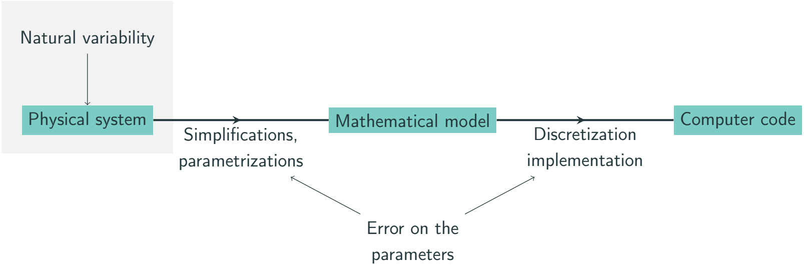
Robust calibration of numerical models based on Relative-regret

Robust Estimation of bottom friction

Victor Trappler victor.trappler@univ-grenoble-alpes.fr
É. Arnaud, L. Debreu, A. Vidard
AIRSEA Research team (Inria) team.inria.fr/airsea/en/
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 ?

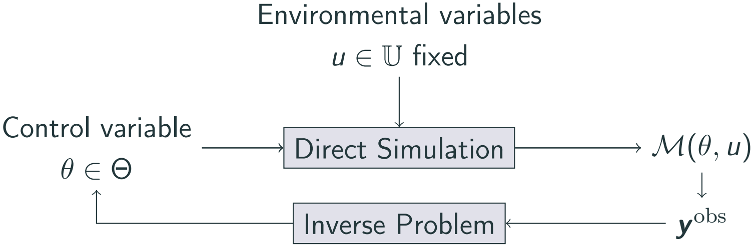
Outline

Calibration problem

Computer code and inverse problem

Input
- θ: Control parameter
- u: Environmental variables (fixed and known)

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



Data assimilation framework

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

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

$\rightarrow$ Deterministic optimization problem

$\rightarrow$ Possibly add regularization

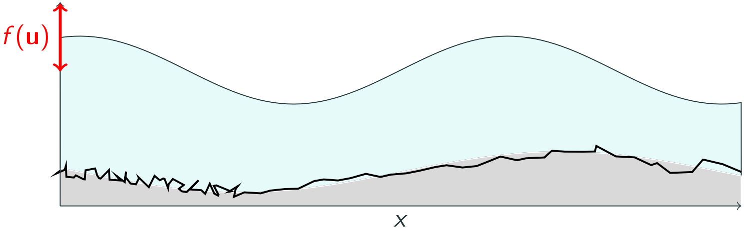
$\rightarrow$ Classical methods: Adjoint gradient and Gradient-descent

BUT

- What if u does not reflect accurately the observations?
- Does $\hat{\theta}$ compensate the errors brought by this random misspecification? ($\sim$overfitting)

Context

- The friction θ 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]

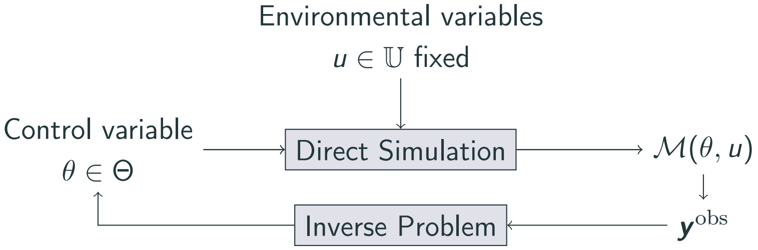
- Epistemic uncertainties: From a lack of knowledge, that can be reduced with more research/exploration
- Aleatoric uncertainties: From the inherent variability of the system studied, operating conditions

$\rightarrow$ But where to draw the line?

Our goal is to take into account the aleatoric uncertainties in the estimation of our parameter.

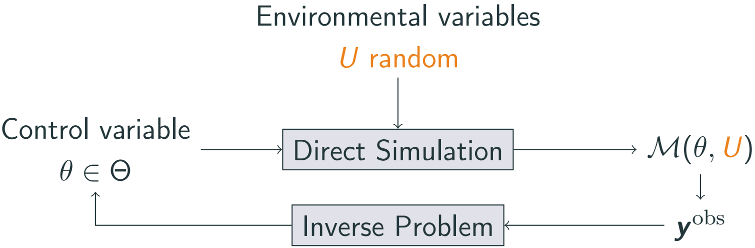
Aleatoric uncertainties

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.



Aleatoric uncertainties

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 θ and u as input:

$$\mathcal{M}(\theta, u)$$

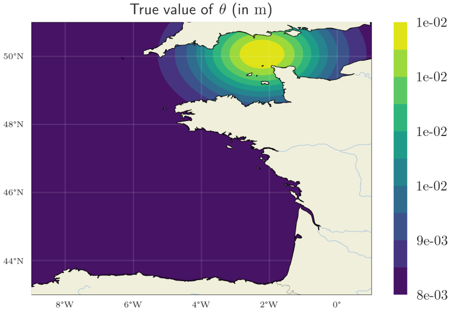
- The deterministic quadratic error is now

$$J(\theta, u) = \frac{1}{2}\|\mathcal{M}(\theta, u) - \boldsymbol{y}^{\mathrm{obs}}\|^2$$

$$''\hat{\theta} = \underset{\theta\in\Theta}{\arg\min}\, J(\theta, u)'' \text{ but what can we do about } u?$$

Misspecification of u: twin experiment setup

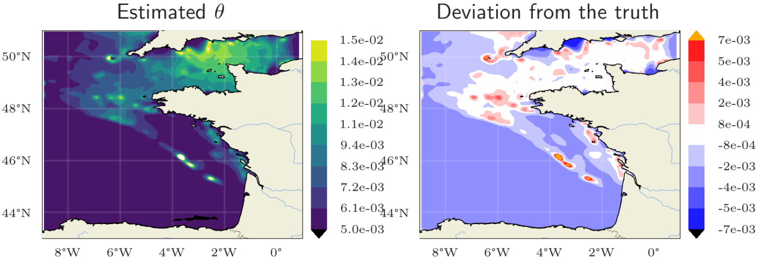
Minimization performed on $\theta \mapsto J(\theta, u)$, for different u:



Misspecification of u: twin experiment setup

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

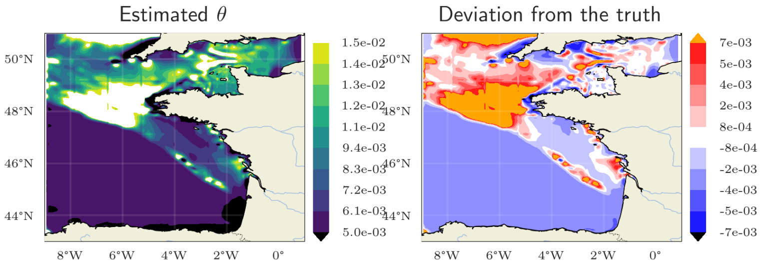
Well-specified model



Misspecification of u: twin experiment setup

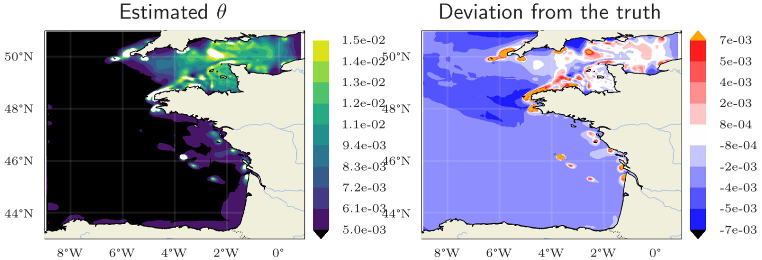
Minimization performed on $\theta \mapsto J(\theta, u)$, for different u:

1% error on the amplitude of the M2 tide



Minimization performed on $\theta \mapsto J(\theta, u)$, 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]:

$$\min_{\theta \in \Theta} \left\{ \max_{u \in \mathbb{U}} J(\theta, u) \right\}$$

- M-robustness [LSN04]:

$$\min_{\theta \in \Theta} \mathbb{E}_U \left[J(\theta, U) \right]$$

- V-robustness [LSN04]:

$$\min_{\theta \in \Theta} \mathbb{V}\mathrm{ar}_U \left[J(\theta, U) \right]$$

- Multiobjective [Bau12]:

Pareto frontier

- Best performance acheivable 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:

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

$\longrightarrow$ estimate its density (how often is the value θ a minimizer)

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

"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:

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

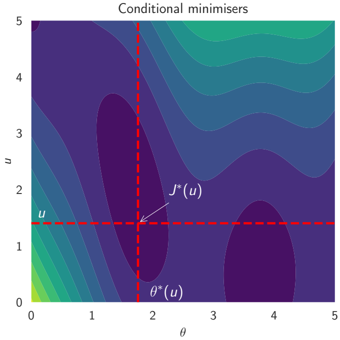
$\longrightarrow$ estimate its density (how often is the value θ a minimizer)

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

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

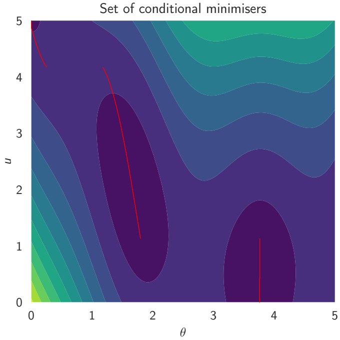
$$\Gamma_\alpha(\theta) = \mathbb{P}_U\left[J(\theta, U) \leq \alpha J^*(U)\right]$$

Illustration



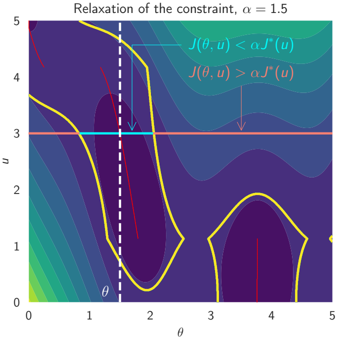
- Sample $u \sim U$, and solve $\theta^*(u) = \arg\min_{\theta \in \Theta} J(\theta, u)$

Illustration



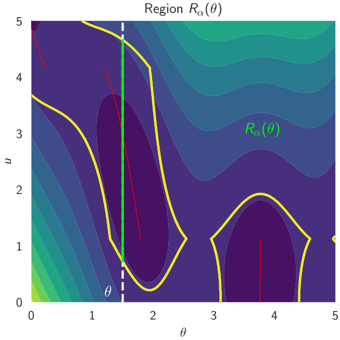
- Sample $u \sim U$, and solve $\theta^*(u) = \arg\min_{\theta \in \Theta} J(\theta, u)$
- Set of conditional minimisers: $\{(\theta^*(u), u) \mid u \in \mathbb{U}\}$

Illustration



- 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 $\alpha \geq 1$

Illustration



- 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 θ) is α-acceptable

- If α known, maximize the probability that θ gives acceptable values:

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

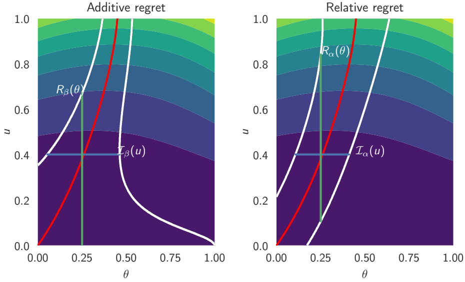
- Set a target probability $1 - \eta$, and find the smallest α.

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

More generally, let us define the RR family

$$\left\{ \hat{\theta} \mid \hat{\theta} = \arg\max_{\theta\in\Theta} \Gamma_\alpha(\theta), \alpha > 1 \right\} \tag{3}$$

Why the relative regret ?



- Relative regret
  - α-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 ($\sim$ plug-in approach)
- Adapted sequential procedures e.g. EGO

$\rightarrow$ Kriging (Gaussian Process Regression) [Mat62, Kri51]

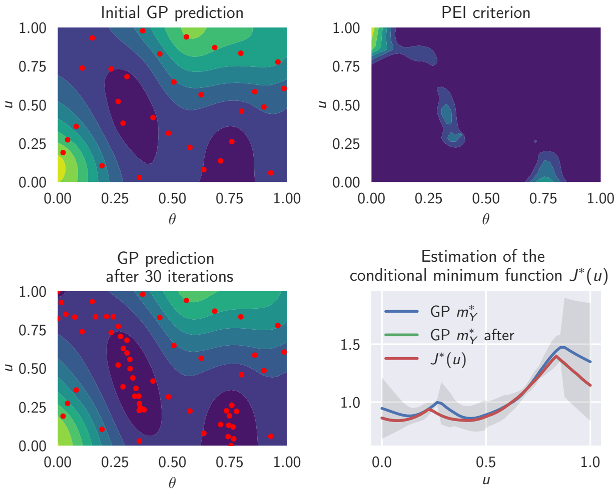
Surrogates, and cost function

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

$\rightarrow$ Kriging (Gaussian Process Regression) [Mat62, Kri51]
$Y \sim \mathrm{GP}\left(m_Y(\cdot), C_Y(\cdot,\cdot)\right)$ GP regression of J on $\Theta \times \mathbb{U}$, using an initial design $\mathcal{X} = \{((\theta_i, u_i), J(\theta_i, u_i))\}$

Estimation of θ^*, $J^*(u)$

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



GP of the "penalized" cost function

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

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

$$\Delta_\alpha = Y - \alpha Y^* \tag{5}$$

Still a GP

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

$$m_\alpha(\theta, u) = m_Y(\theta, u) - \alpha m_Y^*(u) \tag{7}$$

$$C_\alpha(\theta, u) = \sigma_Y^2(\theta, u) + \alpha^2 \sigma_{Y^*}^2(u) - 2\alpha C_Y\left((\theta, u), (\theta_Y^*(u), u)\right) \tag{8}$$

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

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

$$\Delta_\alpha = Y - \alpha Y^* \tag{5}$$

Still a GP

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

$$m_\alpha(\theta, u) = m_Y(\theta, u) - \alpha m_Y^*(u) \tag{7}$$

$$C_\alpha(\theta, u) = \sigma_Y^2(\theta, u) + \alpha^2 \sigma_{Y^*}^2(u) - 2\alpha C_Y\left((\theta, u), (\theta_Y^*(u), u)\right) \tag{8}$$

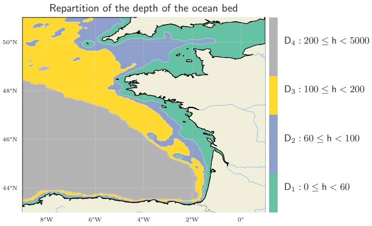
Estimate the "probability of failure" [BGL+12, EGL11]
$\mathbb{P}_U\left[J(\theta, U) - \alpha J^*(U) \leq 0\right] \approx \mathbb{P}_U\left[\mathbb{P}_Y\left[\Delta_\alpha \leq 0\right]\right]$

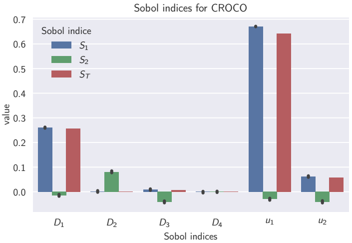
Joint space or objective-oriented exploration

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 Γ_α
  $\rightarrow$ reduce uncertainty on the whole space
- Select a candidate $\tilde{\theta}$, such that uncertainty on the estimation of $\Gamma_\alpha(\tilde{\theta})$ is reduced
  $\rightarrow$ reduce uncertainty on $\{\tilde{\theta}\} \times \mathbb{U}$, with $\tilde{\theta}$ well-chosen.

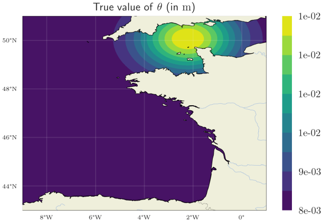
Application to CROCO: Dimension reduction

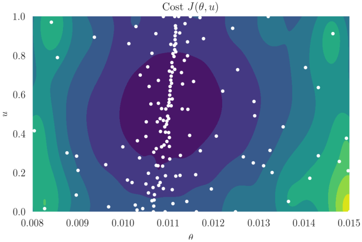




Ad-hoc segmentation according to the depth, and sensitivity analysis: only the shallow coastal regions seem to have an influence.

Robust optimization





- $U \sim \mathrm{U}[-1, 1]$ 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 α or η 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?

References i

Vincent Baudoui.
***Optimisation Robuste Multiobjectifs Par Modèles de Substitution*.**
PhD thesis, Toulouse, ISAE, 2012.

Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez.
Sequential design of computer experiments for the estimation of a probability of failure.
Statistics and Computing, 22(3):773–793, May 2012.

B. Echard, N. Gayton, and M. Lemaire.
AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation.
Structural Safety, 33(2):145–154, March 2011.

David Ginsbourger, Jean Baccou, Clément Chevalier, Frédéric Perales, Nicolas Garland, and Yann Monerie.
Bayesian Adaptive Reconstruction of Profile Optima and Optimizers.
SIAM/ASA Journal on Uncertainty Quantification, 2(1):490–510, January 2014.

Daniel G. Krige.
A statistical approach to some basic mine valuation problems on the Witwatersrand.
Journal of the Southern African Institute of Mining and Metallurgy, 52(6):119–139, 1951.

Jeffrey S. Lehman, Thomas J. Santner, and William I. Notz.
Designing computer experiments to determine robust control variables.
Statistica Sinica, pages 571–590, 2004.

Georges Matheron.
***Traité de Géostatistique Appliquée. 1 (1962)*, volume 1.**
Editions Technip, 1962.

Julien Marzat, Eric Walter, and Hélène Piet-Lahanier.
Worst-case global optimization of black-box functions through Kriging and relaxation.
Journal of Global Optimization, 55(4):707–727, April 2013.

Warren E. Walker, Poul Harremoës, Jan Rotmans, Jeroen P. van der Sluijs, Marjolein BA van Asselt, Peter Janssen, and Martin P. Krayer von Krauss.
Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support.
Integrated assessment, 4(1):5–17, 2003.

Notions of regret

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

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

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

$\rightarrow$ linked to misspecified LRT: maximize the probability of keeping $\mathcal{H}_0$: θ valid instead of $\arg\max\mathcal{L}$.

PEI criterion

$Y \sim \mathrm{GP}(m_Y(\cdot), C_Y(\cdot,\cdot))$ on $\Theta \times \mathbb{U}$

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

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