Parameter control in the presence of uncertainties

Victor Trappler
Supervisors: Élise Arnaud, Laurent Debreu, Arthur Vidard

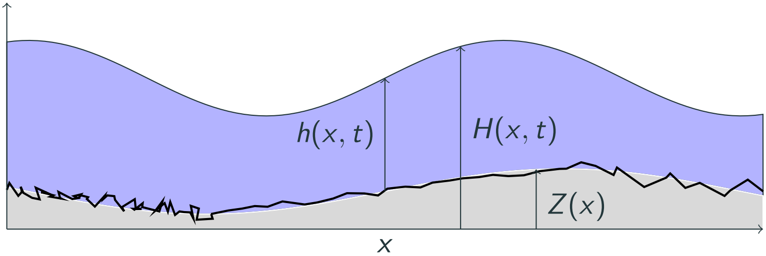
April 9, 2018

AIRSEA Research team (Inria)– Laboratoire Jean Kuntzmann

Introduction

Bottom friction

- The friction of the ocean bed has an influence on the water circulation
- Depends on the type and/or characteristic length of the asperities
- Subgrid phenomenon



Outline

Introduction

Deterministic problem

Dealing with uncertainties

Robust minimization

Surrogates and optimization

Conclusion

Deterministic problem

Computer code : the Shallow Water Equations

Input
- $\boldsymbol{k}$: Bottom friction (spatially distributed)
- $\boldsymbol{u}$: Environmental variables (fixed and known)

Output
- $W(\boldsymbol{k}) = \{W_i^n(\boldsymbol{k})\}_{i,n}$, where $W_i^n(\boldsymbol{k}) = [h_i^n(\boldsymbol{k}) \quad q_i^n(\boldsymbol{k})]^T$ for $0 \leq i \leq N_x$ and $0 \leq n \leq N_t$

Environmental variables

$\boldsymbol{u} \in \mathcal{U}$ fixed

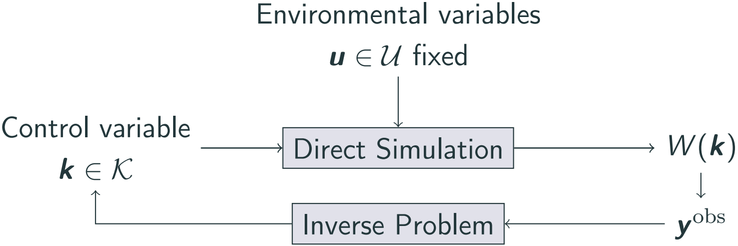
Control variable

$\boldsymbol{k} \in \mathcal{K}$

$$\boldsymbol{k} \longrightarrow \boxed{\text{Direct Simulation}} \longrightarrow W(\boldsymbol{k})$$

Computer code : the Shallow Water Equations

Input	• $\boldsymbol{k}$: Bottom friction (spatially distributed)
	• $\boldsymbol{u}$: Environmental variables (fixed and known)
Output	• $W(\boldsymbol{k}) = \{W_i^n(\boldsymbol{k})\}_{i,n}$, where $W_i^n(\boldsymbol{k}) = [h_i^n(\boldsymbol{k}) \quad q_i^n(\boldsymbol{k})]^T$ for $0 \leq i \leq N_x$ and $0 \leq n \leq N_t$



Data assimilation framework: Twin experiments

Let us set $\boldsymbol{k}_{\mathrm{ref}}$

We have $\boldsymbol{y}^{\mathrm{obs}} = M(\boldsymbol{k}_{\mathrm{ref}}) = \{h_i^n(\boldsymbol{k}_{\mathrm{ref}})\}_{i,n}$

$$j(\boldsymbol{k}) = \frac{1}{2}\|M(\boldsymbol{k}) - \boldsymbol{y}^{\mathrm{obs}}\|^2$$

Data assimilation framework: Twin experiments

Let us set $\boldsymbol{k}_{\mathrm{ref}}$

We have $\boldsymbol{y}^{\mathrm{obs}} = M(\boldsymbol{k}_{\mathrm{ref}}) = \{h_i^n(\boldsymbol{k}_{\mathrm{ref}})\}_{i,n}$

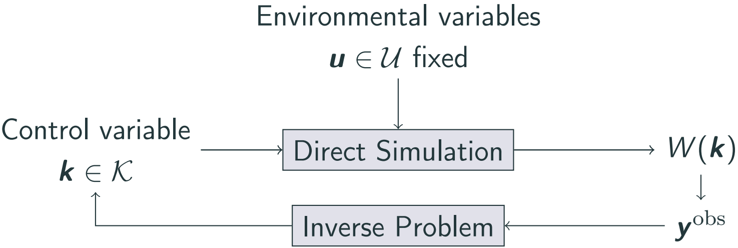
$$j(\boldsymbol{k}) = \frac{1}{2}\|M(\boldsymbol{k}) - \boldsymbol{y}^{\mathrm{obs}}\|^2$$

$$\arg\min_{\boldsymbol{k}\in\mathcal{K}} j(\boldsymbol{k})?$$

Dealing with uncertainties

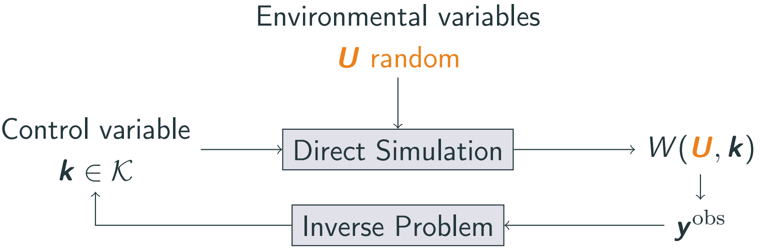
Introducing the uncertainties

Instead of considering $\boldsymbol{u}$ fixed, we consider that $\boldsymbol{U}$ is a random variable (pdf $\pi(\boldsymbol{u})$), and the output of the model depends on its realization.



Introducing the uncertainties

Instead of considering $\boldsymbol{u}$ fixed, we consider that $\boldsymbol{U}$ is a random variable (pdf $\pi(\boldsymbol{u})$), and the output of the model depends on its realization.



The cost function as a random variable

- Output of the computer code ($\boldsymbol{u}$ is an input):

$$M(\boldsymbol{k}) \quad \text{becomes} \quad M(\boldsymbol{k}, \boldsymbol{u})$$

- The (deterministic) quadratic error is now

$$j(\boldsymbol{k}, \boldsymbol{u}) = \frac{1}{2} \| M(\boldsymbol{k}, \boldsymbol{u}) - \boldsymbol{y}^{\text{obs}} \|^2$$

What to do with $j(\boldsymbol{k}, \boldsymbol{U})$ (random variable) ?

Variational approach or Bayesian approach ?

- **Variational**: Minimize a function of $j(\boldsymbol{k}, \boldsymbol{U})$,
  e.g. Minimize $\mathbb{E}[j(\boldsymbol{K}, \boldsymbol{U}) | \boldsymbol{K} = \boldsymbol{k}]$.
  $\longrightarrow$ Precise objective

Variational approach or Bayesian approach ?

- **Variational**: Minimize a function of $j(\boldsymbol{k}, \boldsymbol{U})$,
  e.g. Minimize $\mathbb{E}[j(\boldsymbol{K}, \boldsymbol{U}) | \boldsymbol{K} = \boldsymbol{k}]$.
  $\longrightarrow$ Precise objective
- **Bayesian**: let us assume $e^{-j(\boldsymbol{k},\boldsymbol{u})} \propto p(\boldsymbol{y}^{\mathrm{obs}} | \boldsymbol{k}, \boldsymbol{u})$
  Work around the likelihood and posterior distributions
  $p(\boldsymbol{k} | \boldsymbol{y}^{\mathrm{obs}})$
  $\longrightarrow$ More general method

Variational approach or Bayesian approach ?

- **Variational**: Minimize a function of $j(\boldsymbol{k}, \boldsymbol{U})$,
  e.g. Minimize $\mathbb{E}[j(\boldsymbol{K}, \boldsymbol{U}) | \boldsymbol{K} = \boldsymbol{k}]$.
  $\longrightarrow$ Precise objective
- **Bayesian**: let us assume $e^{-j(\boldsymbol{k},\boldsymbol{u})} \propto p(\boldsymbol{y}^{\mathrm{obs}} | \boldsymbol{k}, \boldsymbol{u})$
  Work around the likelihood and posterior distributions $p(\boldsymbol{k} | \boldsymbol{y}^{\mathrm{obs}})$
  $\longrightarrow$ More general method

But

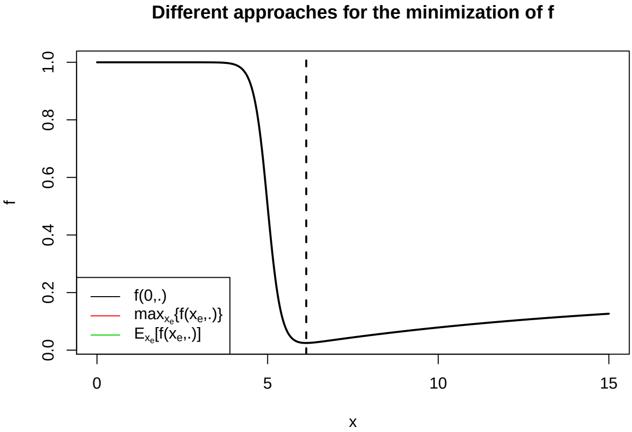
- Dependent on the efficiency of the statistical estimators
- Knowledge of $\boldsymbol{U}$? Assumptions on error ?
- Computational cost ?

Robust minimization

An illustration

$$(\boldsymbol{u}, \boldsymbol{k}) \mapsto f(\boldsymbol{u}, \boldsymbol{k}) = \tilde{f}(\boldsymbol{u} + \boldsymbol{k})$$

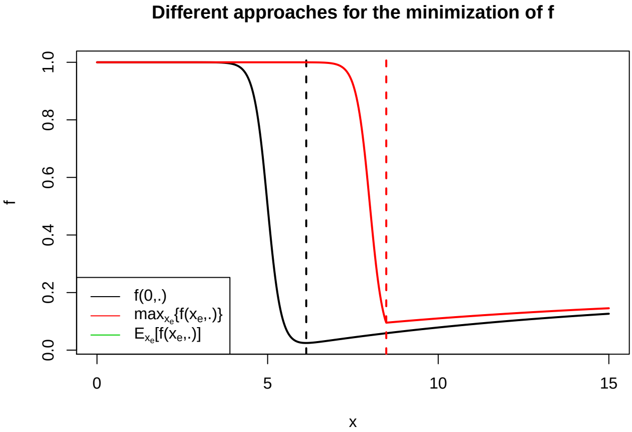
$\boldsymbol{U} \sim \mathcal{N}(0, s^2)$ truncated on $[-3; 3]$. Plot of $f(0, \cdot) = \tilde{f}(\cdot)$



An illustration

$$(\boldsymbol{u}, \boldsymbol{k}) \mapsto f(\boldsymbol{u}, \boldsymbol{k}) = \tilde{f}(\boldsymbol{u} + \boldsymbol{k})$$

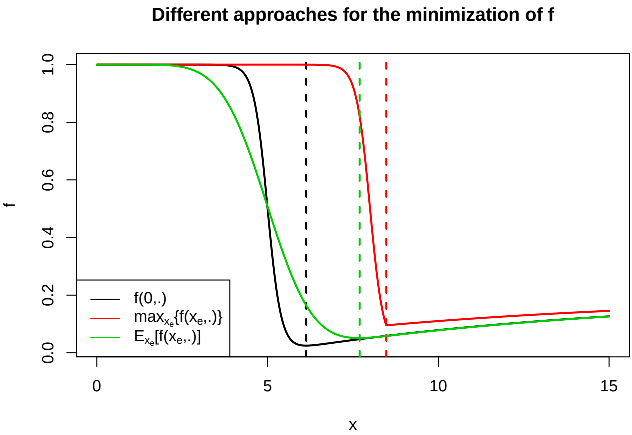
$\boldsymbol{U} \sim \mathcal{N}(0, s^2)$ truncated on $[-3; 3]$. Plot of $\max_{\boldsymbol{u}}\{f(\boldsymbol{u}, \cdot)\}$



An illustration

$$(\boldsymbol{u}, \boldsymbol{k}) \mapsto f(\boldsymbol{u}, \boldsymbol{k}) = \tilde{f}(\boldsymbol{u} + \boldsymbol{k})$$

$\boldsymbol{U} \sim \mathcal{N}(0, s^2)$ truncated on $[-3; 3]$. Plot of $\mathbb{E}_{\boldsymbol{u}}[f(\boldsymbol{u}, \cdot)]$



Non-exhaustive list of "Robust" Objectives

- Global Optimum: $\min_{(\boldsymbol{k},\boldsymbol{u})} j(\boldsymbol{u},\boldsymbol{k}) \longrightarrow$ EGO
- Worst case: $\min_{\boldsymbol{k}} \max_{\boldsymbol{u}} j(\boldsymbol{u},\boldsymbol{k}) \longrightarrow$ Explorative EGO
- M-robustness: $\min_{\boldsymbol{k}} \mathbb{E}\left[j(\boldsymbol{U},\boldsymbol{k})\right] \longrightarrow$ iterated LHS
- V-robustness: $\min_{\boldsymbol{k}} \mathbb{V}\mathrm{ar}\left[j(\boldsymbol{U},\boldsymbol{k})\right] \longrightarrow$ gradient-descent with PCE
- ρ-robustness: $\min \rho(j(\boldsymbol{U},\boldsymbol{k})) \longrightarrow$ gradient-descent with PCE
- Multiobjective: choice within Pareto frontier $\longrightarrow$ 1L/2L kriging

Bayesian approach

Let us suppose $\boldsymbol{K} \sim \pi(\boldsymbol{k})$.

Having observed $\boldsymbol{y}^{\mathrm{obs}}$, joint distribution of $(\boldsymbol{K}, \boldsymbol{U})$: $p(\boldsymbol{k}, \boldsymbol{u} | \boldsymbol{y}^{\mathrm{obs}})$?

Bayesian approach

Let us suppose $\boldsymbol{K} \sim \pi(\boldsymbol{k})$.

Having observed $\boldsymbol{y}^{\mathrm{obs}}$, joint distribution of $(\boldsymbol{K}, \boldsymbol{U})$: $p(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}})$?

Bayes' Theorem

$$
\begin{aligned}
p(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}}) &\propto p(\boldsymbol{y}^{\mathrm{obs}}|\boldsymbol{k}, \boldsymbol{u})\pi(\boldsymbol{k}, \boldsymbol{u}) \\
&\propto L(\boldsymbol{k}, \boldsymbol{u}; \boldsymbol{y}^{\mathrm{obs}})\pi(\boldsymbol{k})\pi(\boldsymbol{u})
\end{aligned}
$$

Bayesian approach

Let us suppose $\boldsymbol{K} \sim \pi(\boldsymbol{k})$.

Having observed $\boldsymbol{y}^{\mathrm{obs}}$, joint distribution of $(\boldsymbol{K}, \boldsymbol{U})$: $p(\boldsymbol{k}, \boldsymbol{u} | \boldsymbol{y}^{\mathrm{obs}})$?

Bayes' Theorem

$$\begin{aligned} p(\boldsymbol{k}, \boldsymbol{u} | \boldsymbol{y}^{\mathrm{obs}}) &\propto p(\boldsymbol{y}^{\mathrm{obs}} | \boldsymbol{k}, \boldsymbol{u}) \pi(\boldsymbol{k}, \boldsymbol{u}) \\ &\propto L(\boldsymbol{k}, \boldsymbol{u}; \boldsymbol{y}^{\mathrm{obs}}) \pi(\boldsymbol{k}) \pi(\boldsymbol{u}) \end{aligned}$$

Link with cost function j : Squared error $\leftrightarrow$ Gaussian errors

$$L(\boldsymbol{k}, \boldsymbol{u}; \boldsymbol{y}^{\mathrm{obs}}) \propto \exp\left[-\frac{1}{2} \| M(\boldsymbol{k}, \boldsymbol{u}) - \boldsymbol{y}^{\mathrm{obs}} \|^2_{\boldsymbol{\Sigma}^{-1}}\right] = \exp\left[-j(\boldsymbol{k}, \boldsymbol{u})\right]$$

Bayesian Quantities of interest

Bayes' theorem

$$p(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}}) \propto L(\boldsymbol{k}, \boldsymbol{u}; \boldsymbol{y}^{\mathrm{obs}})\pi(\boldsymbol{k})\pi(\boldsymbol{u}) \propto p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{u})\pi(\boldsymbol{u})$$

$$\begin{aligned}
\text{ML}: &\quad \operatorname*{arg\,max}_{(\boldsymbol{k},\boldsymbol{u})} L(\boldsymbol{k}, \boldsymbol{u}; \boldsymbol{y}^{\mathrm{obs}}) \\
\text{MAP}: &\quad \operatorname*{arg\,max}_{(\boldsymbol{k},\boldsymbol{u})} p(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}}) = L(\boldsymbol{k}, \boldsymbol{u}; \boldsymbol{y}^{\mathrm{obs}})\pi(\boldsymbol{k})\pi(\boldsymbol{u}) \\
\text{MMAP}: &\quad \operatorname*{arg\,max}_{\boldsymbol{k}} p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}) = \int_{\mathcal{U}} p(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}})\,\mathrm{d}\boldsymbol{u} \\
\text{Min of variance}: &\quad \operatorname*{arg\,min}_{\boldsymbol{k}} \mathbb{V}\mathrm{ar}_U\left[p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{U})\right] \\
\text{Worst Case}: &\quad \operatorname*{arg\,max}_{\boldsymbol{k}}\{\min_{\boldsymbol{u}} p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{u})\} \\
\text{MPE}: &\quad \text{Mode of } \boldsymbol{K}_{\arg\max} = \operatorname*{arg\,max}_{\boldsymbol{k}} p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{U})
\end{aligned}$$

Bayesian Quantities of interest

Bayes' theorem

$$p(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}}) \propto L(\boldsymbol{k}, \boldsymbol{u}; \boldsymbol{y}^{\mathrm{obs}})\pi(\boldsymbol{k})\pi(\boldsymbol{u}) \propto p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{u})\pi(\boldsymbol{u})$$

$$\begin{aligned}
\text{ML}:&\quad \operatorname*{arg\,max}_{(\boldsymbol{k},\boldsymbol{u})} L(\boldsymbol{k}, \boldsymbol{u}; \boldsymbol{y}^{\mathrm{obs}}) \\
\text{MAP}:&\quad \operatorname*{arg\,max}_{(\boldsymbol{k},\boldsymbol{u})} p(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}}) = L(\boldsymbol{k}, \boldsymbol{u}; \boldsymbol{y}^{\mathrm{obs}})\pi(\boldsymbol{k})\pi(\boldsymbol{u}) \\
\text{MMAP}:&\quad \operatorname*{arg\,max}_{\boldsymbol{k}} p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}) = \int_{\mathcal{U}} p(\boldsymbol{k}, \boldsymbol{u}|\boldsymbol{y}^{\mathrm{obs}})\,\mathrm{d}\boldsymbol{u} \\
\text{Min of variance}:&\quad \operatorname*{arg\,min}_{\boldsymbol{k}} \mathbb{V}\mathrm{ar}_{U}\left[p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{U})\right] \\
\text{Worst Case}:&\quad \operatorname*{arg\,max}_{\boldsymbol{k}}\{\min_{\boldsymbol{u}} p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{u})\} \\
\text{MPE}:&\quad \text{Mode of } \boldsymbol{K}_{\arg\max} = \operatorname*{arg\,max}_{\boldsymbol{k}} p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{U})
\end{aligned}$$

Illustration on the SWE

Family of densities: $\{p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{u}); \boldsymbol{u} \in \mathcal{U}\}$

MMAP:

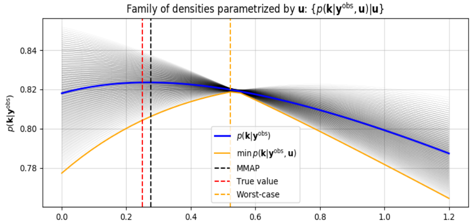
$\arg\max_{\boldsymbol{k}} p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}})$

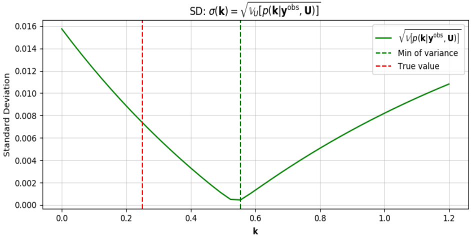
Min Var:

$\arg\min_{\boldsymbol{k}} \mathbb{V}\mathrm{ar}_U \left[p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{U})\right]$

Worst case:

$\arg\max_{\boldsymbol{k}}\{\min_{\boldsymbol{u}} p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{u})\}$





"Most Probable Estimate"

$\boldsymbol{K}_{\arg\max} = \arg\max_{\boldsymbol{k}\in\mathcal{K}} p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{U})$ random variable
$\longrightarrow$ estimate its density (how often is the value $\boldsymbol{k}$ a maximizer)

Straightforward algorithm:

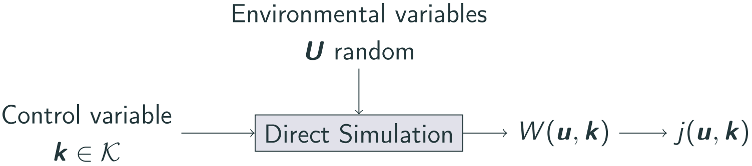
- For $i = 1 \dots N$:
  - Sample $\boldsymbol{u}^{(i)}$ from $\pi(\boldsymbol{u})$ / Adapted space-filling designs
  - Maximize $p(\boldsymbol{k}|\boldsymbol{y}^{\mathrm{obs}}, \boldsymbol{u}^{(i)})$ yielding $\boldsymbol{k}^{(i)}_{\arg\max}$ (adjoint method)
- Estimate density (KDE) / Mode

Illustration of MPE

Surrogates and optimization

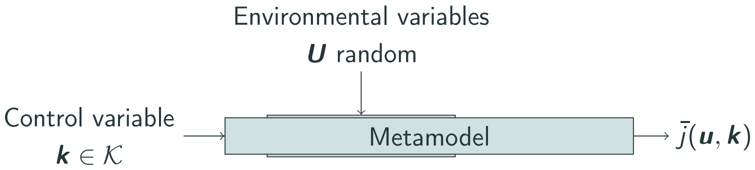
Why surrogates?

- Computer model: expensive to run
- $\dim \mathcal{K}$, $\dim \mathcal{U}$ can be very large: curse of dimensionality
- Uncertainties upon $\boldsymbol{u}$ directly in the surrogate



Why surrogates?

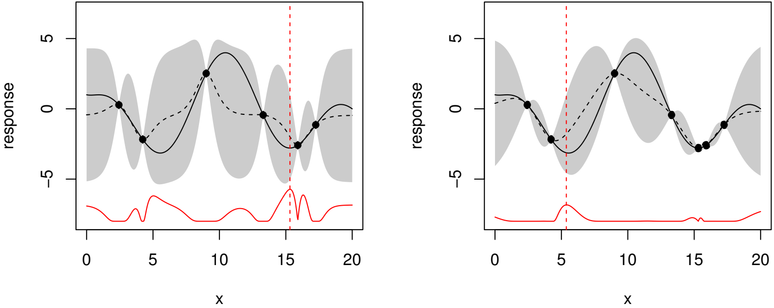
- Computer model: expensive to run
- $\dim \mathcal{K}$, $\dim \mathcal{U}$ can be very large: curse of dimensionality
- Uncertainties upon $\boldsymbol{u}$ directly in the surrogate



Based on kriging model (=Gaussian Process Regression) $\longrightarrow$ mean and variance

How to choose a new point to evaluate ? Criterion $\kappa(\boldsymbol{x}) \longrightarrow$ "potential" of the point

$$\boldsymbol{x}_{\text{new}} = \arg\max \kappa(\boldsymbol{x})$$



Conclusion

Conclusion

Wrapping up

- Variational and bayesian approaches for this inverse problem results in different methods
- In both case, strategies rely heavily on surrogate models $\longrightarrow$ Kriging, Polynomial chaos

Perspective and future work

- Cost of computer evaluations $\rightarrow$ limit the total number of runs
- Dimensionality of the input space $\rightarrow$ reduction of the input space ?
- How to deal with uncontrollable errors $\rightarrow$ errors between model and reality ?