

Titre – 05/12/2018

Victor Trappler, Élise Arnaud, Laurent Debrou, and Arthur Vidard

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

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

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

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

This will then be applied on a academic problem of robust estimation of bottom friction for a numerical model of the one dimensional shallow water equations.

Copyright statement. TEXT

10 ToDo

	P.
1. Notations \check{k}/\hat{k} ?	4
2. Notation MPE ?	5
3. Parler plutôt des quantiles de $\frac{J(\mathbf{k}, \mathbf{U})}{J(\mathbf{k}^*(\mathbf{U}), \mathbf{U})}$?	5

15 1 Introduction

1.1 Motivation

The process of modelling a physical phenomenon usually implies to add some errors. These uncertainties come from different mandatory steps, such as the discretisation of the domain, or simplifications to take into account subscale phenomena. These unresolved processes are usually represented by corrective terms involving parametrisation. Once those parameters have been
20 identified, they have to be estimated, so that the numerical model matches as closely as possible the reality. This is done by the minimisation of a cost function.

The uncertainties in the model are twofold: the “true” values of the parameters are yet to be evaluated, while some other parameters are fixed but possibly to a wrong value. A rough distinction between those types of uncertainty can be found in Walker et al. (2003):

- Epistemic uncertainty is the uncertainty coming from a lack of knowledge, and can be reduced. This is the ignorance of the true value of parameter that we aim at calibrating.
- Aleatoric uncertainty is the uncertainty coming from the inherent variability of the system studied. This is the “natural” variability of some environmental variables.

In this paper, the estimation will consist in reducing the epistemic uncertainty, while taking into account the aleatoric uncertainty. This is what we define as robust estimation.

10 The definition of robustness differs depending of the context in which it is used. Indeed, a definition of the robustness of an estimator is a measure of the sensibility of said estimator to outliers (Huber, 2011). This lead to the introduction of robust norms in data assimilation, such as the Huber norm in Rao et al. (2015). The point of this norm is to assume Gaussian distribution for small errors (the usual L_2 norm), while the large errors are assumed to be Laplace distributed (so the L_1 norm). In other cases, robustness in a Bayesian framework refers to the sensitivity to a wrong specification of the priors (Berger et al., 1994).

15 In our application, the cost function to “minimise” is a random variable. We are going to focus on two different notions of robustness: robustness by minimising the mean and variance, as it has been applied to portfolio selection in Markowitz (1952), and robustness based on the distribution of the minimisers of the function. This information on the location of the minimisers is not a novel idea, as it had been applied as a criterion in sequential Bayesian optimisation (Villemonteix et al., 2006; Hennig and Schuler, 2011). More specifically, the constraint of optimality will be relaxed to define a new estimate.

20 After having properly defined those robust estimates, an academic problem of estimation of bottom friction will be introduced, and the robust estimates will be computed in Section 5.

1.2 Context and notations

We assume that we have access to a model \mathcal{G} , a computer code simulating a physical phenomenon over a specific time window. Different types of parameters for this model can be established (Walker et al., 2003):

- 25
- A priori chosen parameters, which are chosen, and considered invariant
 - Calibrated parameters, which are unknown and needs to be calibrated.

Based on this classification, we define $\mathbf{k} \in \mathcal{K} \subset \mathbb{R}^p$ as the parameter to calibrate, i.e. the control variable, while $\mathbf{u} \in \mathcal{U} \subset \mathbb{R}^l$ is a priori chosen, and set at \mathbf{u}^b . This parameter represents the environmental conditions of the phenomenon studied. The computer code maps those parameters to an observable physical quantity (such as the sea surface height) that will be compared to an observation $\mathbf{y}^o \in \mathcal{Y}$:

$$\mathcal{G} : (\mathbf{k}, \mathbf{u}) \in \mathcal{K} \times \mathcal{U} \longmapsto \mathcal{G}(\mathbf{k}, \mathbf{u}) \in \mathcal{Y} \quad (1)$$

2 Calibration: deterministic twin experiments

For the estimation of the “best” value of \mathbf{k} , we compute first an observation \mathbf{y}^o , generated with the couple $(\mathbf{k}^t, \mathbf{u}^t)$, supposed unknown (where $(\cdot)^t$ stands for “truth”):

$$\mathbf{y}^o = \mathcal{G}(\mathbf{k}^t, \mathbf{u}^t) \quad (2)$$

- 5 In a data assimilation setting, this calibration consists in the minimisation of the cost function J defined by

$$J(\mathbf{k}) = \frac{1}{2} \|\mathcal{G}(\mathbf{k}, \mathbf{u}^b) - \mathbf{y}^o\|_{\mathbf{R}^{-1}}^2 + \text{regul}(\mathbf{k}) \quad (3)$$

where regul is a regularisation function taking possibly into account a background value \mathbf{k}^b . The matrix \mathbf{R} is the covariance matrix of the observation error. Assuming a well-posed problem, the function J reaches a unique global minimum at $\hat{\mathbf{k}}$:

$$\hat{\mathbf{k}} = \arg \min_{\mathbf{k} \in \mathcal{K}} J(\mathbf{k}) \quad (4)$$

- 10 This method brings different problems: minimising J will for instance compensate the error introduced by the parametrisation \mathbf{k} , but will also compensate another misspecification: choosing \mathbf{u}^b instead of \mathbf{u}^t in the computer code. In the end, the value $\hat{\mathbf{k}}$ obtained will be optimal for this configuration. However, if the operating conditions do change, there is no information on the performance of $\hat{\mathbf{k}}$ for this other setting. As a result, this may lead to poor results when the calibrated model is used for predictions. This naïve approach is treated in Section 5.2.
- 15 Another strategy, minimising J over $\mathcal{K} \times \mathcal{U}$ is not always possible, and not always relevant. The minimisation over the whole joint space may increase the complexity of the problem, and the parameter computed may not be robust in the end.

The random nature of the environmental variables is treated by introducing a probabilistic framework.

3 Stochastic modelling of the cost function

- The aleatoric uncertainty on the environmental variables is modelled by a random variable \mathbf{U} whose sample space is \mathcal{U} . The probability measure of \mathbf{U} is \mathbb{P}_U , and its density, if it exists, is p_U . Furthermore, \mathbf{u}^t (used to generate \mathbf{y}^{obs}) is sampled from \mathbf{U} . The cost function defined in Eq. (3) is no longer a function of solely \mathbf{k} , but of \mathbf{U} as well:

$$J(\mathbf{k}, \mathbf{U}) = \frac{1}{2} \|\mathcal{G}(\mathbf{k}, \mathbf{U}) - \mathbf{y}^o\|_{\mathbf{R}^{-1}}^2 + \text{regul}(\mathbf{k}) \quad (5)$$

For a given \mathbf{k} , $J(\mathbf{k}, \mathbf{U})$ is a random variable. The main point is to be able to compute a value \mathbf{k} that “minimises” in a certain sense the random variable $J(\mathbf{k}, \mathbf{U})$.

- 25 For a practical purpose, we assume that $\forall \mathbf{u} \in \mathcal{U}$, the minimisation of $\mathbf{k} \mapsto J(\mathbf{k}, \mathbf{u})$ is a well-posed problem, and that the minimiser is unique. Additionally, the following assumption are made: the cost function is smooth and strictly positive, and $\forall \mathbf{k} \in \mathcal{K}$, the random variable $J(\mathbf{k}, \mathbf{U})$ has finite first- and second-order moments.

4 Definitions of robustness

1.

In this paper, we define robustness as the ability of \mathbf{k} to perform reasonably well under different operating conditions (i.e. different \mathbf{u}). In that sense, we introduce two different aspects:

- 5 – Optimisation of the moments of the random variable: usually minimisation of the expected value, or of the variance
– Study of the distribution of the minimisers, assuming the inverse problem is well-posed to ensure the existence and the unicity of the minimiser for each \mathbf{u} . The function of condiditional minimisers is then defined by:

$$\mathbf{k}^* : \mathbf{u} \in \mathcal{U} \longmapsto \mathbf{k}^*(\mathbf{u}) = \arg \min_{\mathbf{k} \in \mathcal{K}} J(\mathbf{k}, \mathbf{u}) \in \mathcal{K} \quad (6)$$

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

10 $\mathbf{K}^* = \mathbf{k}^*(\mathbf{U}) \quad (7)$

of generalized probability density function

$$p_{K^*}(\mathbf{k}) \propto \mathbb{P}_U [\mathbf{k} = \mathbf{k}^*(\mathbf{U})] = R_1(\mathbf{k}) \quad (8)$$

4.1 Optimisation of the moments

Optimising the moments is usually a sound approach, when one aims at optimising in a robust manner. Let us first define $\mu(\mathbf{k})$,
15 and $\sigma^2(\mathbf{k})$, the expected value and the variance of the cost function for a given \mathbf{k} as

$$\mu(\mathbf{k}) = \mathbb{E}_U [J(\mathbf{k}, \mathbf{U})] = \int_{\mathcal{U}} J(\mathbf{k}, \mathbf{u}) d\mathbb{P}_U(\mathbf{u}) \quad (9)$$

$$\sigma^2(\mathbf{k}) = \text{Var}_U [J(\mathbf{k}, \mathbf{U})] = \int_{\mathcal{U}} (J(\mathbf{k}, \mathbf{u}) - \mu(\mathbf{k}))^2 d\mathbb{P}_U(\mathbf{u}) \quad (10)$$

Mean response optimisation leads to the estimate $\hat{\mathbf{k}}_{\text{mean}}$ defined by:

$$\hat{\mathbf{k}}_{\text{mean}} = \arg \min_{\mathbf{k} \in \mathcal{K}} \mu(\mathbf{k}) \quad (11)$$

20 Minimising the variance, leads to $\hat{\mathbf{k}}_{\text{var}}$:

$$\hat{\mathbf{k}}_{\text{var}} = \arg \min_{\mathbf{k} \in \mathcal{K}} \sigma^2(\mathbf{k}) \quad (12)$$

From a user point of view, using $\hat{\mathbf{k}}_{\text{mean}}$ will give good results in the long run, so when a lot of different \mathbf{u} are considered. On the other hand, using $\hat{\mathbf{k}}_{\text{var}}$ will ensure that the output of the cost function will be as constant as possible when \mathbf{u} varies, but without any information on its performance.

For a single \mathbf{u} sampled, Chebyshev's inequality gives an upper bound on the probability of being more than $\lambda > 0$ units from the mean.

$$\mathbb{P}_U [|J(\mathbf{k}, \mathbf{U}) - \mu(\mathbf{k})| > \lambda] \leq \frac{\sigma^2(\mathbf{k})}{\lambda^2} \quad (13)$$

This inequality ties the mean and variance, so it is usual to consider mean and variance simultaneously. This multiobjective optimisation problem is usually tackled by computing the Pareto front of $(\mu(\mathbf{k}), \sigma^2(\mathbf{k}))$ (Baudouin, 2012). As the multiobjective problem is sometimes hard to solve, alternative strategies based on the minimisation of a scalarized version of the vector of objectives are often considered. Those strategies involve for instance a weighted sum of the objectives (Grodzevich and Romanko, 2006), or the minimisation of one of the objectives under constraint on the other, named ϵ -constrained method (applied for instance in Lehman et al., 2004).

10 4.2 Distribution of the minimisers

4.2.1 Most probable estimate

2. We are now interested in the distribution of the minimisers of the cost function:

$$p_{K^*}(\mathbf{k}) \propto R_1(\mathbf{k}) = \mathbb{P}_U \left[J(\mathbf{k}, \mathbf{U}) = \min_{\tilde{\mathbf{k}}} J(\tilde{\mathbf{k}}, \mathbf{U}) \right] = \mathbb{P}_U [\mathbf{k} = \mathbf{k}^*(\mathbf{U})] \quad (14)$$

From this definition, $R_1(\mathbf{k})$ is the probability that $\mathbf{k} \in \mathcal{K}$ is optimal. To have a first insight on the interest of applying robust estimation, we can compute the information brought by this distribution by means of Shannon's entropy defined by

$$H[\mathbf{K}^*] = - \int_{\mathcal{K}} \log \left(\frac{d\mathbb{P}_{K^*}}{d\lambda} \right) d\mathbb{P}_{K^*} \quad (15)$$

where \mathbb{P}_{K^*} is the image measure of the random variable \mathbf{K}^* , and λ is a base measure on which we compare \mathbb{P}_{K^*} . The distribution is the most informative if \mathbf{K}^* is equal to a constant \mathbf{k} almost surely (with entropy tending to $-\infty$). On the other hand, the distribution is the least informative if \mathbf{K}^* takes values uniformly spread over \mathcal{K} .

20 From the definition of R_1 in Eq. (14), we can define the mode of the distribution of the minimisers (if it exists) as $\hat{\mathbf{k}}_{MPE}$ (where MPE stands for Most Probable Estimate):

$$\hat{\mathbf{k}}_{MPE} = \arg \max_{k \in \mathcal{K}} R_1(k) \quad (16)$$

However, if the value of this maximum is not high, it may not be considered robust: this formulation does not provide information on the performance of the estimator for the configuration where it is not optimal (situation that happens with probability 25 $1 - R_1(\hat{\mathbf{k}}_{MPE})$).

4.2.2 Relaxation of the constraint

3. For the MPE, the distribution of the conditional minimisers has been estimated. Instead of only considering the optimal values, we are going to consider optimal values and their “neighbourhood” as well.

For that purpose, $\alpha \geq 1$ is introduced to relax the equality constraint of Eq. (14). In that sense, for a given \mathbf{u} , \mathbf{k} is deemed acceptable if $J(\mathbf{k}, \mathbf{u}) \leq \alpha J(\mathbf{k}^*(\mathbf{u}), \mathbf{u})$. Taking into account the random nature of \mathbf{U} , $R_\alpha(\mathbf{k})$ is defined as the probability that \mathbf{k} is acceptable:

$$R_\alpha(\mathbf{k}) = \mathbb{P}_U [J(\mathbf{k}, \mathbf{U}) \leq \alpha J(\mathbf{k}^*(\mathbf{U}), \mathbf{U})] \quad (17)$$

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

The motivation behind this relaxation is to take into account the local behaviour of the function around the conditional minimisers: For a given \mathbf{u} , if the function $\mathbf{k} \mapsto J(\mathbf{k}, \mathbf{u})$ is flat around its minimum $\mathbf{k}^*(\mathbf{u})$, then choosing $\mathbf{k}^*(\mathbf{u}) + \epsilon$ (for a small ϵ) will produce a value closer to the minimum than when the function has a high curvature. In addition to that, relaxing the constraint using a multiplicative constant puts more weight (i.e. a slower increase of the acceptable region) on the values of
10 $\mathbf{k}^*(\mathbf{u})$ when $J(\mathbf{k}^*(\mathbf{u}), \mathbf{u})$ is close to zero.

Under the assumption that $J > 0$, $R_\alpha(\mathbf{k})$ is increasing with respect to α for any $\mathbf{k} \in \mathcal{K}$. From Eq. (17), we can obtain an upper bound for relevant values of α , provided that the maximum and minimum of J is attained on $\mathcal{K} \times \mathcal{U}$:

$$\alpha \geq \frac{\max_{(\mathbf{k}, \mathbf{u})} J(\mathbf{k}, \mathbf{u})}{\min_{(\mathbf{k}, \mathbf{u})} J(\mathbf{k}, \mathbf{u})} \iff R_\alpha(\mathbf{k}) = 1 \quad \forall \mathbf{k} \in \mathcal{K} \quad (18)$$

Let us set a level of confidence $0 < p \leq 1$. We are now looking for $\check{\alpha}_p$, the smallest α , such that there exists a particular
15 $\check{\mathbf{k}}_p \in \mathcal{K}$ for which $J(\check{\mathbf{k}}_p, \mathbf{U}) \leq \check{\alpha}_p J(\mathbf{k}^*(\mathbf{U}), \mathbf{U})$ with probability p . As R_α is increasing with respect to α , $\check{\alpha}_p$ can then be defined as

$$\check{\alpha}_p = \inf_{\alpha \geq 1} \{ \alpha \mid \exists \check{\mathbf{k}}_p \in \mathcal{K}, R_\alpha(\check{\mathbf{k}}_p) \geq p \} \quad (19)$$

In other words, this definition can be seen as the construction of the smallest confidence interval of level p :

$$\mathbb{P}_U \left[\frac{J(\check{\mathbf{k}}_p, \mathbf{U})}{J(\mathbf{k}^*(\mathbf{U}), \mathbf{U})} \in [1, \alpha_p] \right] = p \quad (20)$$

20 By considering the particular case of $\check{\alpha}_1$ a relation between $\check{\mathbf{k}}_1$ and $\check{\alpha}_1$ can be derived:

$$\check{\alpha}_1 = \sup_{\mathbf{u} \in \mathcal{U}} \left\{ \frac{J(\check{\mathbf{k}}_1, \mathbf{u})}{J(\mathbf{k}^*(\mathbf{u}), \mathbf{u})} \right\} \quad (21)$$

A high value of $\check{\alpha}_1$ corresponds to a high variability of the response of the function for at least one particular \mathbf{u} . In that sense, $J(\check{\mathbf{k}}_1, \mathbf{u})$ may be possibly very high. For this reason, it is important to study jointly $\check{\alpha}_p$ and $\check{\mathbf{k}}_p$, as the value of $\check{\alpha}_p$ is an information on the “robustness” of the solution as it is linked to the range of $\mathbf{u} \mapsto J(\check{\mathbf{k}}_p, \mathbf{u})$. The different estimators introduced
25 in this paper are summarised in Table 1.

4.3 Illustration

We are going to study the problem of robust optimisation on a toy problem, based on Branin-Hoo’s function defined on $\mathcal{K} = \mathcal{U} = [0, 5]$. This function has been slightly modified to ensure strict positivity.

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

Estimate	Definition	Interpretation
$\hat{\mathbf{k}}_{\text{mean}}$	$\arg \min_{\mathbf{k} \in \mathcal{K}} \mathbb{E}_U [J(\mathbf{k}, \mathbf{U})]$	Long run performances
$\hat{\mathbf{k}}_{\text{var}}$	$\arg \min_{\mathbf{k} \in \mathcal{K}} \text{Var}_U [J(\mathbf{k}, \mathbf{U})]$	Steady performances
$\hat{\mathbf{k}}_{\text{MPE}}$	$\arg \max_{\mathbf{k} \in \mathcal{K}} \mathbb{P}_U [\mathbf{k} = \mathbf{k}^*(\mathbf{U})]$	Most often minimiser
$(\bar{\mathbf{k}}_p, \check{\alpha}_p)$	$\inf_{\alpha \geq 1} \{ \alpha \mid \exists \bar{\mathbf{k}}_p \in \mathcal{K}, R_\alpha(\bar{\mathbf{k}}_p) \geq p \}$	Acceptable values with fixed probability p

$$J_{\text{BH}}(x_1, x_2) = \frac{1}{51.95} \left[\left(\bar{x}_2 - \frac{5.1\bar{x}_1^2}{4\pi^2} + \frac{5\bar{x}_1}{\pi} - 6 \right)^2 + \left(10 - \frac{10}{8\pi} \right) \cos(\bar{x}_1) - 44.81 \right] + 2 \quad (22)$$

$$\text{with } \bar{x}_1 = 3x_1 - 5, \quad \bar{x}_2 = 3x_2 \quad (23)$$

From the function in Eq. (22), we define the BH function as $(\mathbf{k}, \mathbf{u}) \mapsto J_{\text{BH}}(\mathbf{k}, \mathbf{u})$, and BHs as $(\mathbf{k}, \mathbf{u}) \mapsto J_{\text{BH}}(\mathbf{u}, \mathbf{k})$. Although

the functions are quite similar, the asymmetric roles of \mathbf{k} and \mathbf{u} result in different behaviour for the different estimations.

5 Finally, we assume that $\mathbf{U} \sim \text{Unif}([0; 5])$

4.3.1 Moments and conditional minimisers

To compute the MPE, we used different methods to estimate the distribution of the minimisers: KDE (kernel density estimation) (Parzen, 1962; Bishop, 2006) and the EM algorithm (Dempster et al., 1977; Borman, 2004) to fit a mixture of three Gaussian distributions. Figure 1 shows the contour plot, conditional mean and conditional minimisers of the BH function (right)

10 and the BHs function (left).

For the BHs function, we can observe that those two methods give the same results, as the MPE can be estimated at approximately 0.8. However for the BH function, this gives different results. Using KDE, the MPE is found to be at around 1.5, while using EM, the MPE is 3.8. This difference is explained by the accumulation of minimisers at 3.8: the EM algorithm fits this accumulation with a normal distribution with a very small variance. In the meantime, we can consider the mixing coefficients:

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

4.3.2 Relaxation of the constraint

Following its definition in Eq. (21), $\check{\alpha}_1$ has been numerically evaluated at 1.9506 for the BH function. Figure 2 shows the

20 contour plot, and the frontier corresponding to the couples of points (\mathbf{k}, \mathbf{u}) such that $J(\mathbf{k}, \mathbf{u}) = \check{\alpha}_1 J(\mathbf{k}^*(\mathbf{u}), \mathbf{u})$, and the curve $\mathbf{k} \mapsto R_{\check{\alpha}_1}(\mathbf{k})$. In Figure 3, the profiles refers to the function $\mathbf{u} \mapsto J(\mathbf{k}, \mathbf{u})$ for a given \mathbf{k} , and is useful to visualise the range of values taken by the cost function for different estimates: $\hat{\mathbf{k}}_{\text{mean}}$ minimising the mean value, $\hat{\mathbf{k}}_{\text{var}}$ minimising the variance and

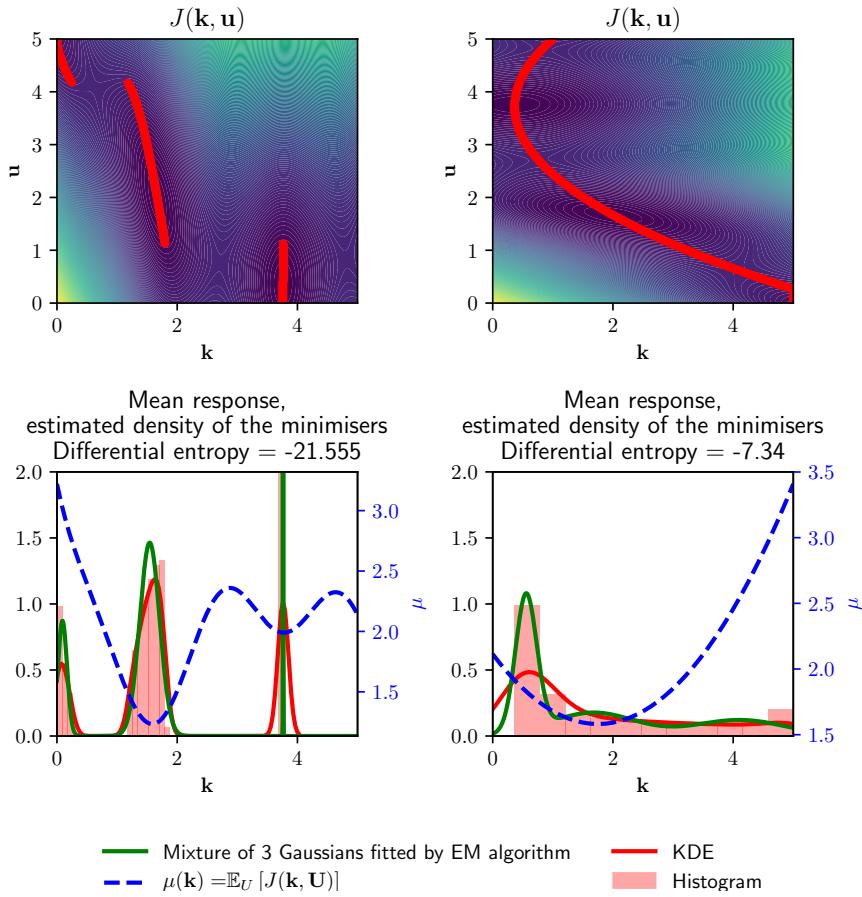


Figure 1. Top Left: BH function along with conditional minimisers in red. Bottom left: Estimated density using KDE and EM algorithm and conditional mean. Right: Same quantities with BHs

\check{k}_1 as defined above. The shaded regions correspond to the interval $[J(\mathbf{k}^*(\mathbf{u}), \mathbf{u}), \check{\alpha}_1 J(\mathbf{k}^*(\mathbf{u}), \mathbf{u})]$. For this particular problem, the minimiser of the mean response and \check{k}_1 are very close.

The fact that \hat{k}_{mean} and \check{k}_1 are similar is specific to the BH function: For the BHs function, the two estimates are quite different, as seen on Figure 2.

5 The profiles shown Figure 4 explain this difference. In this case \hat{k}_{mean} performs very well for $\mathbf{u} > 2$, yet for $\mathbf{u} < 2$, it produces high values of the function. In that case, performances of \check{k}_1 are closer to the performances of \hat{k}_{var} , but performs worse than \hat{k}_{mean} and \hat{k}_{var} for $\mathbf{u} > 2$.

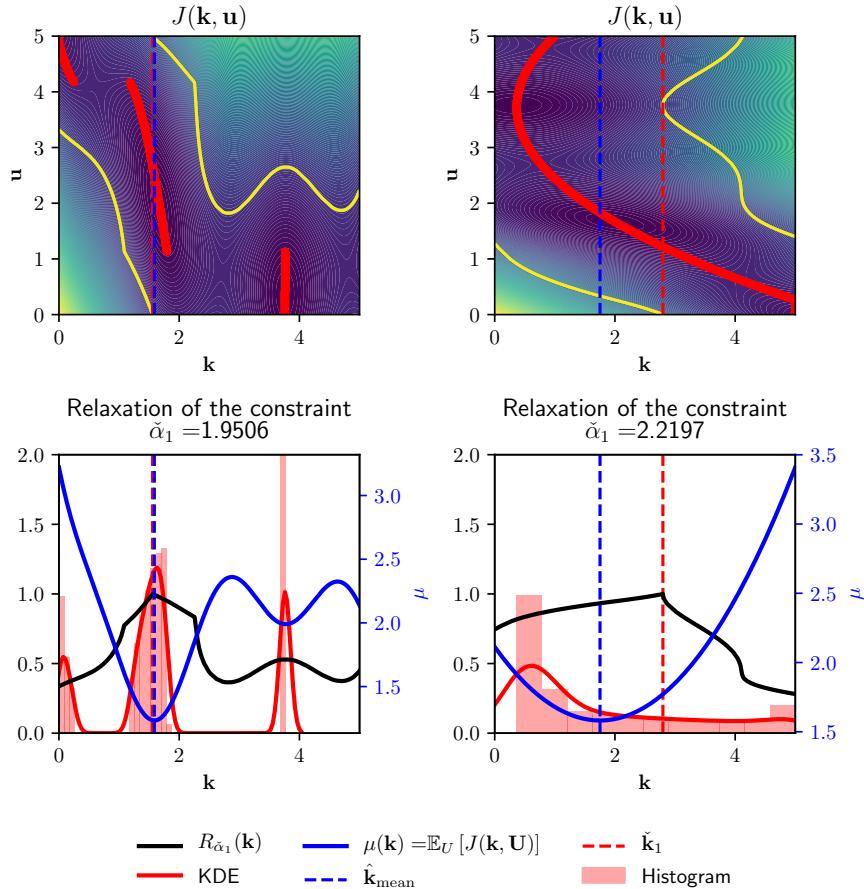


Figure 2. Top: BH function and its conditional minimisers. The thick yellow lines the boundary of the acceptable region, where $\phi_\alpha = 0$. The red dashed line is the estimate \check{k}_1 . Bottom: $R_{\check{\alpha}}$, conditional mean and estimated density of the minimisers

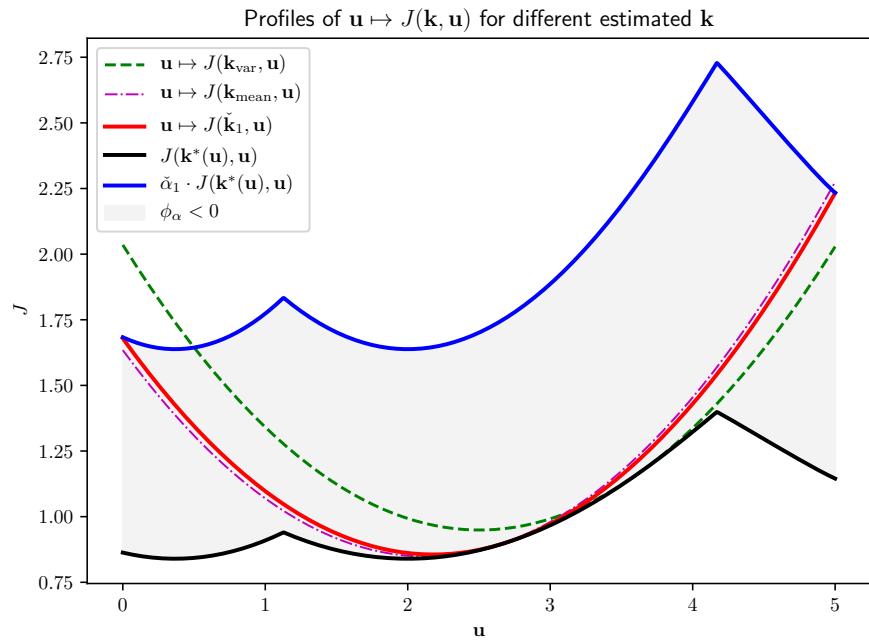


Figure 3. Profiles of BH function for different estimates of \mathbf{k}

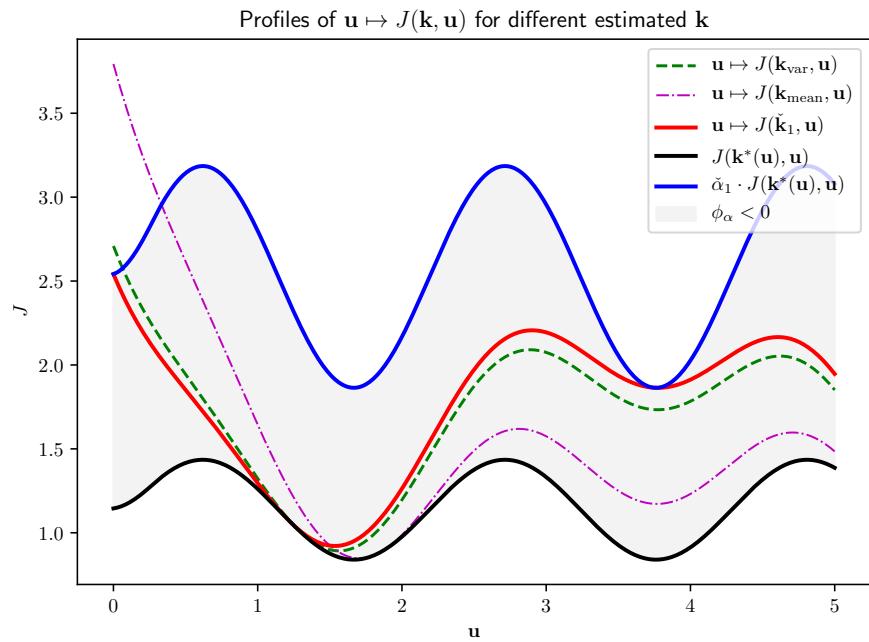


Figure 4. Profiles of BHs for different estimates of \mathbf{k}

5 Application: calibration of a numerical model

5.1 Context: Shallow water equations and bottom friction

The quality of water circulation modelling in coastal areas is largely dependent on the boundary conditions, (e.g. the tide), the bathymetry and the friction induced by the rugosity of the ocean bed (Boutet, 2015). The latter will then be estimated in a robust manner, by applying the criteria introduced above.

The model studied is based on the one-dimensional shallow water equations, described Eq. (24), solved numerically on a domain D . h is the height of the water column, q is the discharge, and z is the bathymetry, while g is the usual gravitation constant. \mathbf{k} is a quadratic friction term, proportional to the square of the inverse of Manning-Strickler coefficient.

$$\begin{cases} \partial_t h + \partial_x q = 0 \\ \partial_t q + \partial_x \left(\frac{q^2}{h} + \frac{g}{2} h^2 \right) = -gh\partial_x z - \mathbf{k}q|q|h^{-7/3} \\ h(0, t) = 20.0 + \text{Amplitude} \cdot \sin \left(\frac{2\pi}{\mathbf{u}} t \right) \\ q(0, t) = 0 \end{cases} \quad (24)$$

10 The estimation of the bottom friction and/or topography has been treated in a deterministic setting, using tools from variational data assimilation (e.g. Das and Lardner, 1991, 1992; Honnorat et al., 2010). If the boundary conditions are not well-known, the estimation will then compensate this misspecification. The resulting estimate will then be optimal for this particular configuration, but will not acknowledge the uncertainty upon the boundary conditions. Instead, we are looking for an estimate that is robust considering the variability of the uncertain boundary conditions.

15 The problem of the estimation of the bottom friction will be studied using twin experiments. For this purpose, a few references simulations have been performed in order to produce different observations \mathbf{y}^o . Those scenarios are detailed Table 3. The true bottom friction \mathbf{k}^t is the same for all simulations, and is a non-constant function of the spatial domain $D = [0, L]$ defined as:

$$\mathbf{k}_t = \mathbf{k}_t(x) = 0.2 \cdot (1 + \sin(2\pi x/L)) \quad (25)$$

20 However, the space \mathcal{K} in which we wish to estimate the bottom friction is one-dimensional. This leads to the strict positivity of the cost function, as no perfect fit can be achieved.

As presented on Table 2, we consider two a priori chosen variable: $\mathbf{u} \in \mathcal{U}$ that is the period of the sine wave on the left boundary. The aleatoric uncertainty on this variable is modelled by a random variable distributed uniformly on \mathcal{U} . The amplitude of the sine wave is also chosen a priori and set to an arbitrary value, that may be different from the true amplitude used to generate \mathbf{y}^o .

Table 2. Modelling of the estimation problem

Parameter	Modelling	Type of parameter
Bottom friction: k	$\mathcal{K} = [0, 1]$	Calibrated parameter
Period: U	Unif ([14.7, 15.3])	A priori chosen and uncertain
Amplitude	5.0	A priori chosen and fixed

5.2 Deterministic optimisation

The problem has first been studied in a deterministic setting. The following optimisation problem has been performed numerically:

$$\text{Minimise } k \in \mathcal{K} \mapsto J(k, u^b) \quad \text{for } u^b = \mathbb{E}_U[U] = 15.0 \quad (26)$$

- 5 At the same time, the background value of the amplitude is fixed and set to 5.0. Different scenario have been established (i.e. different u^t and amplitude producing different y^o), and are detailed in Table 3. The configurations correspond to different values of the amplitude (letter “A”) and the period (letter “P”). If the truth value is greater (resp. smaller) than the background value, the lowercase letter “p” follows (resp. “m”).

On the same table, are shown the result \hat{k} of the optimisation defined in Eq. (26) for each one of the different operating conditions.

Table 3. Settings for reference simulations, and optimisation results

Scenario for u^t :	$\mathbb{E}_U[U]$	Pm	ApPm	AmPm	Pp	ApPp	AmPp
$u^t - u^b$	0.0	-0.1	-0.2	-0.1	+0.2	+0.1	+0.2
Amplitude - 5.0	0.0	0.0	+0.1	-0.2	0.0	+0.2	-0.1
\hat{k}	0.1043	0.2028	0.2709	0.2427	0.0	0.0	0.0015

Optimising without uncertainties (meaning that the $u^b = u^t$) yields an optimal value \hat{k} close to 0.10, that we define to be \hat{k}_{opt} . We can observe two different regimes, depending on the sign of $u^t - \mathbb{E}_U[U]$. If this quantity is positive, optimisation will produce an estimation close to 0. If it is negative, the estimate is larger than 0.20.

5.3 Robust estimates

- 15 In order to compute the robust estimates, the cost function has been evaluated on a regular grid of 150×150 points over $\mathcal{K} \times \mathcal{U}$. \hat{k}_{mean} , \hat{k}_{var} , $\check{k}_{.9}$, $\check{k}_{.95}$ and \check{k}_1 (along with their corresponding factors $\check{\alpha}_p$) have been evaluated for the different configurations, detailed in Table 3. As shown on Table 4, \check{k}_p does not vary significantly when the level of confidence p increases.

Table 4. Mean and variance optimisers, and \check{k}_p for different levels of confidence p

Scenario	$\mathbb{E}_U[\mathbf{U}]$	Pm	ApPm	AmPm	Pp	ApPp	AmPp
\hat{k}_{mean}	0.1678	0.2214	0.2617	0.2685	0.0872	0.0738	0.1275
\hat{k}_{var}	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$\check{k}_{.90}$	0.1074	0.1074	0.0872	0.1678	0.1074	0.0738	0.1275
$\check{\alpha}_{.90}$	1.1754	1.1718	1.1779	1.1293	1.1768	1.1807	1.1406
$\check{k}_{.95}$	0.1074	0.1074	0.0872	0.1611	0.1074	0.0671	0.1275
$\check{\alpha}_{.95}$	1.1802	1.1806	1.1868	1.1320	1.1861	1.2027	1.1525
\check{k}_1	0.1074	0.1074	0.0872	0.1611	0.1074	0.0671	0.1275
$\check{\alpha}_1$	1.1944	1.1922	1.1961	1.1321	1.1976	1.2035	1.1682

In that case, the minimiser of the variance \hat{k}_{var} is the right boundary of \mathcal{K} , regardless of the value of \mathbf{u}^t . This is due to the fact that at this point, the cost function produces bad, yet steady performances. From a physical point of view, increasing the bottom friction will decrease the momentum. The movement of the water surface will be damped, and the operating conditions will have less effects.

5 Looking at \hat{k}_{mean} , we can see that the estimates obtained are larger than 0.22 when $\mathbf{u}^t < \mathbb{E}_U[\mathbf{U}]$ (that is for Pm, ApPm, and AmPm). Those results are similar to the one obtained in the deterministic setting of Table 3. However, for $\mathbf{u}^t > \mathbb{E}_U[\mathbf{U}]$ (Pp, ApPp, and AmPp), \hat{k}_{mean} is much closer to \hat{k}_{opt} , the optimal value without uncertainties.

Now for \check{k}_1 , while the estimates for scenarios Pp, ApPp and AmPp are similar to \hat{k}_{mean} , the obtained values for the other scenarios are close to \hat{k} . Figures 5 and 6 show the contour plot and profiles of the cost function in centered conditions ($\mathbb{E}_U[\mathbf{U}]$ scenario). Figures 7 and 8 (resp. Figures 9 and 10) show the contour plot and profiles of the cost function for the ApPp scenario (resp. AmPm scenario).

The concentration of \check{k}_p close to \hat{k}_{opt} is caused by the very small value of the cost function at its minimum over $\mathcal{K} \times \mathcal{U}$. Close to this minimum, the region of acceptable \mathbf{k} will be growing slowly as α increases, weighing the estimation toward \hat{k}_{opt} .

The fact that we retrieved values close to the optimal one should not overlook that a robust estimate is not necessarily \hat{k}_{opt} ,
15 and that this result is specific to this problem.

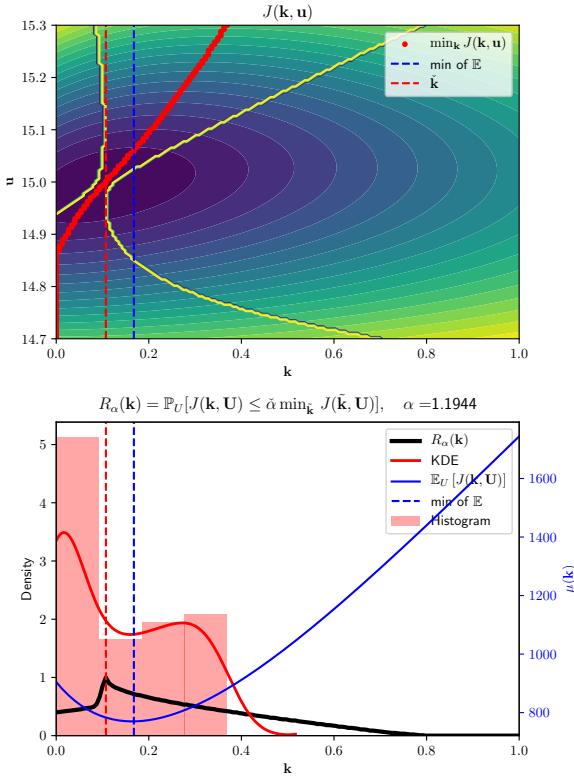


Figure 5. Top: Cost function of the SWE problem ($\mathbb{E}_U[U]$ scenario) and its conditional minimisers. The thick yellow lines represent the boundary of the region where $\phi_\alpha < 0$. The red dashed line is the estimate \check{k}_1 . Bottom: $R_{\check{\alpha}}$

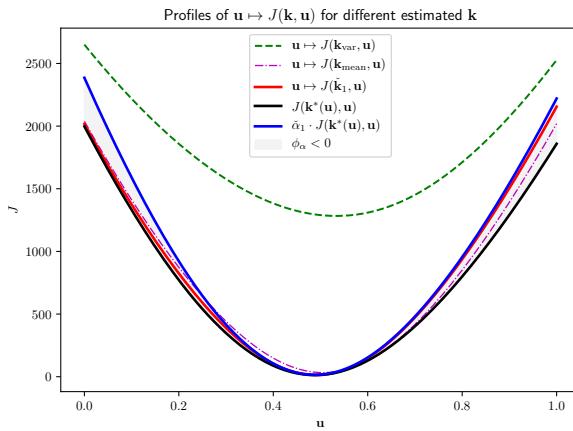


Figure 6. Profiles of different estimates of k for the $\mathbb{E}_U[U]$ scenario

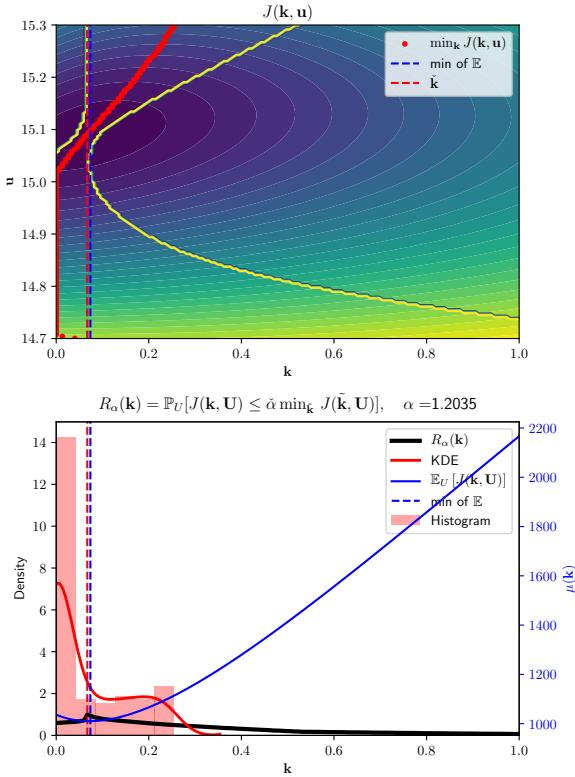


Figure 7. Top: Cost function of the SWE problem (ApPp scenario) and its conditional minimisers. The thick yellow lines represent the boundary of the region where $\phi_\alpha < 0$. The red dashed line is the estimate \check{k}_1 . Bottom: $R_{\bar{\alpha}}$

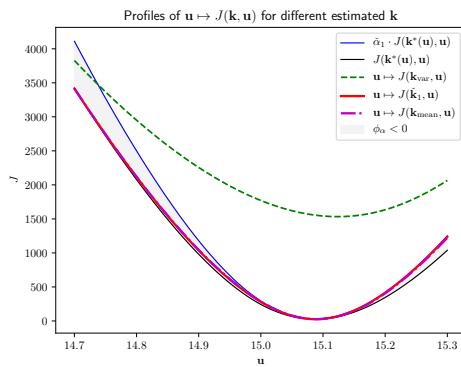


Figure 8. Profiles of different estimates of k for the ApPp scenario

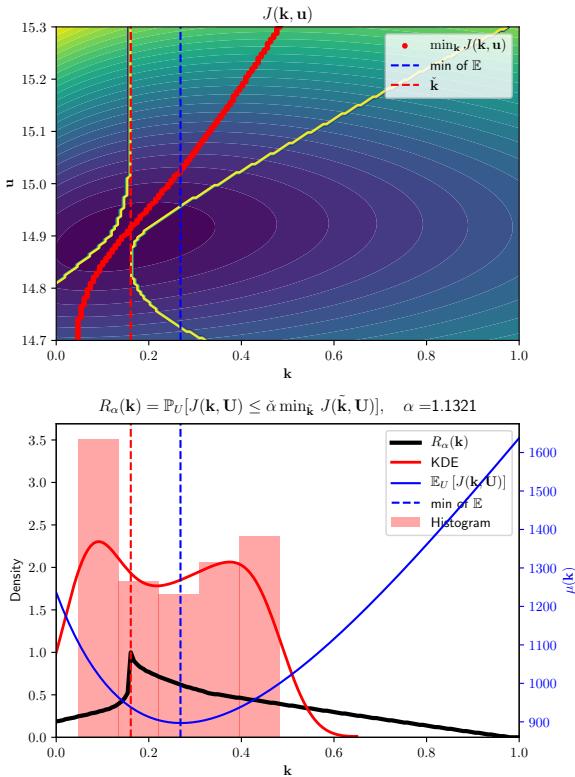


Figure 9. Top: Cost function of the SWE problem (AmPm scenario) and its conditional minimisers. The thick yellow lines represent the boundary of the region where $\phi_\alpha < 0$. The red dashed line is the estimate $\check{\mathbf{k}}_1$. Bottom: $R_{\check{\alpha}}$

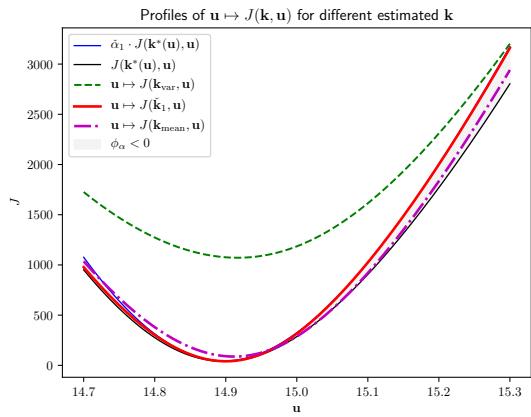


Figure 10. Profiles of different estimates of \mathbf{k} for the AmPm scenario

6 Conclusions

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

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

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

A solution to explore is the use of surrogate models. Based on an initial design of experiment, the (expensive) cost function is replaced by a function, that is cheap to evaluate. For instance, Ginsbourger et al. (2014) uses Gaussian Process regression and sequential methods to explore the conditional minimisers.

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

Code availability. TEXT

Code and data availability. TEXT

Author contributions. TEXT

Competing interests. TEXT

20 *Disclaimer.* TEXT

Acknowledgements. TEXT

References

- Baudouï, V.: Optimisation Robuste Multiobjectifs Par Modèles de Substitution, Ph.D. thesis, Toulouse, ISAE, 2012.
- Berger, J. O., Moreno, E., Pericchi, L. R., Bayarri, M. J., Bernardo, J. M., Cano, J. A., De la Horra, J., Martín, J., Ríos-Insúa, D., and Betrò, B.: An Overview of Robust Bayesian Analysis, *Test*, 3, 5–124, 1994.
- 5 Bishop, C. M.: *Pattern Recognition and Machine Learning*, Springer, 2006.
- Borman, S.: The Expectation Maximization Algorithm-a Short Tutorial, Submitted for publication, pp. 1–9, 2004.
- Boutet, M.: Estimation Du Frottement Sur Le Fond Pour La Modélisation de La Marée Barotrope, Ph.D. thesis, Université d'Aix Marseille, 2015.
- Das, S. K. and Lardner, R. W.: On the Estimation of Parameters of Hydraulic Models by Assimilation of Periodic Tidal Data, *Journal of Geophysical Research*, 96, 15 187, <https://doi.org/10.1029/91JC01318>, 1991.
- 10 Das, S. K. and Lardner, R. W.: Variational Parameter Estimation for a Two-Dimensional Numerical Tidal Model, *International Journal for Numerical Methods in Fluids*, 15, 313–327, <https://doi.org/10.1002/fld.1650150305>, 1992.
- Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.
- 15 Ginsbourger, D., Baccou, J., Chevalier, C., Perales, F., Garland, N., and Monerie, Y.: Bayesian Adaptive Reconstruction of Profile Optima and Optimizers, *SIAM/ASA Journal on Uncertainty Quantification*, 2, 490–510, <https://doi.org/10.1137/130949555>, 2014.
- Grodzevich, O. and Romanko, O.: Normalization and Other Topics in Multi-Objective Optimization, 2006.
- Hennig, P. and Schuler, C. J.: Entropy Search for Information-Efficient Global Optimization, 2011.
- Honorat, M., Monnier, J., Rivière, N., Huot, E., and Le Dimet, F.-X.: Identification of Equivalent Topography in an Open Channel Flow 20 Using Lagrangian Data Assimilation, *Computing and Visualization in Science*, 13, 111–119, <https://doi.org/10.1007/s00791-009-0130-8>, 2010.
- Huber, P. J.: Robust Statistics, in: *International Encyclopedia of Statistical Science*, pp. 1248–1251, Springer, 2011.
- Lehman, J. S., Santner, T. J., and Notz, W. I.: Designing Computer Experiments to Determine Robust Control Variables, *Statistica Sinica*, pp. 571–590, 2004.
- 25 Markowitz, H.: Portfolio Selection, *The journal of finance*, 7, 77–91, 1952.
- Parzen, E.: On Estimation of a Probability Density Function and Mode, *The annals of mathematical statistics*, 33, 1065–1076, 1962.
- Rao, V., Sandu, A., Ng, M., and Nino-Ruiz, E.: Robust Data Assimilation Using \$L_1\$ and Huber Norms, SciRate, 2015.
- Villemonteix, J., Vazquez, E., and Walter, E.: An Informational Approach to the Global Optimization of Expensive-to-Evaluate Functions, arXiv:cs/0611143, 2006.
- 30 Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B., Janssen, P., and Krayer von Krauss, M. P.: Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support, *Integrated assessment*, 4, 5–17, 2003.