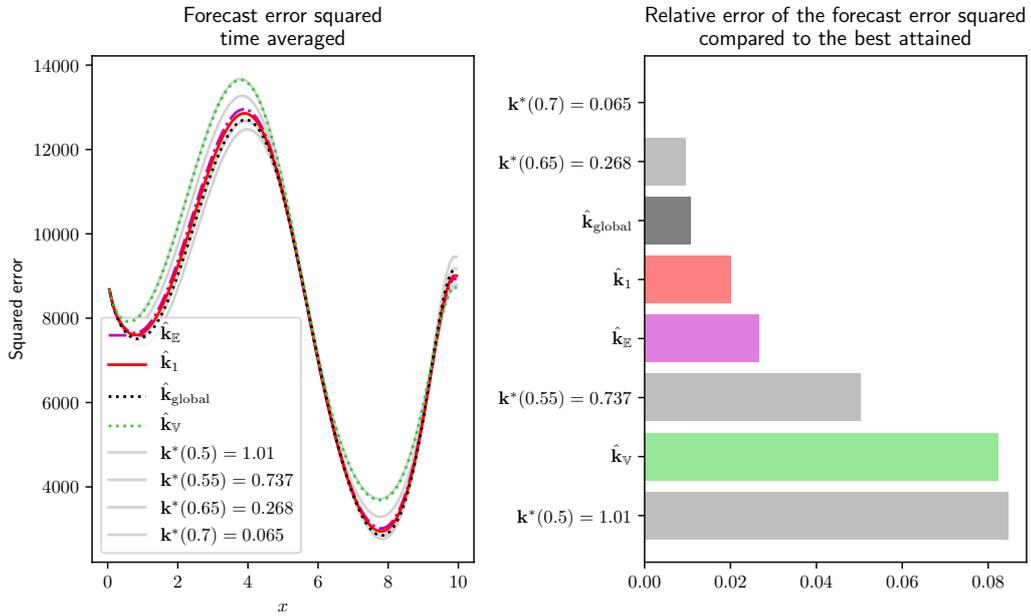


Figure 13: Squared forecast error for the shallow water case, depending on the calibrated parameter. The left figure shows the squared error time-averaged as a function of the spatial coordinate x , while the right barplot shows the relative change of the mean forecast error, averaged over time and space,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

516 We can see that the mean forecast error squared, on the right side, averaged over
517 time and space is the smallest for the conditional minimisers $\hat{\mathbf{k}}^*(0.7)$ and $\hat{\mathbf{k}}^*(0.65)$, then
518 $\hat{\mathbf{k}}_{\text{global}}$, while $\hat{\mathbf{k}}_{1.0}$ performs slightly better than $\hat{\mathbf{k}}_{\mathbb{E}}$.

519 It may seem surprising that some parameters calibrated without a robustness aspect
520 perform better than $\hat{\mathbf{k}}_1$ and $\hat{\mathbf{k}}_{\mathbb{E}}$, but their performances are largely dependent on the
521 choice of \mathbf{u} and their associated conditional minimisers. Good forecasts can be achieved

522 as well as bad ones, as shown for $\mathbf{u} = 0.5$, that leads to bad forecasts.

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

527 4.4.2. Continuous Ranked Probability Score

528 Given the random variables $\mathcal{M}_{\text{pred}}(\hat{\mathbf{k}}, \mathbf{U})$ and $\mathcal{M}_{\text{pred}}^o(\mathbf{U})$, representing the probabilis-
 529 tic forecast and the probabilistic observations, we can define the cumulative distribu-
 530 tion functions (CDF) $F_{\text{pred}}(\cdot, \hat{\mathbf{k}})$ and $F_{\text{pred}}^o(\cdot)$. The Continuous ranked probability score
 531 (CRPS) measures the squared difference between the predicted CDF F_{pred} using a cali-
 532 brated value, and the CDF of the observations F_{pred}^o .

$$\text{CRPS}(\mathbf{k}) = \int_{\mathbb{R}} (F_{\text{pred}}(\xi, \mathbf{k}) - F_{\text{pred}}^o(\xi))^2 d\xi \quad (33)$$

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

545 The numerical evaluations of the CRPS for different parameters show the same order
 546 of performances observed for the squared error: the calibrated parameters that present
 547 the best performances for forecasts according to those two metrics are $\hat{\mathbf{k}}^*(0.7)$, $\hat{\mathbf{k}}^*(0.65)$
 548 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}
 549 to choose for the conditional minimisation, thus having a strong insight on the value of
 550 the parameters in the first place.

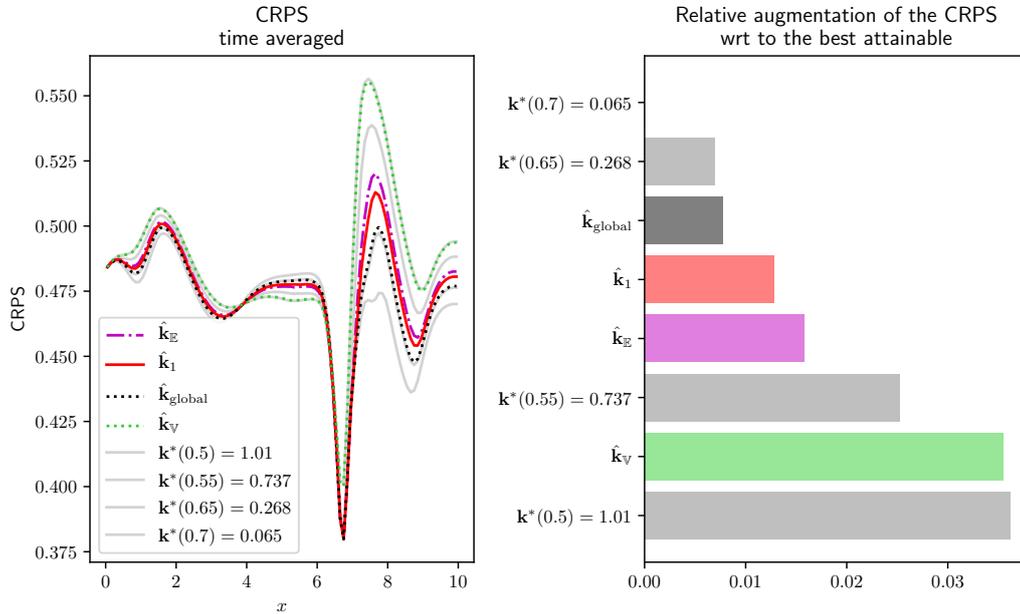


Figure 14: CRPS computed for different calibrated parameters for the shallow water case. The left figure shows the CRPS time-averaged as a function of the spatial coordinate x , and the right figure shows the relative change of the CRPS averaged over time and space, taken with respect to the best attained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

551 Conclusion

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

557 Our approach is based on the comparison of the objective function to the optimal
 558 value it can attain, given a realisation of the uncertain variable. By constraining this
 559 ratio in terms of probability, the modeller can then minimize the frequency at which the
 560 relative error of the objective function will exceed a prescribed threshold. Alternatively,
 561 at a level of confidence p , the associated threshold α_p bounds the relative deviation of
 562 the objective function for the proportion p of the most favourable cases. The resulting

563 estimate can then be assessed robust, as it corresponds to values of the objective function
564 as close as possible to the optimal values.

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

571 For the calibration of the shallow water model presented in Section 4, we chose to
572 construct a response surface based on Gaussian processes, which is then used for the
573 extensive computations. This construction is first based on an initial design, that is
574 enriched to better capture the locus of the conditional minimisers using the PEI criterion
575 introduced in [42].

576 Practically speaking, the estimation of those quantities is very crude, so the computa-
577 tional cost may be very expensive. As perspective, specific strategies have to be created
578 in order to carefully select each additional point that will be evaluated, and thus reduce
579 the amount of evaluations needed of the numerical model. This selection can be based
580 for instance on sequential design of computer experiments using Gaussian processes [45].

581 Finally, the models upon which this calibration procedure have been applied are very
582 simplistic. We plan in the future to apply this to the robust calibration of the bottom
583 friction of a realistic model of the ocean.

584 References

- 585 [1] W. E. Walker, P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. van Asselt, P. Janssen,
586 M. P. Kreyer von Krauss, Defining uncertainty: A conceptual basis for uncertainty management in
587 model-based decision support, *Integrated assessment* 4 (2003) 5–17.
- 588 [2] S. K. Das, R. W. Lardner, On the estimation of parameters of hydraulic models by assimilation of
589 periodic tidal data, *Journal of Geophysical Research* 96 (1991) 15187.
- 590 [3] S. K. Das, R. W. Lardner, Variational parameter estimation for a two-dimensional numerical tidal
591 model, *International Journal for Numerical Methods in Fluids* 15 (1992) 313–327.
- 592 [4] M. Boutet, Estimation Du Frottement Sur Le Fond Pour La Modélisation de La Marée Barotrope,
593 Ph.D. thesis, Université d’Aix Marseille, 2015.

- 594 [5] P. G. J. ten Brummelhuis, A. W. Heemink, Parameter identification in tidal models with uncertain
595 boundary conditions, *Stochastic Hydrology and Hydraulics* 4 (1990) 193–208.
- 596 [6] L. Huyse, D. M. Bushnell, Free-form airfoil shape optimization under uncertainty using maximum
597 expected value and second-order second-moment strategies (2001).
- 598 [7] G. Kuczera, B. Renard, M. Thyer, D. Kavetski, There are no hydrological monsters, just models
599 and observations with large uncertainties!, *Hydrological Sciences Journal* 55 (2010) 980–991.
- 600 [8] T. E. Wong, A. Klufas, V. Srikrishnan, K. Keller, Neglecting model structural uncertainty under-
601 estimates upper tails of flood hazard, *Environmental Research Letters* 13 (2018) 074019.
- 602 [9] Y. Hu, S. Rao, Robust Design of Horizontal Axis Wind Turbines Using Taguchi Method, *Journal*
603 *of Mechanical Design* 133 (2011) 111009.
- 604 [10] L. Brevault, Contributions to Multidisciplinary Design Optimization under Uncertainty, Application
605 to Launch Vehicle Design., Theses, Ecole Nationale Supérieure des Mines de Saint-Étienne, 2015.
- 606 [11] N. Lelièvre, P. Beaurepaire, C. Mattrand, N. Gayton, A. Otsmane, On the consideration of uncer-
607 tainty in design: Optimization-reliability-robustness, *Structural and Multidisciplinary Optimization*
608 54 (2016) 1423–1437.
- 609 [12] G. Petrone, G. Iaccarino, D. Quagliarella, Robustness criteria in optimization under uncertainty,
610 Evolutionary and deterministic methods for design, optimization and control (EUROGEN 2011).
611 CIRA, Capua (2011) 244–252.
- 612 [13] P. Seshadri, P. Constantine, G. Iaccarino, G. Parks, A density-matching approach for optimization
613 under uncertainty, arXiv:1409.7089 [math, stat] (2014).
- 614 [14] L. W. Cook, J. P. Jarrett, Horsetail matching: A flexible approach to optimization under uncer-
615 tainty, *Engineering Optimization* 50 (2018) 549–567.
- 616 [15] C. Ning, F. You, Optimization under uncertainty in the era of big data and deep learning: When
617 machine learning meets mathematical programming, *Computers & Chemical Engineering* 125 (2019)
618 434–448.
- 619 [16] P. J. Huber, Robust statistics, in: *International Encyclopedia of Statistical Science*, Springer, 2011,
620 pp. 1248–1251.
- 621 [17] V. Rao, A. Sandu, M. Ng, E. Nino-Ruiz, Robust data assimilation using L_1 and Huber norms,
622 *SciRate* (2015).
- 623 [18] J. O. Berger, E. Moreno, L. R. Pericchi, M. J. Bayarri, J. M. Bernardo, J. A. Cano, J. De la Horra,
624 J. Martín, D. Ríos-Insúa, B. Betrò, An overview of robust Bayesian analysis, *Test* 3 (1994) 5–124.
- 625 [19] H. Rahimian, S. Mehrotra, Distributionally Robust Optimization: A Review, arXiv:1908.05659 [cs,
626 math, stat] (2019).
- 627 [20] J. Marzat, E. Walter, H. Piet-Lahanier, Worst-case global optimization of black-box functions
628 through Kriging and relaxation, *Journal of Global Optimization* 55 (2013) 707–727.
- 629 [21] J. Villemonteix, E. Vazquez, E. Walter, An informational approach to the global optimization of
630 expensive-to-evaluate functions, arXiv:cs/0611143 (2006).
- 631 [22] P. Hennig, C. J. Schuler, Entropy Search for Information-Efficient Global Optimization (2011).
- 632 [23] J. M. Buhmann, M. Mihalak, R. Sramek, P. Widmayer, Robust optimization in the presence of

- 633 uncertainty, in: Proceedings of the 4th Conference on Innovations in Theoretical Computer Science
634 - ITCS '13, ACM Press, Berkeley, California, USA, 2013, p. 505. doi:10.1145/2422436.2422491.
- 635 [24] P. Kouvelis, A. A. Kurawarwala, G. J. Gutiérrez, Algorithms for robust single and multiple period
636 layout planning for manufacturing systems, *European Journal of Operational Research* 63 (1992)
637 287–303.
- 638 [25] L. Snyder, M. Daskin, Stochastic p-robust location problems, *Iie Transactions* 38 (2004).
- 639 [26] A. Juditsky, A. S. Nemirovski, G. Lan, A. Shapiro, Stochastic Approximation Approach to Stochastic
640 Programming, in: ISMP 2009 - 20th International Symposium of Mathematical Programming,
641 2009.
- 642 [27] S. Kim, R. Pasupathy, S. Henderson, A Guide to Sample Average Approximation, volume 216,
643 2015, pp. 207–243. doi:10.1007/978-1-4939-1384-8-8.
- 644 [28] J. Janusevskis, R. Le Riche, Simultaneous Kriging-Based Sampling for Optimization and Uncer-
645 tainty Propagation, Technical Report, 2010.
- 646 [29] V. Baudoui, Optimisation Robuste Multiobjectifs Par Modèles de Substitution, Ph.D. thesis,
647 Toulouse, ISAE, 2012.
- 648 [30] O. Grodzevich, O. Romanko, Normalization and other topics in multi-objective optimization (2006).
- 649 [31] R. T. Marler, J. S. Arora, The weighted sum method for multi-objective optimization: New insights,
650 *Structural and Multidisciplinary Optimization* 41 (2010) 853–862.
- 651 [32] J. S. Lehman, T. J. Santner, W. I. Notz, Designing computer experiments to determine robust
652 control variables, *Statistica Sinica* (2004) 571–590.
- 653 [33] B. Silverman, *Density Estimation: For Statistics and Data Analysis*, 2018. doi:10.1201/
654 9781315140919.
- 655 [34] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM
656 algorithm, *Journal of the royal statistical society. Series B (methodological)* (1977) 1–38.
- 657 [35] D. Freedman, P. Diaconis, On the histogram as a density estimator:L2 theory, *Zeitschrift für*
658 *Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57 (1981) 453–476.
- 659 [36] D. W. Scott, On Optimal and Data-Based Histograms, *Biometrika* 66 (1979) 605.
- 660 [37] R. T. Rockafellar, S. P. Uryasev, M. Zabrankin, Deviation Measures in Risk Analysis and Op-
661 timization, SSRN Scholarly Paper ID 365640, Social Science Research Network, Rochester, NY,
662 2002.
- 663 [38] L. J. Savage, The Theory of Statistical Decision, *Journal of the American Statistical Association*
664 46 (1951) 55–67.
- 665 [39] J. Luedtke, S. Ahmed, A sample approximation approach for optimization with probabilistic con-
666 straints, *SIAM Journal on Optimization* 19 (2008) 674–699.
- 667 [40] M. Thulin, The cost of using exact confidence intervals for a binomial proportion, *Electronic*
668 *Journal of Statistics* 8 (2014) 817–840.
- 669 [41] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, *Journal of the Royal*
670 *Statistical Society: Series B (Statistical Methodology)* 63 (2001) 425–464.
- 671 [42] D. Ginsbourger, J. Baccou, C. Chevalier, F. Perales, N. Garland, Y. Monerie, Bayesian Adaptive

- 672 Reconstruction of Profile Optima and Optimizers, *SIAM/ASA Journal on Uncertainty Quantifica-*
673 *tion* 2 (2014) 490–510.
- 674 [43] J. Bossek, B. Bischl, T. Wagner, G. Rudolph, Learning Feature-Parameter Mappings for Parameter
675 Tuning via the Profile Expected Improvement, in: *Proceedings of the 2015 Annual Conference on*
676 *Genetic and Evolutionary Computation, GECCO '15*, ACM, New York, NY, USA, 2015, pp. 1319–
677 1326. doi:10.1145/2739480.2754673.
- 678 [44] T. Gneiting, M. Katzfuss, Probabilistic Forecasting, *Annual Review of Statistics and Its Application*
679 *1* (2014) 125–151.
- 680 [45] D. Ginsbourger, R. Le Riche, Towards Gaussian Process-based Optimization with Finite Time
681 Horizon, in: A. Giovagnoli, A. C. Atkinson, B. Torsney, C. May (Eds.), *mODa 9 – Advances in*
682 *Model-Oriented Design and Analysis, Contributions to Statistics*, Physica-Verlag HD, Heidelberg,
683 2010, pp. 89–96. doi:10.1007/978-3-7908-2410-0_12.