Robust Estimation of bottom friction

Parameter control in the presence of uncertainties

VICTOR TRAPPLER

s151431

Master Thesis Defence

DTU Aqua

National Institute of Aquatic Resources

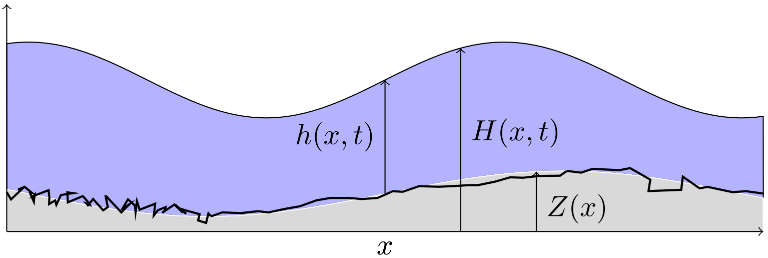
The AIRSEA team, Grenoble, FRANCE

The AIRSEA team, Grenoble, FRANCE

- Modelling Oceanic and Atmospheric flows: parametrization and coupling of the equations
- Model reduction, multiscale algorithms
- High-performance computing
- Dealing with uncertainties

The AIRSEA team, Grenoble, FRANCE

- Modelling Oceanic and Atmospheric flows: parametrization and coupling of the equations
- Model reduction, multiscale algorithms
- High-performance computing
- Dealing with uncertainties


$h(x,t)$
$H(x,t)$
$Z(x)$
x

Outline

Deterministic Framework

- Introduction
- Deterministic Framework
  - The 1D Shallow Water Equations
  - Adjoint-based optimization
- Global sensitivity analysis
- Robust Optimization
- Conclusion

The 1D Shallow Water Equations

1D-SWE

$$\partial_t h + \partial_x q = 0 \qquad \text{(Conservation)}$$

$$\partial_t q + \partial_x \left(\frac{q^2}{h} + \frac{1}{2} g h^2 \right) = -g h \partial_x Z - S \qquad \text{(Momentum)}$$

The 1D Shallow Water Equations

1D-SWE

$$\partial_t h + \partial_x q = 0 \qquad \text{(Conservation)}$$

$$\partial_t q + \partial_x \left(\frac{q^2}{h} + \frac{1}{2} g h^2 \right) = -gh\partial_x Z - S \qquad \text{(Momentum)}$$

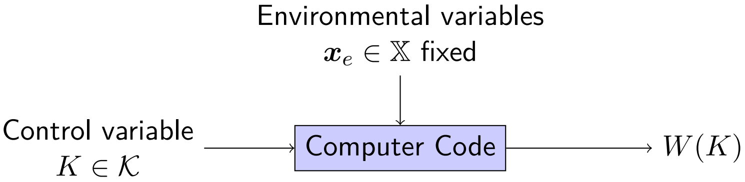
Quadratic Friction

$$S = -K|q|qh^{-\eta}, \quad \eta = 7/3$$

K: control parameter. Either a scalar value or a vector

Computer code

- 1D Shallow water equations
  - K: Bottom friction
  - Boundary conditions (considered fixed and known)
- Output $W(K)$:
  $W_i^n(K) = [h_i^n(K) \quad q_i^n(K)]^T$, for $0 \leq i \leq N_x$ and $0 \leq n \leq N_t$



Data assimilation

K_{ref} and $\mathcal{H}$ observation operator
We have $Y = \mathcal{H}W(K_{\text{ref}}) = \{h_i^n(K_{\text{ref}})\}_{i,n}$

$$j(K) = \frac{1}{2}\|\mathcal{H}W(K) - Y\|^2$$

Data assimilation

K_{ref} and $\mathcal{H}$ observation operator
We have $Y = \mathcal{H}W(K_{\text{ref}}) = \{h_i^n(K_{\text{ref}})\}_{i,n}$

$$j(K) = \frac{1}{2}\|\mathcal{H}W(K) - Y\|^2$$

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

- Gradient-free: Simulated annealing, Nelder-mead,... $\rightarrow$ High number of runs. Very expensive in practice

Data assimilation

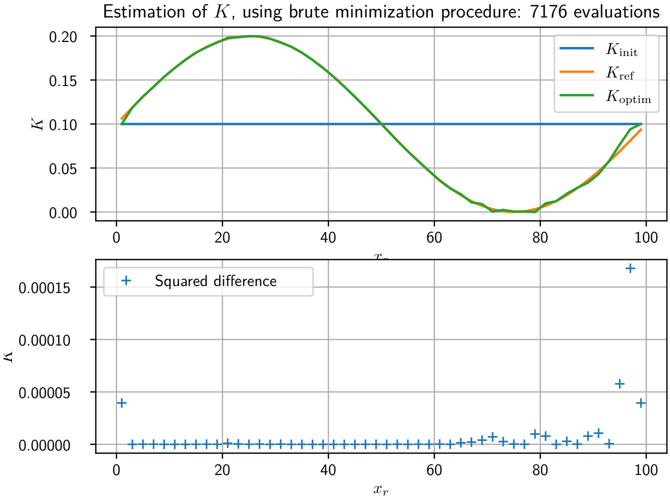
K_{ref} and $\mathcal{H}$ observation operator
We have $Y = \mathcal{H}W(K_{\mathrm{ref}}) = \{h_i^n(K_{\mathrm{ref}})\}_{i,n}$

$$j(K) = \frac{1}{2}\|\mathcal{H}W(K) - Y\|^2$$

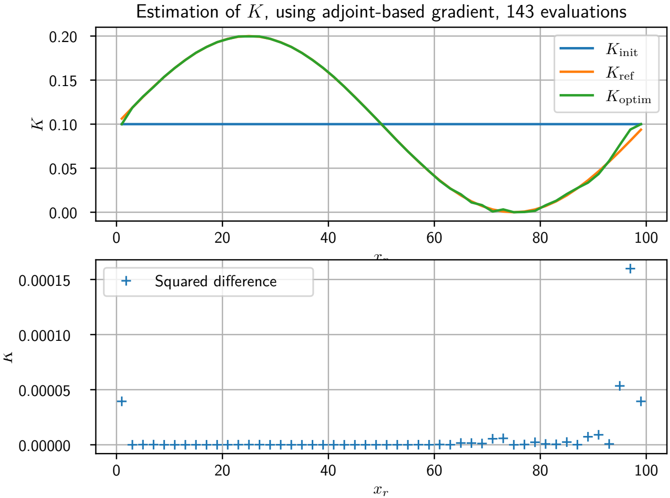
$$\underset{K\in\mathcal{K}}{\arg\min}\, j(K)?$$

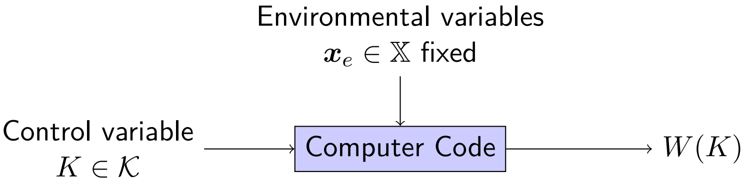
- Gradient-free: Simulated annealing, Nelder-mead,... $\rightarrow$ High number of runs. Very expensive in practice
- Gradient-based: gradient-descent, (quasi-) Newton method ... $\rightarrow$ Less number of runs, but need to derive adjoint code

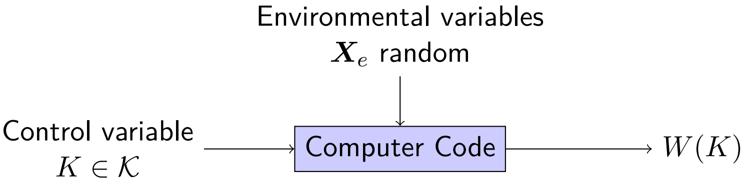
Estimation procedure no gradient



Estimation procedure adjoint-based gradient




Environmental variables
$\boldsymbol{x}_e \in \mathbb{X}$ fixed
Control variable
$K \in \mathcal{K}$
Computer Code
$W(K)$


Environmental variables
$\boldsymbol{X}_e$ random
Control variable
$K \in \mathcal{K}$
Computer Code
$W(K)$

Introducing uncertainties

$\boldsymbol{X}_e$ random vector with realizations $\boldsymbol{x}_e \in \mathbb{X}$

Variable	`mean.h`	`ampli`	`period`	`phase`
$\boldsymbol{X}_e$	$\mathcal{U}([19.5, 20.5])$	$\mathcal{U}([4.9, 5.1])$	$\mathcal{U}([49.9, 50.1])$	$\mathcal{U}([-0.001, 0.001])$
$\boldsymbol{x}_{e,\mathrm{ref}}$	20.0	5.0	50.0	0.000

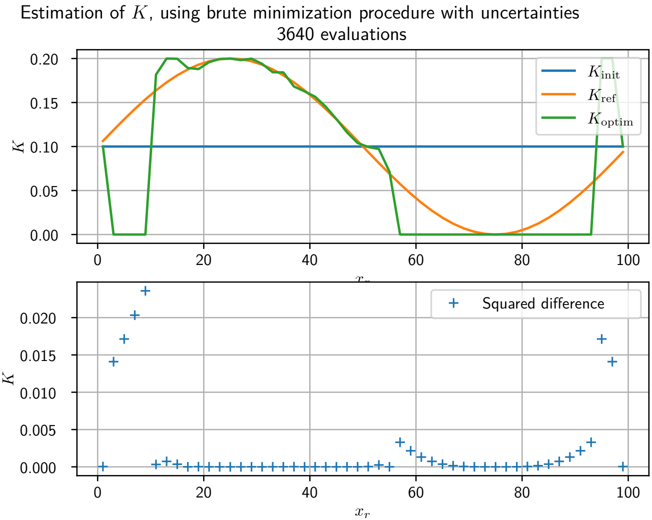
$$W(K) \quad \text{becomes} \quad W(\boldsymbol{x}_e, K)$$

We have $Y = \mathcal{H}W(\boldsymbol{x}_{e,\mathrm{ref}}, K_{\mathrm{ref}})$
The (deterministic) quadratic error is now

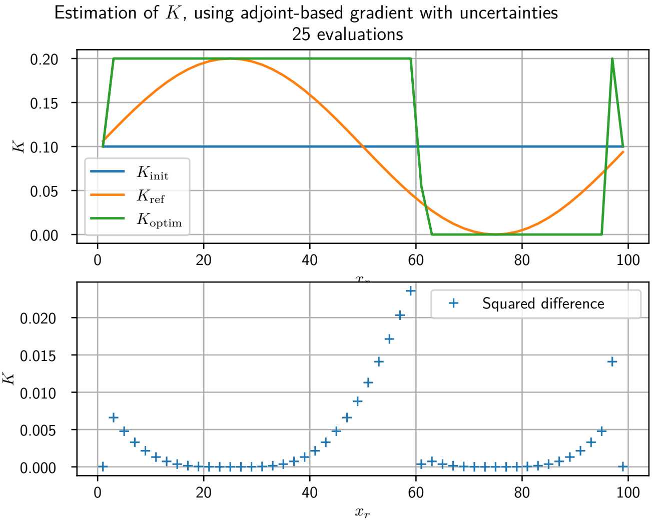
$$j(\boldsymbol{x}_e, K) = \frac{1}{2}\|\mathcal{H}W(\boldsymbol{x}_e, K) - Y\|^2$$

$\longrightarrow$ sample one $\boldsymbol{x}_e$ and min w.r.t. K ?

Estimation procedure with uncertainties no gradient



Estimation procedure with uncertainties adjoint-based gradient



Influence of $\boldsymbol{X}_e$?
Minimizing $j(\boldsymbol{x}_e, K)$ wrt K ?
Computational cost ?

Influence of $\boldsymbol{X}_e$?	$\longrightarrow$ Sensitivity analysis
Minimizing $j(\boldsymbol{x}_e, K)$ wrt K ?	$\longrightarrow$ Robust optimization
Computational cost ?	$\longrightarrow$ Use of surrogate

Sensitivity Analysis

Sobol Indices [Sobol, 2001]

Let $\mathcal{J} = J(\boldsymbol{X})$ a rv, with $\boldsymbol{X} = (X_1, \dots X_p)$ uniformly distributed on $[0;1]^p$ and components independent.

$$\mathbb{E}[\mathcal{J}|X_i = \alpha]$$

Sobol Indices [Sobol, 2001]

Let $\mathcal{J} = J(\boldsymbol{X})$ a rv, with $\boldsymbol{X} = (X_1, \dots X_p)$ uniformly distributed on $[0;1]^p$ and components independent.

$$\mathbb{V}\mathrm{ar}[\mathbb{E}[\mathcal{J}|X_i]]$$

Sobol Indices [Sobol, 2001]

Let $\mathcal{J} = J(\boldsymbol{X})$ a rv, with $\boldsymbol{X} = (X_1, \dots X_p)$ uniformly distributed on $[0;1]^p$ and components independent.

$$S_i = \frac{\mathbb{V}\mathrm{ar}[\mathbb{E}[\mathcal{J}|X_i]]}{\mathbb{V}\mathrm{ar}[\mathcal{J}]}$$

Sobol Indices [Sobol, 2001]

Let $\mathcal{J} = J(\boldsymbol{X})$ a rv, with $\boldsymbol{X} = (X_1, \dots X_p)$ uniformly distributed on $[0;1]^p$ and components independent.

$$S_i = \frac{\mathbb{V}\mathrm{ar}[\mathbb{E}[\mathcal{J}|X_i]]}{\mathbb{V}\mathrm{ar}[\mathcal{J}]}$$

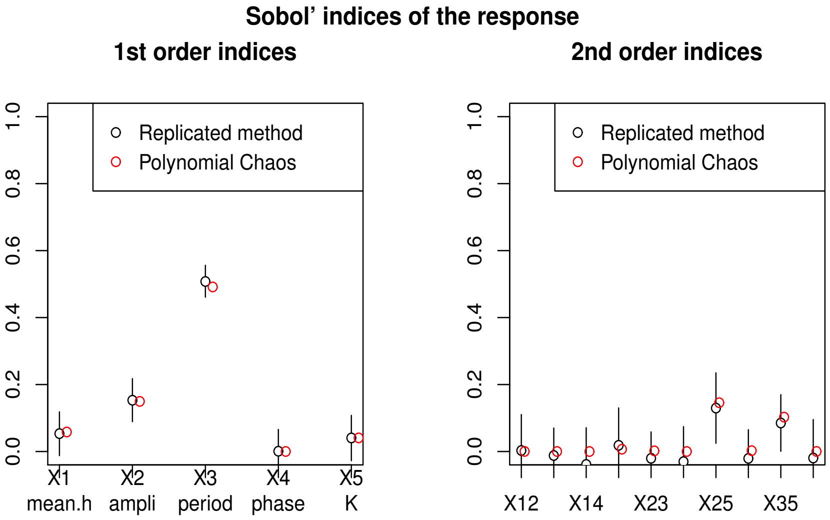
Variance of the ANOVA decomposition

$$1 = \underbrace{\sum_{i=1}^{p} S_i}_{\text{Single variable influence}} + \underbrace{\sum_{1\leq i<j\leq p} S_{ij}}_{\text{Interactions order 2}} + \cdots + \underbrace{S_{1\dots p}}_{\text{Interaction order } p}$$

[Gilquin et al., 2017, Sudret, 2015]

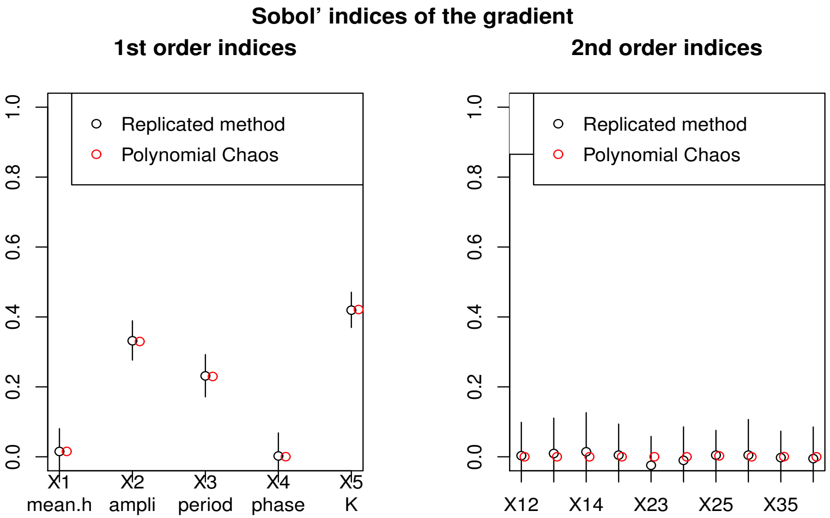
Sobol' indices of the cost function

$j(\boldsymbol{X}_e, K)$



Sobol' indices of the gradient of cost function

$\frac{\mathrm{d}j}{\mathrm{d}K}(\boldsymbol{X}_e, K)$

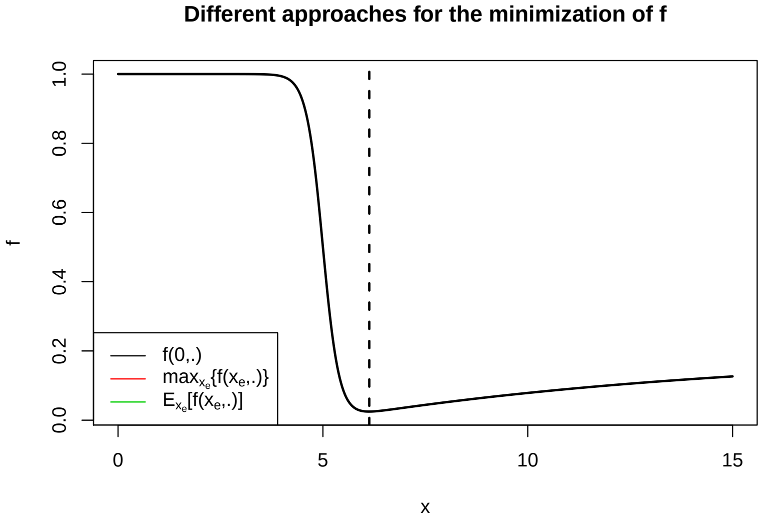


Robust Optimization

- Introduction
- Deterministic Framework
- Global sensitivity analysis
- Robust Optimization
  - Concepts of robustness
  - Metamodeling
  - Adaptative sampling
- Conclusion

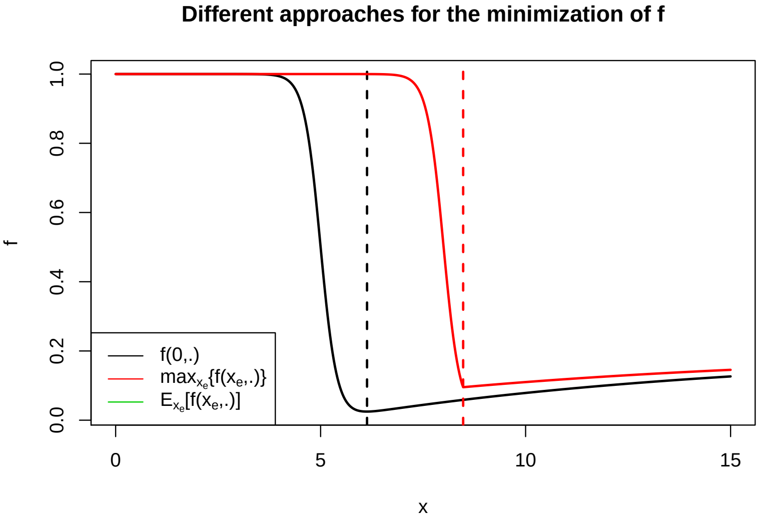
A first example

$(x_e, K) \mapsto f(x_e, K) = \tilde{f}(x_e + K)$ and $X_e \sim \mathcal{N}(0, s^2)$ truncated on $[-3; 3]$



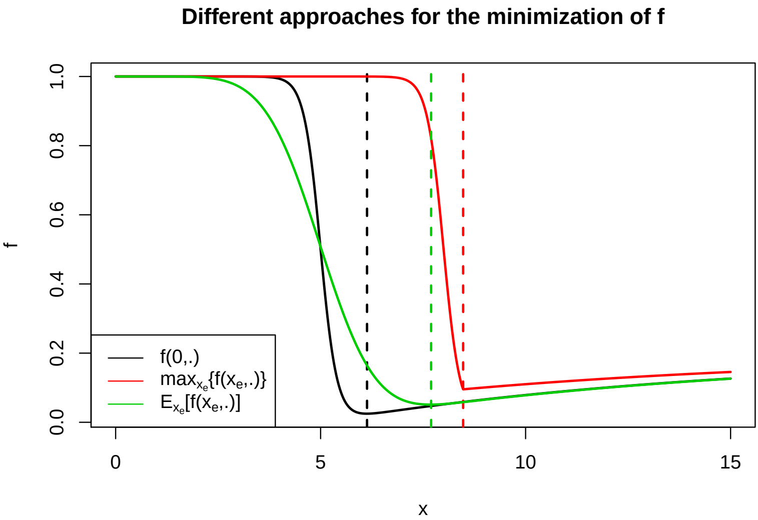
A first example

$(x_e, K) \mapsto f(x_e, K) = \tilde{f}(x_e + K)$ and $X_e \sim \mathcal{N}(0, s^2)$ truncated on $[-3; 3]$



A first example

$(x_e, K) \mapsto f(x_e, K) = \tilde{f}(x_e + K)$ and $X_e \sim \mathcal{N}(0, s^2)$ truncated on $[-3; 3]$



Different Notions of robustness

- Global Optimum: $\min j(\boldsymbol{x}_e, K) \longrightarrow$ EGO

Different Notions of robustness

- Global Optimum: $\min j(\boldsymbol{x}_e, K) \longrightarrow$ EGO
- Worst case: $\min_K \max_{\boldsymbol{x}_e} j(\boldsymbol{x}_e, K) \longrightarrow$ Explorative EGO

Different Notions of robustness

- Global Optimum: $\min j(\boldsymbol{x}_e, K) \longrightarrow$ EGO
- Worst case: $\min_K \max_{\boldsymbol{x}_e} j(\boldsymbol{x}_e, K) \longrightarrow$ Explorative EGO
- M-robustness: $\min \mu(K),$ constraint on $\sigma^2(K) \longrightarrow$ iterated LHS

Different Notions of robustness

- Global Optimum: $\min j(\boldsymbol{x}_e, K) \longrightarrow$ EGO
- Worst case: $\min_K \max_{\boldsymbol{x}_e} j(\boldsymbol{x}_e, K) \longrightarrow$ Explorative EGO
- M-robustness: $\min \mu(K), \quad$ constraint on $\sigma^2(K) \longrightarrow$ iterated LHS
- V-robustness: $\min \sigma^2(K), \quad$ constraint on $\mu(K) \longrightarrow$ gradient-descent with PCE

Different Notions of robustness

- Global Optimum: $\min j(\boldsymbol{x}_e, K) \longrightarrow$ EGO
- Worst case: $\min_K \max_{\boldsymbol{x}_e} j(\boldsymbol{x}_e, K) \longrightarrow$ Explorative EGO
- M-robustness: $\min \mu(K), \quad$ constraint on $\sigma^2(K) \longrightarrow$ iterated LHS
- V-robustness: $\min \sigma^2(K), \quad$ constraint on $\mu(K) \longrightarrow$ gradient-descent with PCE
- ρ-robustness: $\min \rho(\mu(K), \sigma^2(K)) \longrightarrow$ gradient-descent with PCE

Different Notions of robustness

- Global Optimum: $\min j(\boldsymbol{x}_e, K) \longrightarrow$ EGO
- Worst case: $\min_K \max_{\boldsymbol{x}_e} j(\boldsymbol{x}_e, K) \longrightarrow$ Explorative EGO
- M-robustness: $\min \mu(K),\quad$ constraint on $\sigma^2(K) \longrightarrow$ iterated LHS
- V-robustness: $\min \sigma^2(K),\quad$ constraint on $\mu(K) \longrightarrow$ gradient-descent with PCE
- ρ-robustness: $\min \rho(\mu(K), \sigma^2(K)) \longrightarrow$ gradient-descent with PCE
- Multiobjective: choice within Pareto frontier $\longrightarrow$ 1L/2L kriging

Different Notions of robustness

- Global Optimum: $\min j(\boldsymbol{x}_e, K) \longrightarrow$ EGO
- Worst case: $\min_K \max_{\boldsymbol{x}_e} j(\boldsymbol{x}_e, K) \longrightarrow$ Explorative EGO
- M-robustness: $\min \mu(K), \quad$ constraint on $\sigma^2(K) \longrightarrow$ iterated LHS
- V-robustness: $\min \sigma^2(K), \quad$ constraint on $\mu(K) \longrightarrow$ gradient-descent with PCE
- ρ-robustness: $\min \rho(\mu(K), \sigma^2(K)) \longrightarrow$ gradient-descent with PCE
- Multiobjective: choice within Pareto frontier $\longrightarrow$ 1L/2L kriging

Why surrogates ?

- Model $\longrightarrow$ expensive to run
- High dimensional problem + taking into account uncertainties ?

Environmental variables

$\boldsymbol{X}_e$ random

Control variable

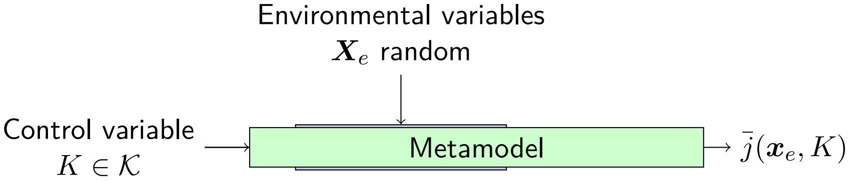
$K \in \mathcal{K}$

Computer Code

$W(\boldsymbol{x}_e, K) \longrightarrow j(\boldsymbol{x}_e, K)$

Why surrogates ?

- Model $\longrightarrow$ expensive to run
- High dimensional problem + taking into account uncertainties ?



Comparison between PCE and Kriging

	Polynomial Chaos	Kriging
Surrogate	$J(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \hat{J}_{\boldsymbol{\alpha}} \boldsymbol{\Phi}_{\boldsymbol{\alpha}}(\boldsymbol{X})$	$\bar{J}(\boldsymbol{x}) \sim \mathcal{N}(\hat{m}(\boldsymbol{x}), \hat{s}^2(\boldsymbol{x}))$
Estim.	Numerical quadrature/Regression	Regression
Quantity	Statistical moments	estimate + CI
Ref.	[Wiener, 1938, Sudret, 2015]	[Krige, 1951, Matheron, 1969]

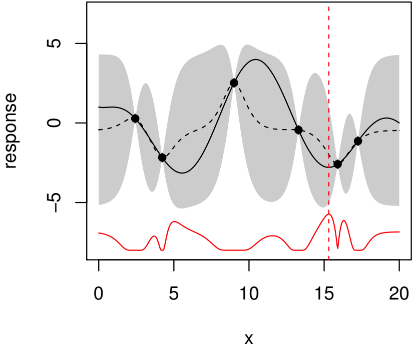
Comparison between PCE and Kriging

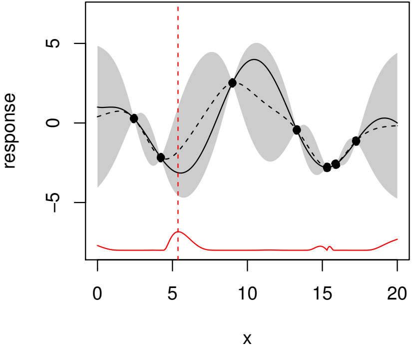
	Polynomial Chaos	Kriging
Surrogate	$J(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \hat{J}_{\boldsymbol{\alpha}} \boldsymbol{\Phi}_{\boldsymbol{\alpha}}(\boldsymbol{X})$	$\bar{J}(\boldsymbol{x}) \sim \mathcal{N}(\hat{m}(\boldsymbol{x}), \hat{s}^2(\boldsymbol{x}))$
Estim.	Numerical quadrature/Regression	Regression
Quantity	Statistical moments	estimate + CI
Ref.	[Wiener, 1938, Sudret, 2015]	[Krige, 1951, Matheron, 1969]

Principle of adaptative sampling

Based on kriging model $\longrightarrow$ mean and variance
How to choose a new point to evaluate ? Criterion $\kappa(\boldsymbol{x}) \longrightarrow$ "potential" of the point

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





EGO [Jones et al., 1998]

Global Optimum

$\mathcal{P}_N$ experimental design on $\mathbb{X} \times \mathcal{K}$, $\mathcal{Y}_N = j(\mathcal{P}_N)$.

$$\bar{J}_N(\boldsymbol{x}_e, K) \sim \mathcal{N}\left(\hat{m}_N(\boldsymbol{x}_e, K), \hat{s}_N^2(\boldsymbol{x}_e, K)\right)$$

EGO [Jones et al., 1998]

Global Optimum

$\mathcal{P}_N$ experimental design on $\mathbb{X} \times \mathcal{K}$, $\mathcal{Y}_N = j(\mathcal{P}_N)$.

$$\bar{J}_N(\boldsymbol{x}_e, K) \sim \mathcal{N}\left(\hat{m}_N(\boldsymbol{x}_e, K), \hat{s}_N^2(\boldsymbol{x}_e, K)\right)$$

$$j_{\min}^N = \min \mathcal{Y}_N$$

Expected improvement

$$j_{\min}^N - \bar{J}_N(\boldsymbol{x}_e, K)$$

EGO [Jones et al., 1998]

Global Optimum

$\mathcal{P}_N$ experimental design on $\mathbb{X} \times \mathcal{K}$, $\mathcal{Y}_N = j(\mathcal{P}_N)$.

$$\bar{J}_N(\boldsymbol{x}_e, K) \sim \mathcal{N}\left(\hat{m}_N(\boldsymbol{x}_e, K), \hat{s}_N^2(\boldsymbol{x}_e, K)\right)$$

$$j_{\min}^N = \min \mathcal{Y}_N$$

Expected improvement

$$\max\{0, j_{\min}^N - \bar{J}_N(\boldsymbol{x}_e, K)\}$$

EGO [Jones et al., 1998]
Global Optimum

$\mathcal{P}_N$ experimental design on $\mathbb{X} \times \mathcal{K}$, $\mathcal{Y}_N = j(\mathcal{P}_N)$.

$$\bar{J}_N(\boldsymbol{x}_e, K) \sim \mathcal{N}\left(\hat{m}_N(\boldsymbol{x}_e, K), \hat{s}_N^2(\boldsymbol{x}_e, K)\right)$$

$$j_{\min}^N = \min \mathcal{Y}_N$$

Expected improvement

$$EI(\boldsymbol{x}_e, K) = \mathbb{E}[\max\{0, j_{\min}^N - \bar{J}_N(\boldsymbol{x}_e, K)\}]$$

EGO [Jones et al., 1998]

Global Optimum

$\mathcal{P}_N$ experimental design on $\mathbb{X} \times \mathcal{K}$, $\mathcal{Y}_N = j(\mathcal{P}_N)$.

$$\bar{J}_N(\boldsymbol{x}_e, K) \sim \mathcal{N}\left(\hat{m}_N(\boldsymbol{x}_e, K), \hat{s}_N^2(\boldsymbol{x}_e, K)\right)$$

$$j_{\min}^N = \min \mathcal{Y}_N$$

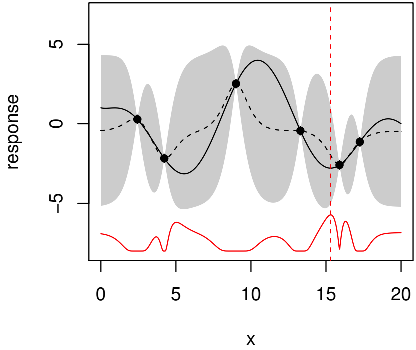
Expected improvement

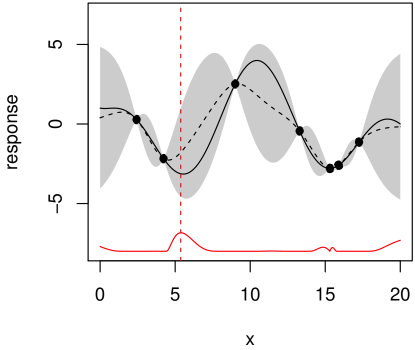
$$EI(\boldsymbol{x}_e, K) = \mathbb{E}[\max\{0, j_{\min}^N - \bar{J}_N(\boldsymbol{x}_e, K)\}]$$

EGO iteration

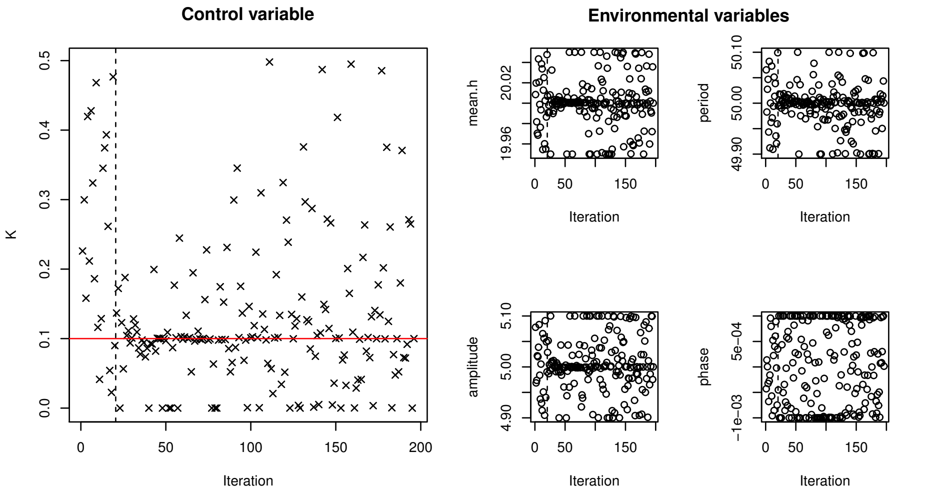
$$\begin{aligned} (\tilde{\boldsymbol{x}}_e, \tilde{K}) &= \arg\max EI(\boldsymbol{x}_e, K) \\ \mathcal{P}_{N+1} &= \mathcal{P}_N \cup (\tilde{\boldsymbol{x}}_e, \tilde{K}) \\ \bar{J}_{N+1}(\boldsymbol{x}_e, K) &\sim \mathcal{N}\left(\hat{m}_{N+1}(\boldsymbol{x}_e, K), \hat{s}_{N+1}^2(\boldsymbol{x}_e, K)\right) \end{aligned}$$

Example of an EGO iteration





EGO


Control variable
Environmental variables
K
0.0
0.1
0.2
0.3
0.4
0.5
0
50
100
150
200
Iteration
mean.h
19.96
20.02
period
49.90
50.00
50.10
amplitude
4.90
5.00
5.10
phase
−1e−03
5e−04
0
50
150
Iteration

Explorative EGO [Lehman et al., 2004]

Worst-case scenario

$\mathcal{P}_N$ experimental design on $\mathbb{X} \times \mathcal{K}$, $\mathcal{Y}_N = j(\mathcal{P}_N)$.

$$\bar{J}_N(\boldsymbol{x}_e, K) \sim \mathcal{N}\left(\hat{m}_N(\boldsymbol{x}_e, K), \hat{s}_N^2(\boldsymbol{x}_e, K)\right)$$

$j_{\min}^N = \min \mathcal{Y}_N$

Expected improvement

$$EI(\boldsymbol{x}_e, K) = \mathbb{E}[\max\{0, j_{\min}^N - \bar{J}_N(\boldsymbol{x}_e, K)\}]$$

Explorative EGO [Lehman et al., 2004]

Worst-case scenario

$\mathcal{P}_N$ experimental design on $\mathbb{X} \times \mathcal{K}$, $\mathcal{Y}_N = j(\mathcal{P}_N)$.

$$\bar{J}_N(\boldsymbol{x}_e, K) \sim \mathcal{N}\left(\hat{m}_N(\boldsymbol{x}_e, K), \hat{s}_N^2(\boldsymbol{x}_e, K)\right)$$

$$j_{\min}^N = \min \mathcal{Y}_N$$

Expected improvement

$$EI(\boldsymbol{x}_e, K) = \mathbb{E}[\max\{0, j_{\min}^N - \bar{J}_N(\boldsymbol{x}_e, K)\}]$$

Explorative EGO iteration

$$\begin{aligned}
(\tilde{\boldsymbol{x}}_e, \tilde{K}) &= \arg\max EI(\boldsymbol{x}_e, K) \\
\boldsymbol{x}_e^* &= \arg\max_{\boldsymbol{x}_e} d\left((\boldsymbol{x}_e, \tilde{K}), \mathcal{P}_N\right) \\
\mathcal{P}_{N+1} &= \mathcal{P}_N \cup (\boldsymbol{x}_e^*, \tilde{K}) \\
\bar{J}_{N+1}(\boldsymbol{x}_e, K) &\sim \mathcal{N}\left(\hat{m}_{N+1}(\boldsymbol{x}_e, K), \hat{s}_{N+1}^2(\boldsymbol{x}_e, K)\right)
\end{aligned}$$

Explorative EGO [Lehman et al., 2004]

Worst-case scenario

$\mathcal{P}_N$ experimental design on $\mathbb{X} \times \mathcal{K}$, $\mathcal{Y}_N = j(\mathcal{P}_N)$.

$$\bar{J}_N(\boldsymbol{x}_e, K) \sim \mathcal{N}\left(\hat{m}_N(\boldsymbol{x}_e, K), \hat{s}^2_N(\boldsymbol{x}_e, K)\right)$$

$j^N_{\min} = \min \mathcal{Y}_N$

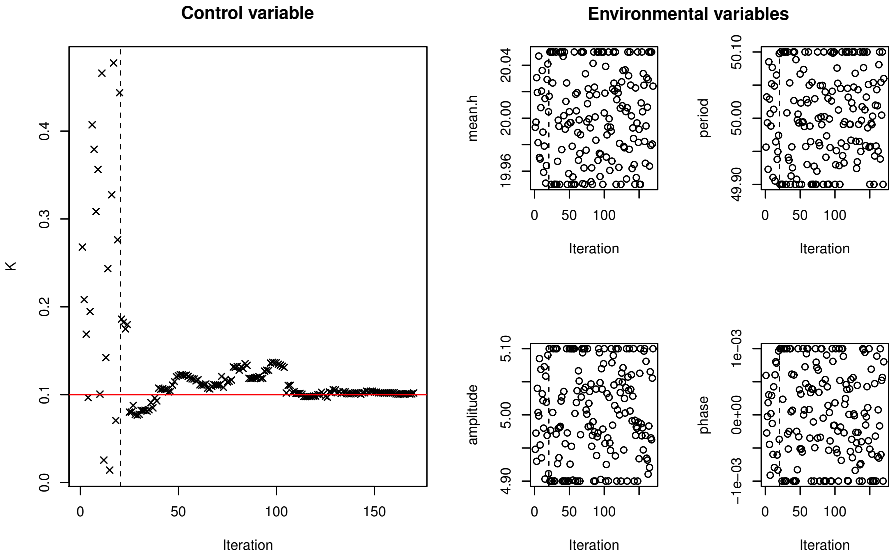
Expected improvement

$$EI(\boldsymbol{x}_e, K) = \mathbb{E}[\max\{0, j^N_{\min} - \bar{J}_N(\boldsymbol{x}_e, K)\}]$$

Explorative EGO iteration

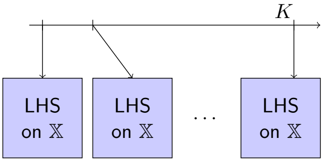
$$(\tilde{\boldsymbol{x}}_e, \tilde{K}) = \arg\max EI(\boldsymbol{x}_e, K)$$
$$\boldsymbol{x}^*_e = \arg\max_{\boldsymbol{x}_e} d\left((\boldsymbol{x}_e, \tilde{K}), \mathcal{P}_N\right)$$
$$\mathcal{P}_{N+1} = \mathcal{P}_N \cup (\boldsymbol{x}^*_e, \tilde{K})$$
$$\bar{J}_{N+1}(\boldsymbol{x}_e, K) \sim \mathcal{N}\left(\hat{m}_{N+1}(\boldsymbol{x}_e, K), \hat{s}^2_{N+1}(\boldsymbol{x}_e, K)\right)$$

Explorative EGO iterations [Lehman et al., 2004]



Iterated LHS

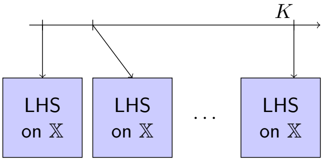
M-robustness



to build $\hat{\mu}(K)$ and $\hat{\sigma}^2(K)$

Iterated LHS

M-robustness



to build $\hat{\mu}(K)$ and $\hat{\sigma}^2(K)$

Estimate of the mean

$$\mathbb{E}[J(\boldsymbol{X}_e, K)|K] \xrightarrow{\text{estimator}} \hat{\mu}(K) = n_e^{-1} \sum_{i=1}^{n_e} J(\boldsymbol{X}_e^i, K) \text{ (r.v.)}$$

$$\mathbb{V}\mathrm{ar}[J(\boldsymbol{X}_e, K)|K] \xrightarrow{\text{estimator}} \hat{\sigma}^2(K) = (n_e - 1)^{-1} \sum_{i=1}^{n_e} (J(\boldsymbol{X}_e^i, K) - \hat{\mu}(K))^2 \text{ (r.v.)}$$

$$\mathbb{E}[\hat{\mu}(K)] = \mathbb{E}[J(\boldsymbol{X}_e, K)|K] \quad \text{and} \quad \mathbb{V}\mathrm{ar}[\hat{\mu}(K)] = \frac{\mathbb{V}\mathrm{ar}[J(\boldsymbol{X}_e, K)|K]}{n_e} \approx \frac{\hat{\sigma}^2(K)}{n_e}$$

Estimated mean as a random variable

$$\hat{\mu}(K) \sim \mathcal{N}\left(\mathbb{E}[J(\boldsymbol{X}_e, K)|K], \frac{\hat{\sigma}^2(K)}{n_e}\right) \qquad \text{(CLT approximation)}$$

Idea [Rullière et al., 2013] :

- Add a new point K_{new} and estimate $\mathbb{E}[J(\boldsymbol{X}_e, K)|K = K_{\mathrm{new}}]$
- OR Reduce the variance by increasing n_e

Estimated mean as a random variable

$$\hat{\mu}(K) \sim \mathcal{N}\left(\mathbb{E}[J(\boldsymbol{X}_e, K)|K], \frac{\hat{\sigma}^2(K)}{n_e}\right) \qquad \text{(CLT approximation)}$$

Idea [Rullière et al., 2013] :

- Add a new point K_{new} and estimate $\mathbb{E}[J(\boldsymbol{X}_e, K)|K = K_{\mathrm{new}}]$
- OR Reduce the variance by increasing n_e

In an adaptative sampling strategy: $K^* = \arg\max_{K \in \mathcal{K}} \kappa(K)$

- if K^* "close" to an existing point $\longrightarrow$ increase n_e
- if not $\longrightarrow$ add K^* to the design

Knowledge-gradient criterion

Metamodel of the *estimated* mean:

Based on $\mathcal{P}_N$, an experimental design of N points on $\mathcal{K}$.

$$\bar{\mu}_N(K) \sim \mathcal{N}\left(\hat{m}_N(K), \hat{s}_N^2(K)\right)$$

Knowledge-gradient criterion

Metamodel of the *estimated* mean:

Based on $\mathcal{P}_N$, an experimental design of N points on $\mathcal{K}$.

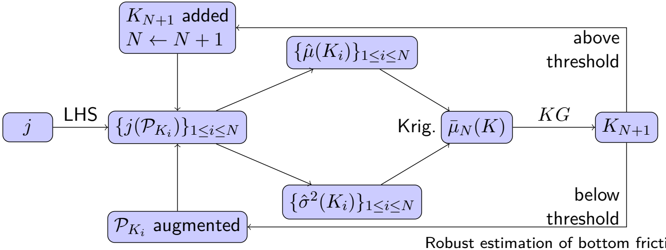
$$\bar{\mu}_N(K) \sim \mathcal{N}\left(\hat{m}_N(K), \hat{s}_N^2(K)\right)$$

In the presence of noise in the kriging model:

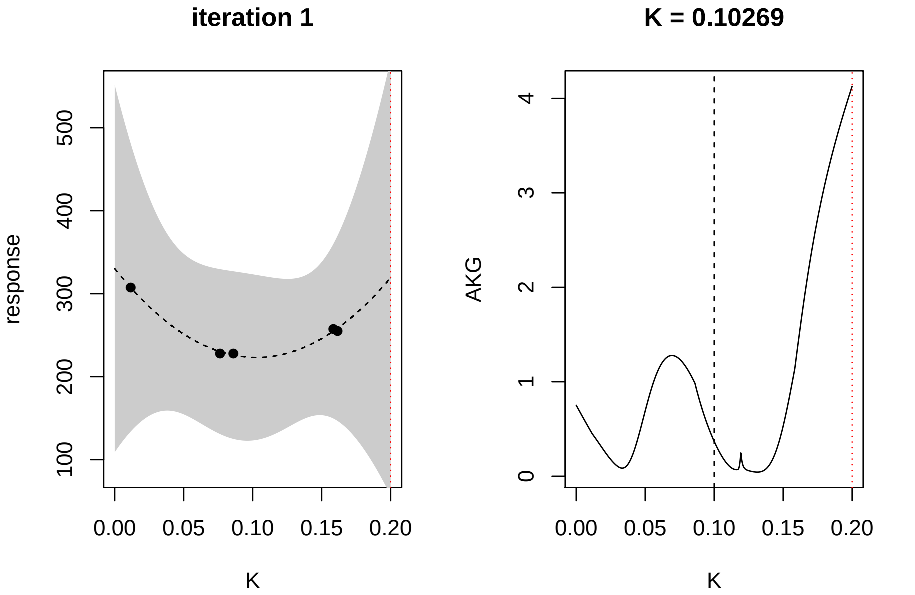
Definition of the KG [Frazier et al., 2008]

$$KG(\tilde{K}) = \min_{K' \in \mathcal{K}} \hat{m}_N(K') - \mathbb{E}\left[\min_{K' \in \mathcal{K}} \hat{m}_{N+1}(K') \middle| \tilde{K}\right]$$

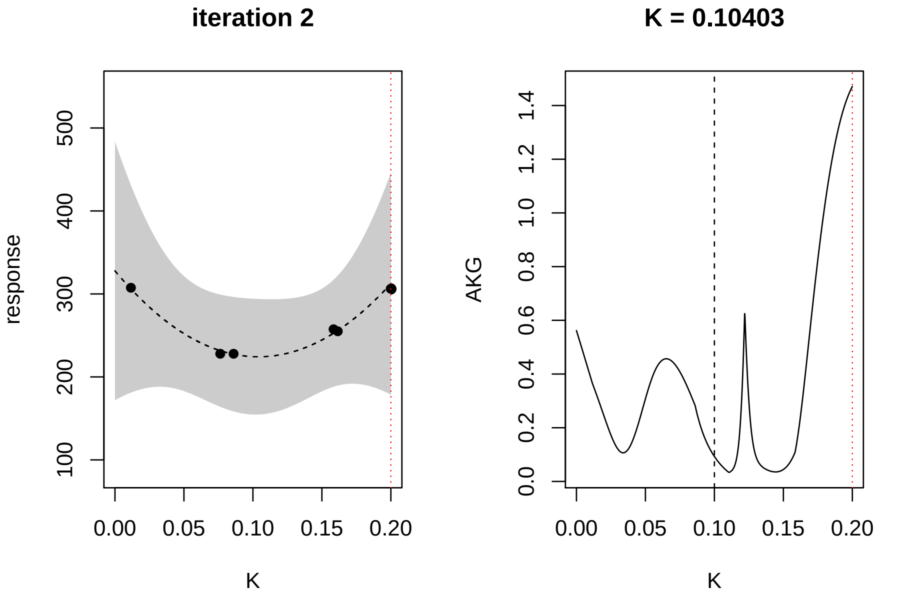
where $\hat{m}_{N+1}$ is the kriging mean computed based on $\mathcal{P}_N \cup \{\tilde{K}\}$



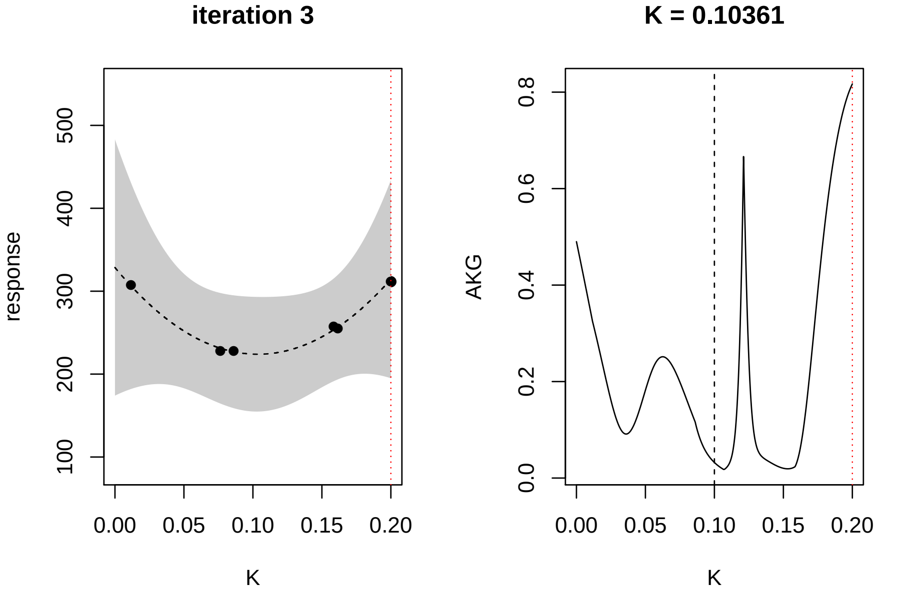
Iterated LHS



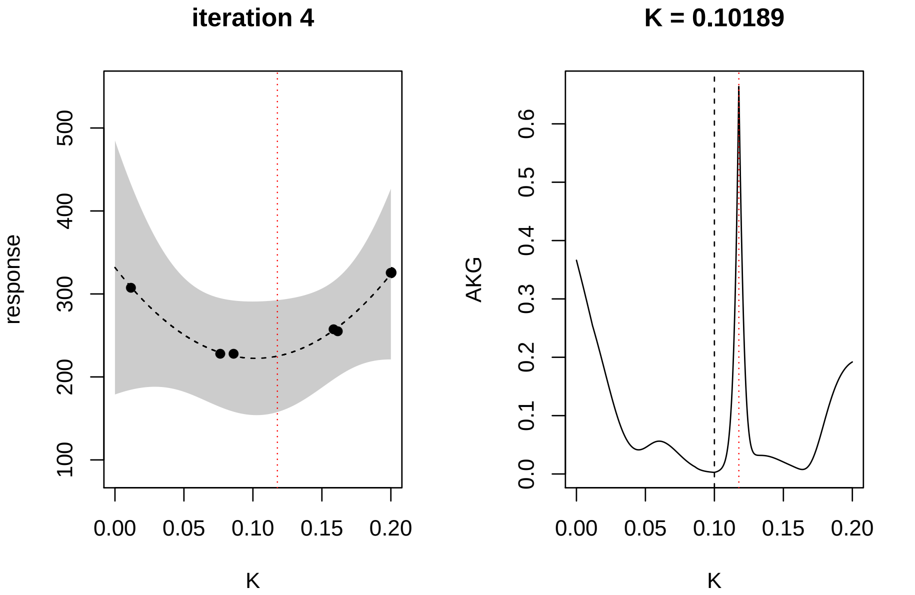
Iterated LHS



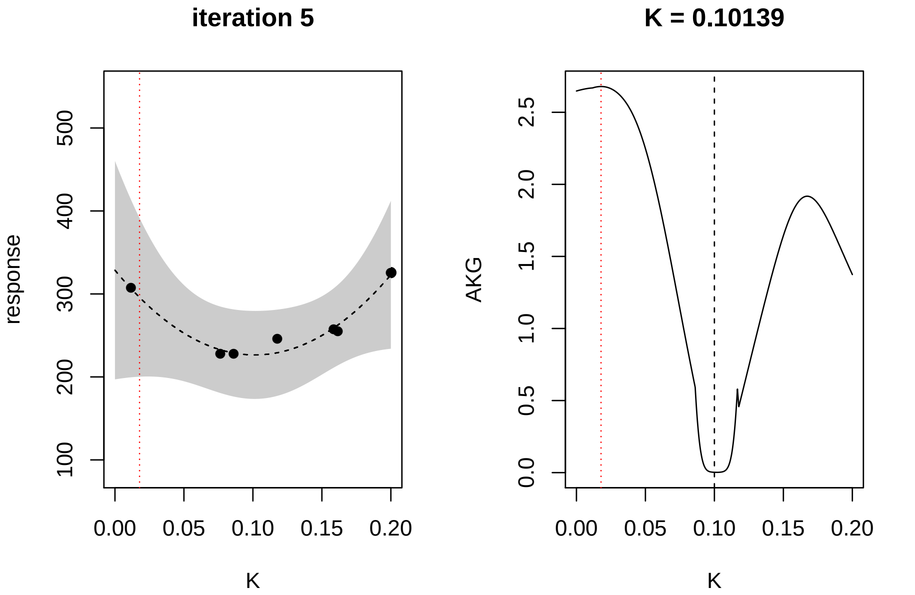
Iterated LHS



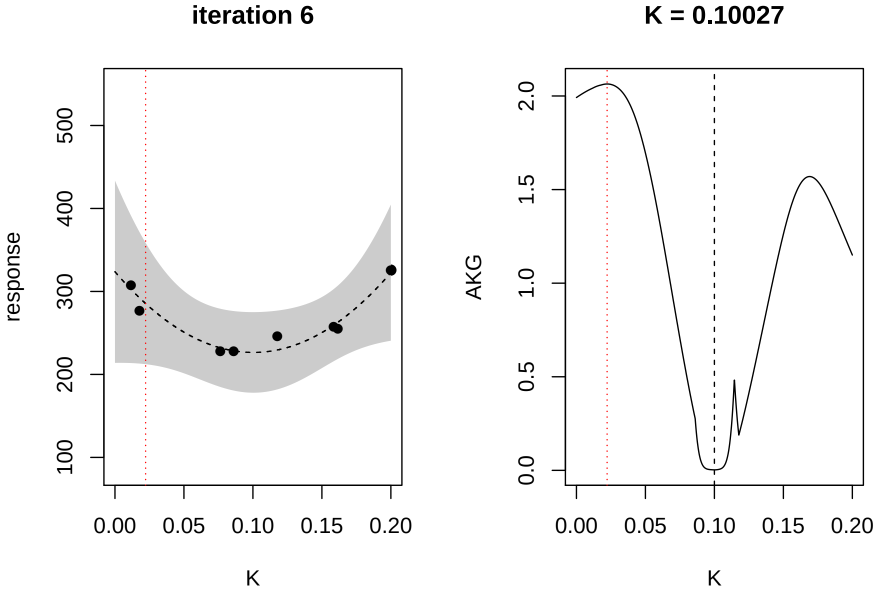
Iterated LHS



Iterated LHS



Iterated LHS



Wrapping up

- Notion of robustness
- Metamodelling techniques $\rightarrow$ include uncertainties
- Balance between precision and number of runs

Perspective and future work

- Better numerical model
  - 2D
  - Better numerical scheme
- Influence of observation operator $\mathcal{H}$
  - Y as real measurements
- Extension to K multidimensional
- Choice of metamodel:
  - Kriging
    - not adapted to high-dimensional input space
    - Adaptative sampling in multidimensional case?
  - PC
    - adapted to K multidimensional
    - "Fixed" grid to evaluate
    - May need more evaluation in some cases + adjoint

Thank you for your attention

- Introduction
  - AIRSEA team
  - Context and scope of the project
- Deterministic Framework
  - The 1D Shallow Water Equations
  - Adjoint-based optimization
- Global sensitivity analysis
  - Definition
  - SA of different outputs of the model
- Robust Optimization
  - Concepts of robustness
  - Metamodeling
  - Adaptative sampling
- Conclusion

Dellino, G., Kleijnen, J. P., and Meloni, C. (2012).
Robust optimization in simulation: Taguchi and krige combined.
INFORMS Journal on Computing, 24(3):471–484.

Frazier, P. I., Powell, W. B., and Dayanik, S. (2008).
A knowledge-gradient policy for sequential information collection.
SIAM Journal on Control and Optimization, 47(5):2410–2439.

Gilquin, L., Arnaud, E., Prieur, C., and Janon, A. (2017).
Making best use of permutations to compute sensitivity indices with replicated designs.
Submitted on June 28th.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998).
Efficient global optimization of expensive black-box functions.
Journal of Global Optimization, 13(4):455–492.

Krige, D. G. (1951).
A statistical approach to some mine valuation and allied problems on the Witwatersrand: By DG Krige.

Lehman, J. S., Santner, T. J., and Notz, W. I. (2004).
Designing computer experiments to determine robust control variables.
Statistica Sinica, 14(2):571–590.

Matheron, G. (1969).
Le krigeage universel.

Miranda, J., Kumar, D., and Lacor, C. (2016).
Adjoint-based robust optimization using polynomial chaos expansions.

Rullière, D., Faleh, A., Planchet, F., and Youssef, W. (2013).
Exploring or reducing noise? A global optimization algorithm in the presence of noise.
Structural and Multidisciplinary Optimization, 47(6):921–936.

Sobol, I. M. (2001).
Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates.
Mathematics and computers in simulation, 55(1):271–280.

Sudret, B. (2015).
Polynomial chaos expansions and stochastic finite element methods.
In Kok-Kwang Phoon, J. C., editor, *Risk and Reliability in Geotechnical Engineering*, pages 265–300. CRC Press.

Wiener, N. (1938).
The homogeneous chaos.
American Journal of Mathematics, 60(4):897–936.

Principle of the adjoint method

Let us assume that we have the (formal) model

$$J(W(K), K) \qquad \text{(Cost function)}$$

$$\text{s.t. } R(W(K), K) = 0 \qquad \text{(Model equation (SWE))}$$

Principle of the adjoint method

Let us assume that we have the (formal) model

$$J(W(K), K) \qquad \text{(Cost function)}$$

$$\text{s.t. } R(W(K), K) = 0 \qquad \text{(Model equation (SWE))}$$

$$\frac{\mathrm{d}J}{\mathrm{d}K} = ? \tag{1}$$

Principle of the adjoint method

Let us assume that we have the (formal) model

$$J(W(K), K) \qquad \text{(Cost function)}$$

$$\text{s.t. } R(W(K), K) = 0 \qquad \text{(Model equation (SWE))}$$

$$\frac{\mathrm{d}J}{\mathrm{d}K} = ? \tag{1}$$

$$\frac{\mathrm{d}J}{\mathrm{d}K} = \frac{\partial J}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial J}{\partial K} \tag{2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}K} = \frac{\partial R}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial R}{\partial K} = 0 \tag{3}$$

Principle of the adjoint method

Let us assume that we have the (formal) model

$$J(W(K), K) \qquad \text{(Cost function)}$$

$$\text{s.t. } R(W(K), K) = 0 \qquad \text{(Model equation (SWE))}$$

$$\frac{\mathrm{d}J}{\mathrm{d}K} = ? \tag{1}$$

$$\frac{\mathrm{d}J}{\mathrm{d}K} = \frac{\partial J}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial J}{\partial K} \tag{2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}K} = \frac{\partial R}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial R}{\partial K} = 0 \tag{3}$$

And by introducing a Lagrange multiplier λ

Principle of the adjoint method

$$\frac{\mathrm{d}J}{\mathrm{d}K} = \frac{\partial J}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial J}{\partial K} - \lambda^T \left(\frac{\partial R}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial R}{\partial K} \right) \tag{4}$$

$$= \underbrace{\left(\frac{\partial J}{\partial W} - \lambda^T \frac{\partial R}{\partial W} \right)}_{=0 \text{ if } \lambda \text{ is choosen wisely}} \frac{\partial W}{\partial K} + \left(\frac{\partial J}{\partial K} - \lambda^T \frac{\partial R}{\partial K} \right) \tag{5}$$

Adjoint model

$$\left(\frac{\partial R}{\partial W} \right)^T \lambda = \left(\frac{\partial J}{\partial W} \right)^T \tag{6}$$

Principle of the adjoint method

$$\frac{\mathrm{d}J}{\mathrm{d}K} = \frac{\partial J}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial J}{\partial K} - \lambda^T \left(\frac{\partial R}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial R}{\partial K}\right) \tag{4}$$

$$= \underbrace{\left(\frac{\partial J}{\partial W} - \lambda^T \frac{\partial R}{\partial W}\right)}_{=0 \text{ if } \lambda \text{ is choosen wisely}} \frac{\partial W}{\partial K} + \left(\frac{\partial J}{\partial K} - \lambda^T \frac{\partial R}{\partial K}\right) \tag{5}$$

Adjoint model

$$\left(\frac{\partial R}{\partial W}\right)^T \lambda = \left(\frac{\partial J}{\partial W}\right)^T \tag{6}$$

For a value K:

Principle of the adjoint method

$$\frac{\mathrm{d}J}{\mathrm{d}K} = \frac{\partial J}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial J}{\partial K} - \lambda^T\left(\frac{\partial R}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial R}{\partial K}\right) \tag{4}$$

$$= \underbrace{\left(\frac{\partial J}{\partial W} - \lambda^T\frac{\partial R}{\partial W}\right)}_{=0 \text{ if } \lambda \text{ is choosen wisely}}\frac{\partial W}{\partial K} + \left(\frac{\partial J}{\partial K} - \lambda^T\frac{\partial R}{\partial K}\right) \tag{5}$$

Adjoint model

$$\left(\frac{\partial R}{\partial W}\right)^T\lambda = \left(\frac{\partial J}{\partial W}\right)^T \tag{6}$$

For a value K:

- solve $R(W(K), K) = 0$ for W

Principle of the adjoint method

$$\frac{\mathrm{d}J}{\mathrm{d}K} = \frac{\partial J}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial J}{\partial K} - \lambda^T \left(\frac{\partial R}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial R}{\partial K}\right) \tag{4}$$

$$= \underbrace{\left(\frac{\partial J}{\partial W} - \lambda^T \frac{\partial R}{\partial W}\right)}_{=0 \text{ if } \lambda \text{ is choosen wisely}} \frac{\partial W}{\partial K} + \left(\frac{\partial J}{\partial K} - \lambda^T \frac{\partial R}{\partial K}\right) \tag{5}$$

Adjoint model

$$\left(\frac{\partial R}{\partial W}\right)^T \lambda = \left(\frac{\partial J}{\partial W}\right)^T \tag{6}$$

For a value K:

- solve $R(W(K), K) = 0$ for W
- solve $\left(\frac{\partial R}{\partial W}\right)^T \lambda = \left(\frac{\partial J}{\partial W}\right)^T$ for λ

Principle of the adjoint method

$$\frac{\mathrm{d}J}{\mathrm{d}K} = \frac{\partial J}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial J}{\partial K} - \lambda^T \left(\frac{\partial R}{\partial W}\frac{\partial W}{\partial K} + \frac{\partial R}{\partial K} \right) \tag{4}$$

$$= \underbrace{\left(\frac{\partial J}{\partial W} - \lambda^T \frac{\partial R}{\partial W} \right)}_{=0 \text{ if } \lambda \text{ is choosen wisely}} \frac{\partial W}{\partial K} + \left(\frac{\partial J}{\partial K} - \lambda^T \frac{\partial R}{\partial K} \right) \tag{5}$$

Adjoint model

$$\left(\frac{\partial R}{\partial W} \right)^T \lambda = \left(\frac{\partial J}{\partial W} \right)^T \tag{6}$$

For a value K:

- solve $R(W(K), K) = 0$ for W
- solve $\left(\frac{\partial R}{\partial W}\right)^T \lambda = \left(\frac{\partial J}{\partial W}\right)^T$ for λ
- compute $\frac{\mathrm{d}J}{\mathrm{d}K}$

Orthogonal Polynomials in a Hilbert space

X, $J(X)$ r.v., such that $\mathbb{E}[X^2], \mathbb{E}[J(X)^2] < +\infty$
$\rightarrow$ Expand J on a basis of orthogonal polynomials with respect to a specific inner product [Wiener, 1938, Sudret, 2015].

Orthogonal Polynomials in a Hilbert space

X, $J(X)$ r.v., such that $\mathbb{E}[X^2], \mathbb{E}[J(X)^2] < +\infty$
$\rightarrow$ Expand J on a basis of orthogonal polynomials with respect to a specific inner product [Wiener, 1938, Sudret, 2015].
Functional inner product:

$$\langle f, g \rangle = \int_{D_X} f(\xi) g(\xi) p_X(\xi) \, \mathrm{d}\xi$$

Family of orthogonal polynomials

$$\langle \varphi_i, \varphi_j \rangle = \|\varphi_i\|^2 \delta_{ij}$$

Family	D_X	p.d.f.	Distribution
Legendre	$[-1;1]$	$p_X(\xi) = {}^1\!/_2$	$\mathcal{U}([-1;1])$
Hermite	$\mathbb{R}$	$p_X(\xi) = e^{-\xi^2/2}$	$\mathcal{N}(0,1)$
Laguerre	$[0, +\infty[$	$p_X(\xi) = e^{-\xi}$	Exp

The expansion as a surrogate model

Polynomial Chaos Expansion, 1D

$$\mathcal{J} = J(X) = \sum_{i=0}^{+\infty} \hat{J}_i \varphi_i(X) \approx \sum_{i=0}^{P} \hat{J}_i \varphi_i(X)$$

The expansion as a surrogate model

Polynomial Chaos Expansion, 1D

$$\mathcal{J} = J(X) = \sum_{i=0}^{+\infty} \hat{J}_i \varphi_i(X) \approx \sum_{i=0}^{P} \hat{J}_i \varphi_i(X)$$

$$\boldsymbol{\alpha} = (\alpha_1, \dots \alpha_p), \quad |\boldsymbol{\alpha}| = \sum_i \alpha_i$$

$$\Phi_{\boldsymbol{\alpha}}(\boldsymbol{X}) = \prod_i \varphi_{\alpha_i}(X_i)$$

Polynomial Chaos Expansion, multidimensional

$$\mathcal{J} = J(\boldsymbol{X}) = \sum_{|\boldsymbol{\alpha}|=0}^{+\infty} \hat{J}_{\boldsymbol{\alpha}} \Phi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \approx \sum_{|\boldsymbol{\alpha}| \leq P} \hat{J}_{\boldsymbol{\alpha}} \Phi_{\boldsymbol{\alpha}}(\boldsymbol{X})$$

Statistical moments using PCE

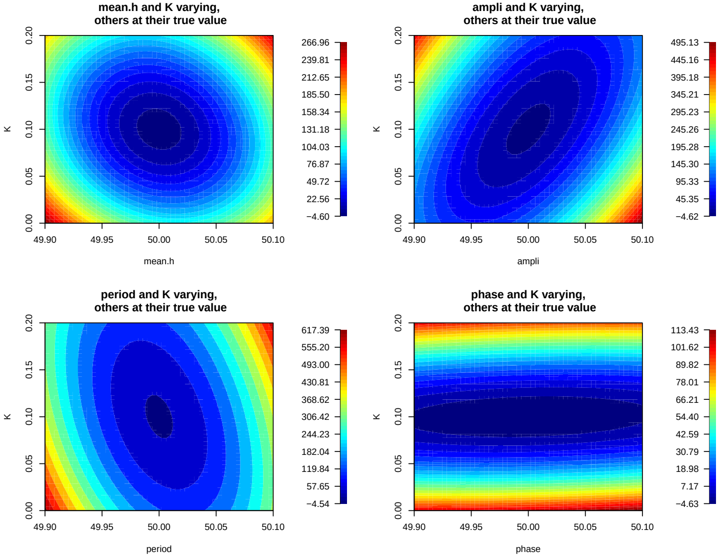
PCE allows us to get easily the statistical moments:

Mean and variance of $\mathcal{J}$ using the coefficients of the expansion

$$\mathbb{E}[\mathcal{J}] = \hat{J}_0 \|\boldsymbol{\Phi}_0\|^2 = \hat{J}_0$$

$$\mathbb{V}\mathrm{ar}[\mathcal{J}] = \mathbb{E}[\mathcal{J}^2] - \mathbb{E}[\mathcal{J}]^2 = \sum_{|\boldsymbol{\alpha}| \leq P} \hat{J}_{\boldsymbol{\alpha}}^2 \|\boldsymbol{\Phi}_{\boldsymbol{\alpha}}\|^2 - \hat{J}_0^2 = \sum_{0 < |\boldsymbol{\alpha}| \leq P} \hat{J}_{\boldsymbol{\alpha}}^2 \|\boldsymbol{\Phi}_{\boldsymbol{\alpha}}\|^2$$

PCE on the response



Principle of Kriging

$\mathcal{X} = \{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(n_s)}\}$ and $\mathcal{Y} = j(\mathcal{X})$
We assume that the deterministic model $j(\boldsymbol{x})$ is the realization of a GP [Krige, 1951, Matheron, 1969]:

Kriging formalism

$$\underbrace{J(\boldsymbol{x})}_{\text{r.v.}} = \underbrace{\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{\beta}}_{\text{deter}} + \underbrace{\varepsilon(\boldsymbol{x})}_{\text{r.v.}}$$

$$\mathbb{E}[\varepsilon(\boldsymbol{x})] = 0 \quad \text{and} \quad \mathrm{Cov}(\varepsilon(\boldsymbol{x}), \varepsilon(\boldsymbol{x'})) = \sigma_J^2 \underbrace{R(\boldsymbol{x}, \boldsymbol{x'})}_{\text{Chosen}}$$

$$\bar{J}(\boldsymbol{x}) \sim J(\boldsymbol{x})|\mathcal{Y}$$

Kriging predicator

$$\boldsymbol{F} = \{f_j(\boldsymbol{x}^{(i)})\}_{\substack{1\le i\le n_s\\ 1\le j\le n_\beta}}$$

$$\boldsymbol{R} = \{R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})\}_{1\le i,j\le n_s}$$

$$\boldsymbol{r}(\boldsymbol{x}) = \left[R(\boldsymbol{x}^{(1)}, \boldsymbol{x}), \dots, R(\boldsymbol{x}^{(n_s)}, \boldsymbol{x})\right]^T$$

Joint distribution of $\mathcal{Y}$ and J

$$\begin{bmatrix} \mathcal{Y} \\ J(\boldsymbol{x}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{F}\hat{\boldsymbol{\beta}} \\ \boldsymbol{f}(\boldsymbol{x})^T\hat{\boldsymbol{\beta}} \end{bmatrix}, \sigma_J^2 \begin{bmatrix} \boldsymbol{R} & \boldsymbol{r}(\boldsymbol{x}) \\ \boldsymbol{r}(\boldsymbol{x})^T & 1 \end{bmatrix} \right)$$

Kriging predicator $\bar{J}$

$$\begin{aligned} \bar{J}(\boldsymbol{x}) &\sim J(\boldsymbol{x})|\mathcal{Y} \\ &\sim \mathcal{N}(\hat{m}_J(\boldsymbol{x}), \hat{s}_J^2(\boldsymbol{x})) \end{aligned}$$

Kriging predicator

$$
\begin{aligned}
\boldsymbol{F} &= \{f_j(\boldsymbol{x}^{(i)})\}_{\substack{1\le i\le n_s\\ 1\le j\le n_\beta}} \\
\boldsymbol{R} &= \{R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})\}_{1\le i,j\le n_s} \\
\boldsymbol{r}(\boldsymbol{x}) &= \left[R(\boldsymbol{x}^{(1)}, \boldsymbol{x}), \ldots, R(\boldsymbol{x}^{(n_s)}, \boldsymbol{x})\right]^T
\end{aligned}
$$

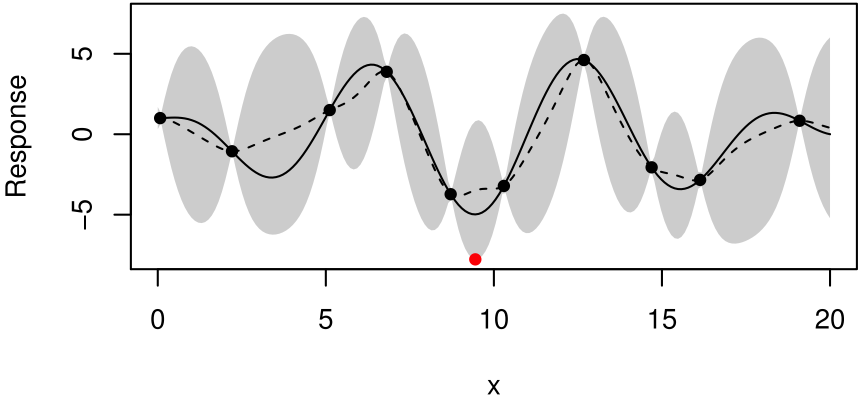
Joint distribution of $\mathcal{Y}$ and J

$$
\begin{bmatrix} \mathcal{Y} \\ J(\boldsymbol{x}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{F}\hat{\boldsymbol{\beta}} \\ \boldsymbol{f}(\boldsymbol{x})^T\hat{\boldsymbol{\beta}} \end{bmatrix}, \sigma_J^2 \begin{bmatrix} \boldsymbol{R} & \boldsymbol{r}(\boldsymbol{x}) \\ \boldsymbol{r}(\boldsymbol{x})^T & 1 \end{bmatrix} \right)
$$

Kriging predicator $\bar{J}$

$$
\begin{aligned}
\bar{J}(\boldsymbol{x}) &\sim J(\boldsymbol{x})|\mathcal{Y} \\
&\sim \mathcal{N}(\hat{m}_J(\boldsymbol{x}), \hat{s}_J^2(\boldsymbol{x}))
\end{aligned}
$$

Example of kriging



Example of kriging

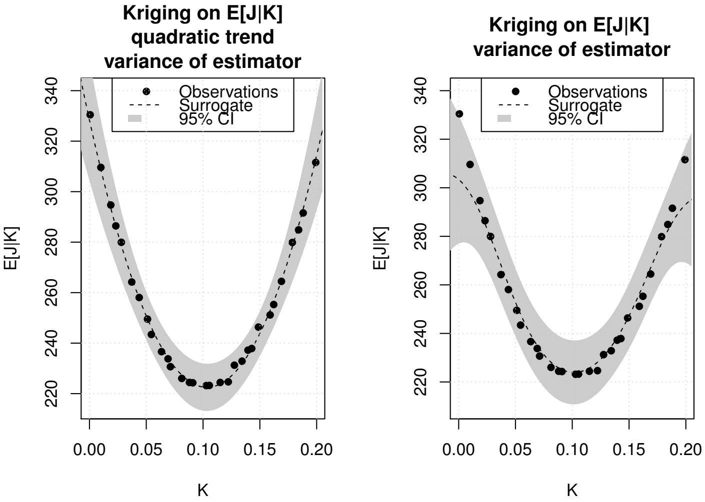


Figure : Kriging with and without trend of the estimate mean. Variance of observations based on the variance of the estimate

Multi-objective problem: general vocabulary

Vector of objective functions $\boldsymbol{f} = (f_1, \ldots, f_r)$:

Pareto domination relation

$$\boldsymbol{f}(\boldsymbol{x}) \prec \boldsymbol{f}(\boldsymbol{x}') \text{ if } \begin{cases} f_j(\boldsymbol{x}) \leq f_j(\boldsymbol{x}') & \forall j \leq r \\ f_j(\boldsymbol{x}) < f_j(\boldsymbol{x}') & \text{for one } j \leq r \end{cases}$$

Multi-objective problem: general vocabulary

Vector of objective functions $\boldsymbol{f} = (f_1, \ldots, f_r)$:

Pareto domination relation

$$\boldsymbol{f}(\boldsymbol{x}) \prec \boldsymbol{f}(\boldsymbol{x}') \text{ if } \begin{cases} f_j(\boldsymbol{x}) \leq f_j(\boldsymbol{x}') & \forall j \leq r \\ f_j(\boldsymbol{x}) < f_j(\boldsymbol{x}') & \text{for one } j \leq r \end{cases}$$

Pareto set, front

Pareto set:

$$\mathfrak{P} = \{\boldsymbol{x} \text{ s.t. } \nexists \boldsymbol{x}', \boldsymbol{f}(\boldsymbol{x}') \prec \boldsymbol{f}(\boldsymbol{x})\}$$

Pareto front:

$$\{\boldsymbol{z} \text{ s.t. } \exists \boldsymbol{x} \in \mathfrak{P}, \boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x})\}$$

Kriging to estimate the Pareto front [Dellino et al., 2012]

Minimization of the vector $(\mu(K), \sigma^2(K))$
Idea: Replace expensive computations of $(\mu(K), \sigma^2(K))$ by cheap computations of the metamodel

Kriging to estimate the Pareto front [Dellino et al., 2012]

Minimization of the vector $(\mu(K), \sigma^2(K))$
Idea: Replace expensive computations of $(\mu(K), \sigma^2(K))$ by cheap computations of the metamodel

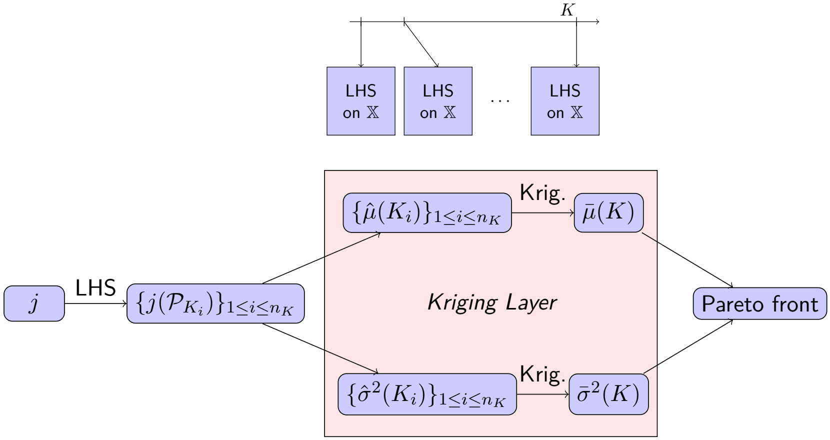
- Initial design on $\mathcal{K}$, compute for each one $\hat{\mu}$ and $\hat{\sigma}^2$, and generate a surrogate

Kriging to estimate the Pareto front [Dellino et al., 2012]

Minimization of the vector $(\mu(K), \sigma^2(K))$
Idea: Replace expensive computations of $(\mu(K), \sigma^2(K))$ by cheap computations of the metamodel

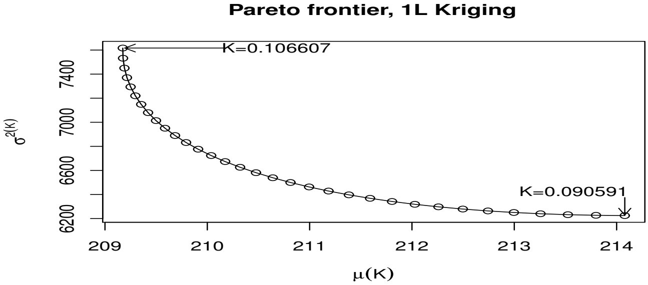
- Initial design on $\mathcal{K}$, compute for each one $\hat{\mu}$ and $\hat{\sigma}^2$, and generate a surrogate
- Initial design on $\mathbb{X} \times \mathcal{K}$, and use a surrogate to estimate mean and variance

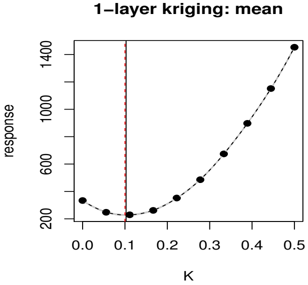
1L-Kriging

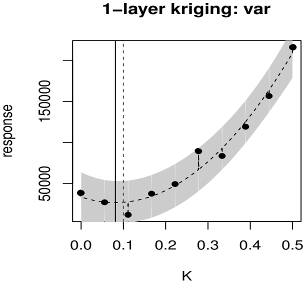


K
LHS on $\mathbb{X}$
LHS on $\mathbb{X}$
...
LHS on $\mathbb{X}$
j
LHS
$\{j(\mathcal{P}_{K_i})\}_{1\leq i\leq n_K}$
$\{\hat{\mu}(K_i)\}_{1\leq i\leq n_K}$
Krig.
$\bar{\mu}(K)$
Kriging Layer
$\{\hat{\sigma}^2(K_i)\}_{1\leq i\leq n_K}$
Krig.
$\bar{\sigma}^2(K)$
Pareto front

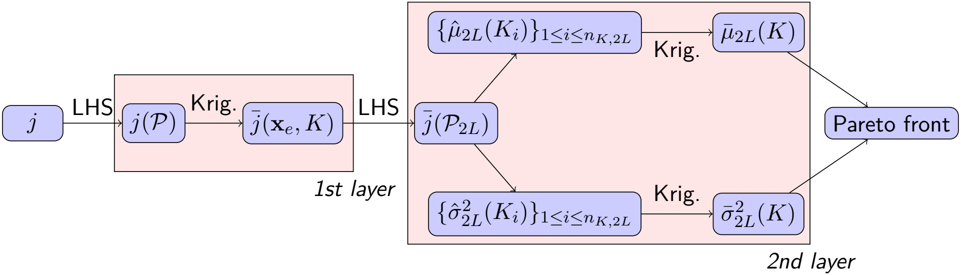
1L-Kriging


Pareto frontier, 1L Kriging
K=0.106607
K=0.090591
σ^2(K)
7400
7000
6600
6200
209
210
211
212
213
214
μ(K)

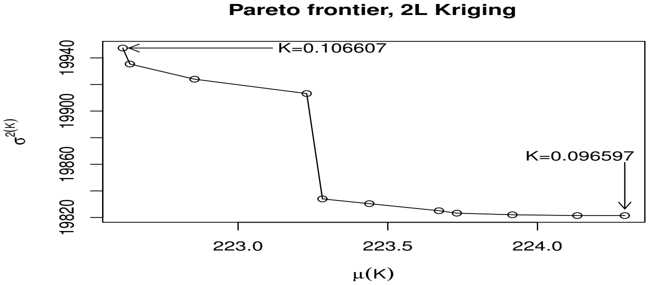

1−layer kriging: mean
response
1400
1000
600
200
0.0
0.1
0.2
0.3
0.4
0.5
K

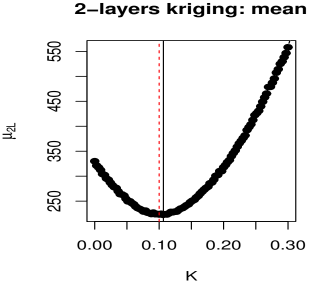

1−layer kriging: var
response
150000
50000
0.0
0.1
0.2
0.3
0.4
0.5
K

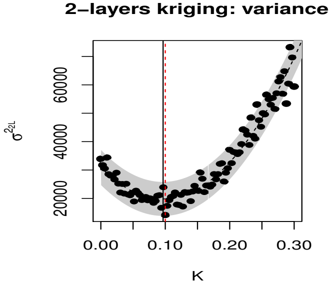
2L-Kriging


j
LHS
$j(\mathcal{P})$
Krig.
$\bar{j}(\mathbf{x}_e, K)$
1st layer
LHS
$\bar{j}(\mathcal{P}_{2L})$
$\{\hat{\mu}_{2L}(K_i)\}_{1\leq i\leq n_{K,2L}}$
Krig.
$\bar{\mu}_{2L}(K)$
$\{\hat{\sigma}^2_{2L}(K_i)\}_{1\leq i\leq n_{K,2L}}$
Krig.
$\bar{\sigma}^2_{2L}(K)$
Pareto front
2nd layer

2L-Kriging


Pareto frontier, 2L Kriging
K=0.106607
K=0.096597
19940
19900
19860
19820
223.0
223.5
224.0
$\mu(K)$
$\sigma^{2}(K)$


2−layers kriging: mean
550
450
350
250
0.00
0.10
0.20
0.30
K
μ_{2L}


2−layers kriging: variance
60000
40000
20000
0.00
0.10
0.20
0.30
K
σ^{2}_{2L}

Steepest descent using PCE [Miranda et al., 2016]

For a given K, expansion of $J(\boldsymbol{X}_e, K)$ and $\frac{\partial J}{\partial K}(\boldsymbol{X}_e, K)$

$$J(\boldsymbol{X}_e, K) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \hat{J}_{\boldsymbol{\alpha}}(K) \Phi_{\boldsymbol{\alpha}}(\boldsymbol{X}_e) \quad \text{and} \quad \frac{\partial J}{\partial K}(\boldsymbol{X}_e, K) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \hat{G}_{\boldsymbol{\alpha}}(K) \Phi_{\boldsymbol{\alpha}}(\boldsymbol{X}_e)$$

Relation between $\hat{J}_{\boldsymbol{\alpha}}(K)$ and $\hat{G}_{\boldsymbol{\alpha}}(K)$

$$\Rightarrow \frac{\mathrm{d}\hat{J}_{\boldsymbol{\alpha}}}{\mathrm{d}K}(K) = \hat{G}_{\boldsymbol{\alpha}}(K), \quad \forall \boldsymbol{\alpha} \in \mathcal{A}$$

Steepest descent using PCE [Miranda et al., 2016]

Recalling that

$$\hat{\sigma}_J^2(K) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \hat{J}_{\boldsymbol{\alpha}}(K)^2 \|\Phi_{\boldsymbol{\alpha}}\|^2$$

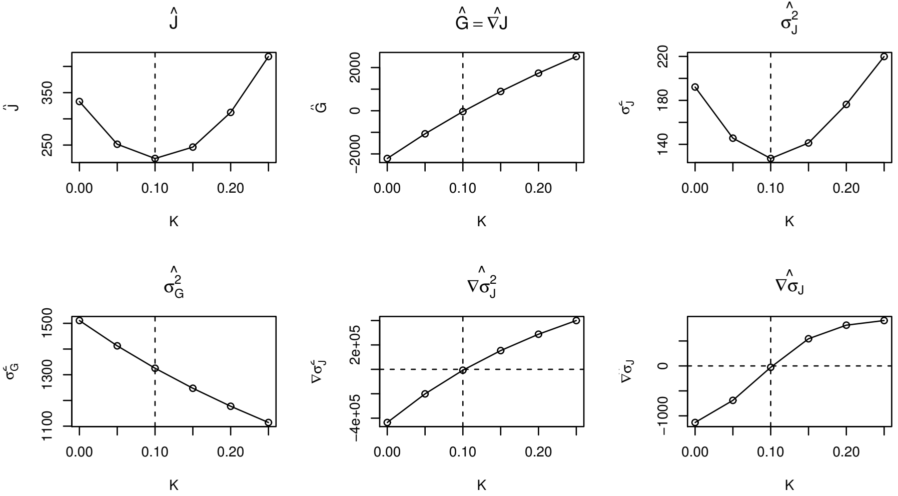
By differentiating with respect to K:

Gradient of the variance

$$\frac{\mathrm{d}\hat{\sigma}_J^2}{\mathrm{d}K}(K) = 2\sum_{\boldsymbol{\alpha}\in\mathcal{A}} \hat{J}_{\boldsymbol{\alpha}}(K)\frac{\mathrm{d}\hat{J}_{\boldsymbol{\alpha}}}{\mathrm{d}K}(K)\|\Phi_{\boldsymbol{\alpha}}\|^2 = 2\sum_{\boldsymbol{\alpha}\in\mathcal{A}} \hat{J}_{\boldsymbol{\alpha}}(K)\hat{G}_{\boldsymbol{\alpha}}(K)\|\Phi_{\boldsymbol{\alpha}}\|^2$$

We have the gradient of the mean $\hat{G}_0$ and the gradient of the variance
$\Rightarrow$ Gradient descent algorithm.

Expansion for different K



Gradient descent algorithm on ρ

$$\rho(K,\lambda) = \lambda\mu(K) + (1-\lambda)\sigma(K)$$