Robust calibration of numerical models based on Relative-regret

Robust Estimation of bottom friction

Victor Trappler  
victor.trappler@univ-grenoble-alpes.fr

É. Arnaud, L. Debreu, A. Vidard
AIRSEA Research team (Inria)
Laboratoire Jean Kuntzmann

GdR MASCOTNUM, Grenoble, 2020
Uncertainties in the modelling

Does reducing the error on the parameters leads to the compensation of the unaccounted natural variability of the physical processes?
Outline

Introduction

Calibration problem

Robust minimization

Surrogates

Conclusion
Calibration problem
Computer code and inverse problem

Input
- $\theta$: Control parameter
- $u$: Environmental variables (fixed and known)

Output
- $M(\theta, u)$: Quantity to be compared to observations

Diagram:
- Environmental variables $u \in \mathbb{U}$ fixed
- Control variable $\theta \in \Theta$
- Direct Simulation
- Inverse Problem
- $M(\theta, u)$
- $y_{obs}$
Data assimilation framework

Let $u \in \mathbb{U}$.

$$\hat{\theta} = \arg \min_{\theta \in \Theta} J(\theta) = \arg \min_{\theta \in \Theta} \frac{1}{2} \| \mathcal{M}(\theta, u) - y^{\text{obs}} \|^2$$

→ Deterministic optimization problem
→ Possibly add regularization
→ Classical methods: Adjoint gradient and Gradient-descent

BUT

○ What if $u$ does not reflect accurately the observations?
○ Does $\hat{\theta}$ compensate the errors brought by this random misspecification? (~overfitting)
• The friction $\theta$ of the ocean bed has an influence on the water circulation
• Depends on the type and/or characteristic length of the asperities
• Subgrid phenomenon
• $u$ parametrizes the BC
Different types of uncertainties

### Epistemic or aleatoric uncertainties? [WHR+03]

- **Epistemic uncertainties**: From a lack of knowledge, that can be reduced with more research/exploration
- **Aleatoric uncertainties**: From the inherent variability of the system studied, operating conditions

→ But where to draw the line?

Our goal is to take into account the aleatoric uncertainties in the estimation of our parameter.
Instead of considering $u$ fixed, we consider that $u \sim U$ r.v. (with known pdf $\pi(u)$), and the output of the model depends on its realization.
Instead of considering $u$ fixed, we consider that $u \sim U$ r.v. (with known pdf $\pi(u)$), and the output of the model depends on its realization.
The cost function as a random variable

- The computer code is deterministic, and takes $\theta$ and $u$ as input:
  \[ M(\theta, u) \]

- The deterministic quadratic error is now
  \[ J(\theta, u) = \frac{1}{2} \| M(\theta, u) - y^{\text{obs}} \|^2 \]

"$\hat{\theta} = \arg \min_{\theta \in \Theta} J(\theta, u)"" but what can we do about $u$?
Misspecification of $u$: twin experiment setup

Minimization performed on $\theta \rightarrow J(\theta, u)$, for different $u$:
Misspecification of $u$: twin experiment setup

Minimization performed on $\theta \rightarrow J(\theta, u)$, for different $u$:

Well-specified model
Misspecification of $u$: twin experiment setup

Minimization performed on $\theta \rightarrow J(\theta, u)$, for different $u$:

1% error on the amplitude of the M2 tide
Misspecification of $u$: twin experiment setup

Minimization performed on $\theta \rightarrow J(\theta, u)$, for different $u$:

1% error on the amplitude of the M2 tide
Robustness and estimation of parameters

Robustness: get good performances when the environmental parameter varies

- Define criteria of robustness, based on $J(\theta; u)$, that will depend on the final application
- Be able to compute them in a reasonable time
Robust minimization

Criteria of robustness
Non-exhaustive list of “Robust” Objectives

- Worst case [MWP13]:
  \[
  \min_{\Theta} \left\{ \max_{u \in \mathcal{U}} J(\theta, u) \right\}
  \]

- M-robustness [LSN04]:
  \[
  \min_{\Theta} \mathbb{E}_U [J(\theta, U)]
  \]

- V-robustness [LSN04]:
  \[
  \min_{\Theta} \text{Var}_U [J(\theta, U)]
  \]

- Multiobjective [Bau12]:
  Pareto frontier

- Best performance achievable given \( u \sim U \)
“Most Probable Estimate”, and relaxation

Given \( u \sim U \), the optimal value is \( J^*(u) \), attained at \( \theta^*(u) = \arg\min_{\theta \in \Theta} J(\theta, u) \).
“Most Probable Estimate”, and relaxation

Given $u \sim U$, the optimal value is $J^*(u)$, attained at $\theta^*(u) = \arg \min_{\theta \in \Theta} J(\theta, u)$.

The minimizer can be seen as a random variable:

$$\theta^*(U) = \arg \min_{\theta \in \Theta} J(\theta, U)$$

→ estimate its density (how often is the value $\theta$ a minimizer)

$$p_{\theta^*}(\theta) = \mathbb{P}_U [J(\theta, U) = J^*(U)]$$
Given $u \sim U$, the optimal value is $J^*(u)$, attained at $\theta^*(u) = \arg\min_{\theta \in \Theta} J(\theta, u)$.

The minimizer can be seen as a random variable:

$$\theta^*(U) = \arg\min_{\theta \in \Theta} J(\theta, U)$$

$\rightarrow$ estimate its density (how often is the value $\theta$ a minimizer)

$$p_{\theta^*}(\theta) = "\mathbb{P}_U [J(\theta, U) = J^*(U)]"$$

How to take into account values not optimal, but not too far either $\rightarrow$ relaxation of the equality with $\alpha > 1$:

$$\Gamma_\alpha(\theta) = \mathbb{P}_U [J(\theta, U) \leq \alpha J^*(U)]$$
• Sample $u \sim U$, and solve
  \[ \theta^*(u) = \arg \min_{\theta \in \Theta} J(\theta, u) \]
• Sample $u \sim \mathcal{U}$, and solve
  \[ \theta^*(u) = \arg\min_{\theta \in \Theta} J(\theta, u) \]

• Set of conditional minimisers:
  \[ \{(\theta^*(u), u) \mid u \in \mathcal{U}\} \]
- Sample $u \sim U$, and solve
  \[ \theta^*(u) = \arg\min_{\theta \in \Theta} J(\theta, u) \]
- Set of conditional minimisers:
  \[ \{(\theta^*(u), u) \mid u \in U\} \]
- Set $\alpha \geq 1$
• Sample $u \sim U$, and solve
  $\theta^*(u) = \arg \min_{\theta \in \Theta} J(\theta, u)$

• Set of conditional minimisers:
  $\{(\theta^*(u), u) \mid u \in \mathbb{U}\}$

• Set $\alpha \geq 1$

• $R_\alpha(\theta) = \{u \mid J(\theta, u) \leq \alpha J^*(u)\}$

• $\Gamma_\alpha(\theta) = \mathbb{P}_U [U \in R_\alpha(\theta)]$
Getting an estimator

$\Gamma_{\alpha}(\theta)$: probability that the cost (thus $\theta$) is $\alpha$-acceptable

- If $\alpha$ known, maximize the probability that $\theta$ gives acceptable values:

$$\max_{\theta \in \Theta} \Gamma_{\alpha}(\theta) = \max_{\theta \in \Theta} \mathbb{P}_{U} [J(\theta, U) \leq \alpha J^*(U)] \quad (1)$$

- Set a target probability $1 - \eta$, and find the smallest $\alpha$.

$$\inf \{ \alpha \mid \max_{\theta \in \Theta} \Gamma_{\alpha}(\theta) \geq 1 - \eta \} \quad (2)$$

More generally, let us define the RR family

$$\left\{ \hat{\theta} \mid \hat{\theta} = \arg \max_{\theta \in \Theta} \Gamma_{\alpha}(\theta), \alpha > 1 \right\} \quad (3)$$
Why the relative regret?

- Relative regret
  - $\alpha$-acceptability regions large for flat and bad situations ($J^*(u)$ large)
  - Conversely, puts high confidence when $J^*(u)$ is small
  - No units $\rightarrow$ ratio of costs
Surrogates

How to compute $\hat{\theta}$ in a reasonable time?
Surrogates, and cost function

- Replace expensive model by a computationally cheap metamodel (≈ plug-in approach)
- Adapted sequential procedures e.g. EGO

→ Kriging (Gaussian Process Regression) [Mat62, Kri51]
Surrogates, and cost function

- Replace expensive model by a computationally cheap metamodel (\(\sim\) plug-in approach)
- Adapted sequential procedures e.g. EGO

\[ Y \sim \text{GP}(m_Y(\cdot), C_Y(\cdot, \cdot)) \] GP regression of \( J \) on \( \Theta \times \mathcal{U} \), using an initial design \( \mathcal{X} = \{((\theta_i, u_i), J(\theta_i, u_i))\} \)
Estimation of $\theta^*(u)$, $J^*(u)$

Estimation of $J^*(u)$ and $\theta^*(u)$: Enrich the design according to PEI criterion [GBC⁺14].
GP of the “penalized” cost function

What about $J(\theta, u) - \alpha J^*(u)$?

$$Y \sim \text{GP} \left( m_Y(\cdot); C_Y(\cdot, \cdot) \right) \text{ on } \Theta \times \mathbb{U} \quad (4)$$

$$\Delta_\alpha = Y - \alpha Y^* \quad (5)$$

Still a GP

$$\Delta_\alpha(\theta, u) \sim \text{GP} \left( m_\alpha(\cdot); C_\alpha(\cdot, \cdot) \right) \quad (6)$$

$$m_\alpha(\theta, u) = m_Y(\theta, u) - \alpha m_Y^*(u) \quad (7)$$

$$C_\alpha(\theta, u) = \sigma_Y^2(\theta, u) + \alpha^2 \sigma_Y^2(\cdot, \cdot) - 2\alpha C_Y(\cdot, \cdot) \approx \sigma_Y(\theta, u) + \alpha^2 \sigma_Y^2(\cdot, \cdot) - 2\alpha C_Y((\theta, u), (\theta^*_Y(u), u)) \quad (8)$$

Estimate the “probability of failure” [BGL+12, EGL11]

$$\mathbb{P}_U \left[ J(\theta, U) - \alpha J^*(U) \leq 0 \right] \approx \mathbb{P}_U \left[ \mathbb{P}_Y \left[ \Delta_\alpha \leq 0 \right] \right]$$
GP of the “penalized” cost function

What about $J(\theta, u) - \alpha J^*(u)$?

$$Y \sim \text{GP} \left( m_Y(\cdot); C_Y(\cdot, \cdot) \right) \text{ on } \Theta \times \mathbb{U}$$

(4)

$$\Delta_\alpha = Y - \alpha Y^*$$

(5)

Still a GP

$$\Delta_\alpha(\theta, u) \sim \text{GP} \left( m_\alpha(\cdot); C_\alpha(\cdot, \cdot) \right)$$

(6)

$$m_\alpha(\theta, u) = m_Y(\theta, u) - \alpha m_Y^*(u)$$

(7)

$$C_\alpha(\theta, u) = \sigma_Y^2(\theta, u) + \alpha^2 \sigma_Y^2(u) - 2\alpha C_Y((\theta, u), (\theta^*_Y(u), u))$$

(8)

Estimate the “probability of failure” [BGL+12, EGL11]

$$\mathbb{P}_U \left[ J(\theta, U) - \alpha J^*(U) \leq 0 \right] \approx \mathbb{P}_U \left[ \mathbb{P}_Y \left[ \Delta_\alpha \leq 0 \right] \right]$$
Joint space or objective-oriented exploration

Because of $J^*(u)$, it is often not enough to select the point where the uncertainty is high. Generally, two main approaches can be considered:

- Estimate the region $\{(\theta, u) \mid J(\theta, u) \leq \alpha J^*(u)\}$, then use the surrogate as a plug-in estimate to compute and maximize $\Gamma_{\alpha}$ → reduce uncertainty on the whole space.

- Select a candidate $\tilde{\theta}$, such that uncertainty on the estimation of $\Gamma_{\alpha}(\tilde{\theta})$ is reduced → reduce uncertainty on $\{\tilde{\theta}\} \times \mathbb{U}$, with $\tilde{\theta}$ well-chosen.
Ad-hoc segmentation according to the depth, and sensitivity analysis: only the shallow coastal regions seem to have an influence.
- $U \sim U[-1, 1]$ uniform r.v. that models the percentage of error on the amplitude of the M2 component of the tide.
- The “truth” ranges from 8mm to 13mm.
- 11.0mm leads to a cost which deviates less than 1% from the optimal value with probability 0.77.
Conclusion
## Conclusion

### Wrapping up
- Problem of a *good* definition of robustness
- Tuning $\alpha$ or $\eta$ reflects risk-seeking or risk-adverse strategies
- Strategies rely heavily on surrogate models, to embed aleatoric uncertainties directly in the modelling

### Perspectives
- Cost of computer evaluations $\rightarrow$ limited number of runs?
- In low dimension, CROCO very well-behaved.
- Dimensionality of the input space $\rightarrow$ reduction of the input space?
Vincent Baudouin.

*Optimisation Robuste Multiobjectifs Par Modèles de Substitution.*

Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez.

*Sequential design of computer experiments for the estimation of a probability of failure.*
B. Echard, N. Gayton, and M. Lemaire.  
**AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation.**  

David Ginsbourger, Jean Baccou, Clément Chevalier, Frédéric Perales, Nicolas Garland, and Yann Monerie.  
**Bayesian Adaptive Reconstruction of Profile Optima and Optimizers.**  
Daniel G. Krige.
A statistical approach to some basic mine valuation problems on the Witwatersrand.

Jeffrey S. Lehman, Thomas J. Santner, and William I. Notz.
Designing computer experiments to determine robust control variables.

Georges Matheron.

Let \( J^*(u) = \min_{\theta \in \Theta} J(\theta, u) \) and \( \theta^*(u) = \arg \min_{\theta \in \Theta} J(\theta, u) \). The regret \( r \):

\[
r(\theta, u) = J(\theta, u) - J^*(u) = - \log \left( \frac{e^{-J(\theta, u)}}{\max_{\theta} \{e^{-J(\theta, u)}\}} \right) \tag{9}
\]

\[
= - \log \left( \frac{\mathcal{L}(\theta, u)}{\max_{\theta \in \Theta} \mathcal{L}(\theta, u)} \right) \tag{10}
\]

→ linked to misspecified LRT: maximize the probability of keeping \( \mathcal{H}_0: \theta \) valid instead of \( \arg \max \mathcal{L} \).
$Y \sim \text{GP}(m_Y(\cdot), C_Y(\cdot, \cdot)) \text{ on } \Theta \times \mathbb{U}$

$$\text{PEI}(\theta, u) = \mathbb{E}_Y \left[ [f_{\min}(u) - Y(\theta, u)]_+ \right]$$ (11)

where $f_{\min}(u) = \max \{ \min_i J(\theta_i, u), \min_{\theta \in \Theta} m_Y(\theta, u) \}$