UQSay #07 Seminar

Iterative estimation in uncertainty and sensitivity analysis

Bertrand Iooss, EDF R&D

Joint works with Y. Fournier, A. Ribés (EDF R&D),
B. Raffin, T. Terraz (INRIA Rhône-Alpes)

& a Univ. Côte d'Azur Master students 2019 team
(T. Bertaina, Z. Mahfoudh, T. Ortais, R. Polizzi)

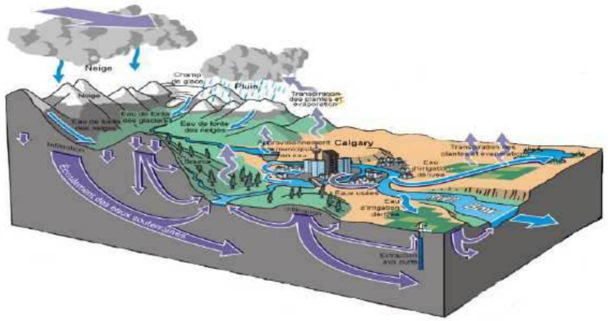
Gif-sur-Yvette

2020, 16th, January

CHANGER L'ÉNERGIE ENSEMBLE

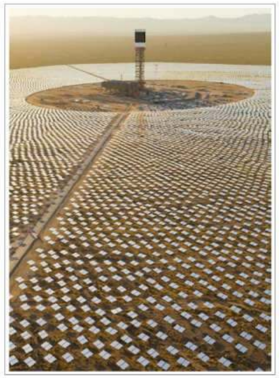
Uncertainty Quantification (UQ) in simulation-based studies

- Exploratory study : understand a phenomena, an experimental or industrial process



- Safety study : evaluate a safety margin (failure probability, rare events)

- Design study : optimizing and control the performances



Uncertainties

- Environmental variables
- Physical parameters
- Process parameters

↓ Design of experiments

Process: simulation code or experiments

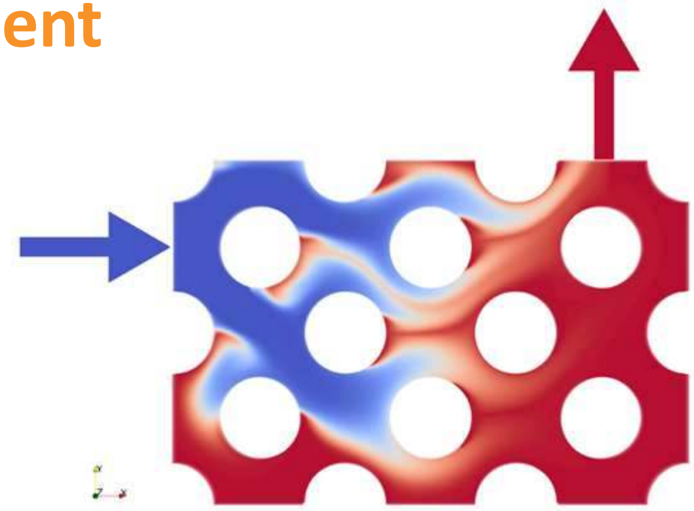
↓ Metamodel

- Output distributions
- Probability of failure
- « Main » influential input parameters

An example of a numerical experiment

<u>CFD computer code:</u>

Code_Saturne (EDF)



Simulation of the purge of hot water by introducing cold water

Example with the following meshing:

10 billion cells, 10x3 vectors per cell, 200 time steps => 12 TB / run

One parametric study would require hundreds of runs with:

- hot water varying from 300°C to 350°C
- cold water varying from 20°C to 30°C

If a probabilistic model is associated to the inputs, **uncertainty propagation aims to provide** the mean, variance, min, max or <u>full pdf</u> for temperature and pressure at each mesh element

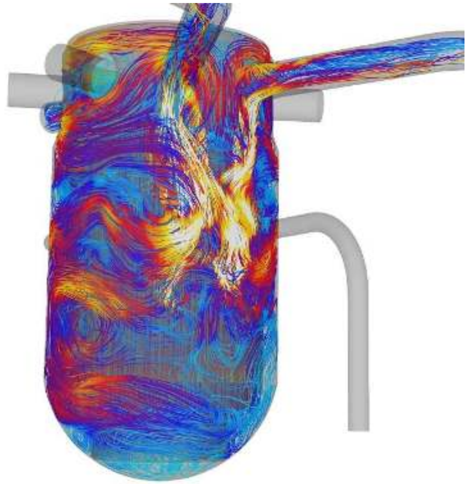
Storage used for *N* = 100 simulations: 1200 TB !

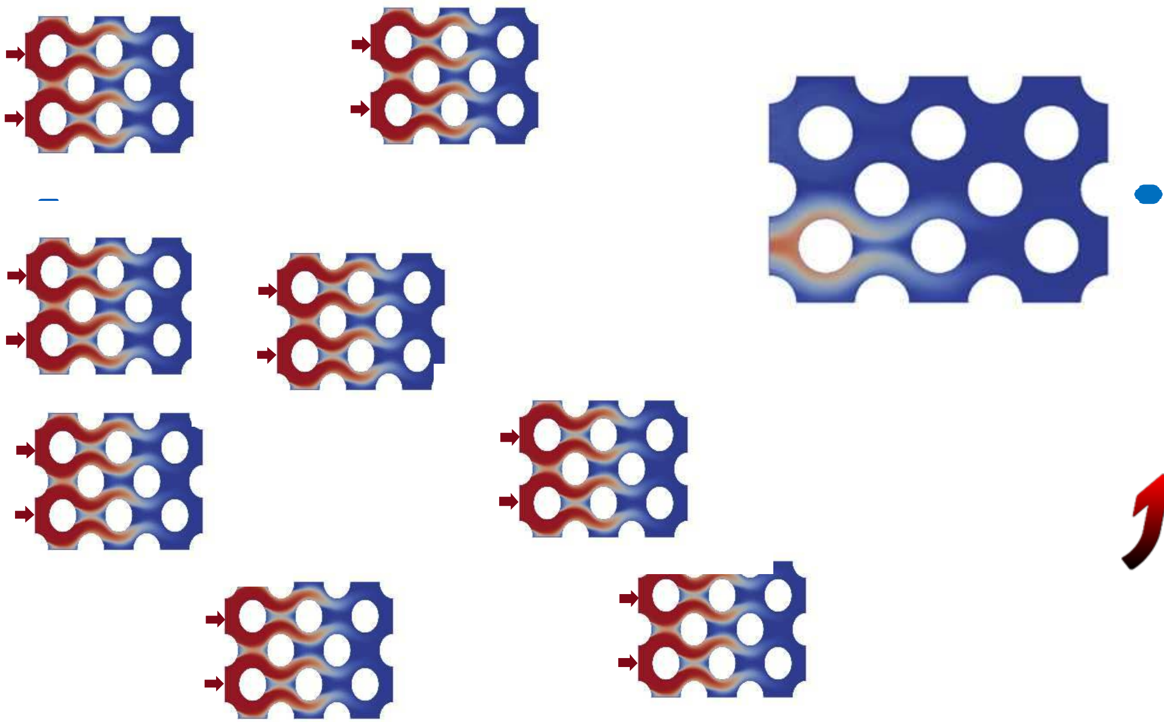
Iterative uncertainty quantification

- <u>Objective</u>: *in-situ* treatment of large volume of data (outputs of computer codes), due to file transfer/access and storage issues

- *In-situ* vs *a posteriori*: performing the data analysis at the same time as the calculation

- <u>Treatment</u>: visualisation, compression and **<u>statistical analysis</u>**



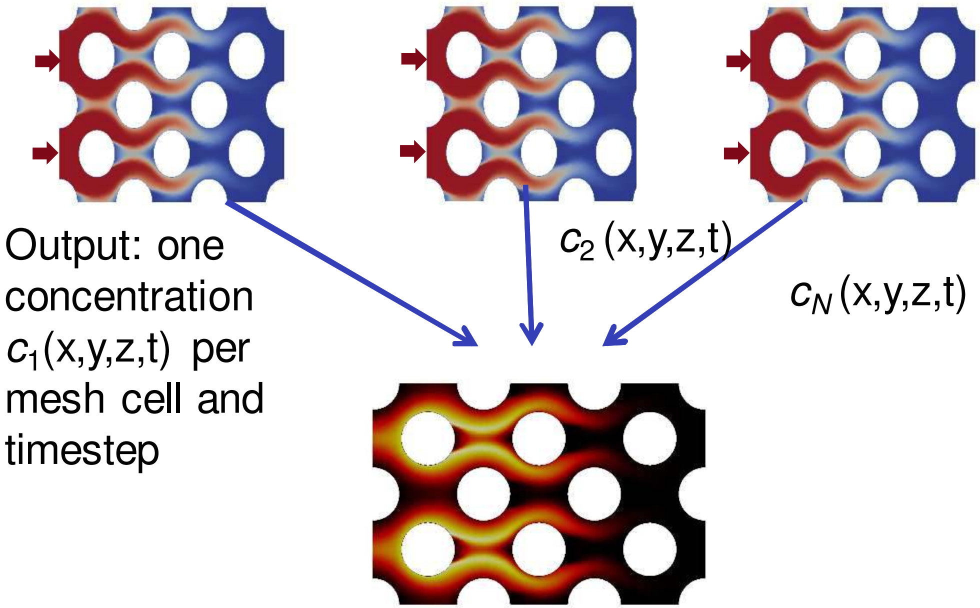


MELISSA

EDF

Iterative (in transit) statistics

N simulations with different parameter values (injection width, duration, dye concentration)



No intermediate files:

- Storage saving
- Time saving

Ubiquitous spatio-temporal **statistics**, i.e. everywhere in space and time

<u>Example on the mean estimation</u>

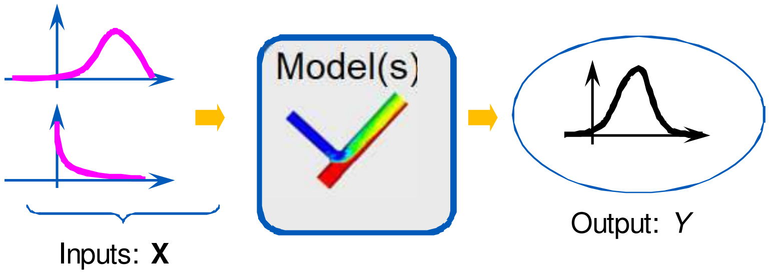
Replace the empirical mean $\mu(x,y,z,t) = \frac{1}{N}\sum_{n=1}^{N} c_n(x,y,z,t)$

by the one-pass average $\mu_n(x,y,z,t) = \mu_{n-1}(x,y,z,t) + \frac{1}{n}[c_n(x,y,z,t) - \mu_{n-1}(x,y,z,t)]$

with $n = 1, \ldots, N$ and $\mu_0(x,y,z,t) = 0$

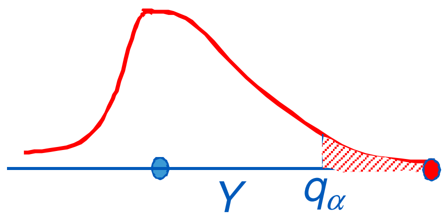
UQ methods considered in this talk

Uncertainty propagation



Quantities of interest:

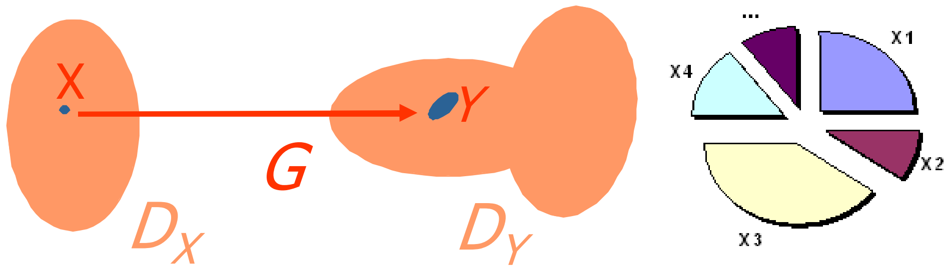
- Quantile of order α of Y



$$q^{\alpha} = \inf\{y : \mathrm{P}(Y \leq y) \geq \alpha\}$$

- Quantile function $Q(\alpha)$, $\alpha \in]0,1[$

Global sensitivity analysis



$\mathbf{X} = (X_1, \ldots, X_p)$

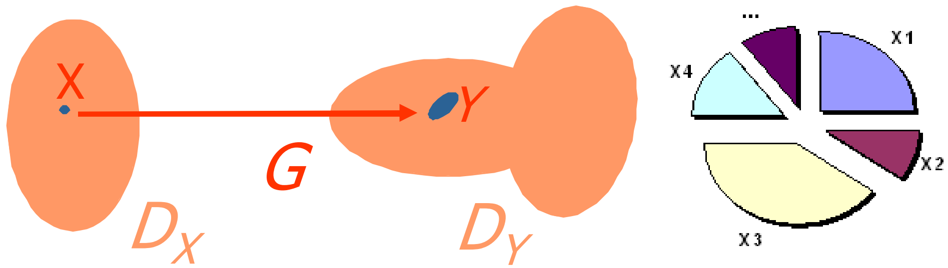
Quantity of interest: Sobol' indices ($i = 1, \ldots, p$)

$$S_i = \frac{\mathrm{Var}(\mathrm{E}[Y|X_i])}{\mathrm{Var}(Y)}$$

$$S_{Ti} = S_i + \sum_j S_{ij} + \sum_{j<k} S_{ijk} + \ldots$$

Part 1: Global sensitivity analysis

Global sensitivity analysis



$\mathbf{X} = (X_1, \ldots, X_p)$

Quantity of interest: Sobol' indices
($i = 1, \ldots, p$)

$$S_i = \frac{\mathrm{Var}(\mathrm{E}[Y|X_i])}{\mathrm{Var}(Y)}$$

$$S_{Ti} = S_i + \sum_j S_{ij} + \sum_{j<k} S_{ijk} + \ldots$$

Sobol' Index Estimation: pick-freeze method

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,p} \\ \vdots & \ddots & \\ a_{n,1} & & a_{n,p} \end{pmatrix}; B = \begin{pmatrix} b_{1,1} & \cdots & b_{1,p} \\ \vdots & \ddots & \\ b_{n,1} & & b_{n,p} \end{pmatrix}$$

A and B are independent random matrices

$$C^k = \begin{pmatrix} a_{1,1} & \cdots & a_{1,k-1} & b_{1,k} & a_{1,k+1} & \cdots & a_{1,p} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i,1} & \cdots & a_{i,k-1} & b_{i,k} & a_{n,k+1} & \cdots & a_{i,p} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,k-1} & b_{n,k} & a_{n,k+1} & \cdots & a_{n,p} \end{pmatrix}$$

C^k built from A and B

It requires running $n\,(p+2)$ simulations, with values given by each raw of A, B, C^k ($k = 1,\ldots,p$)

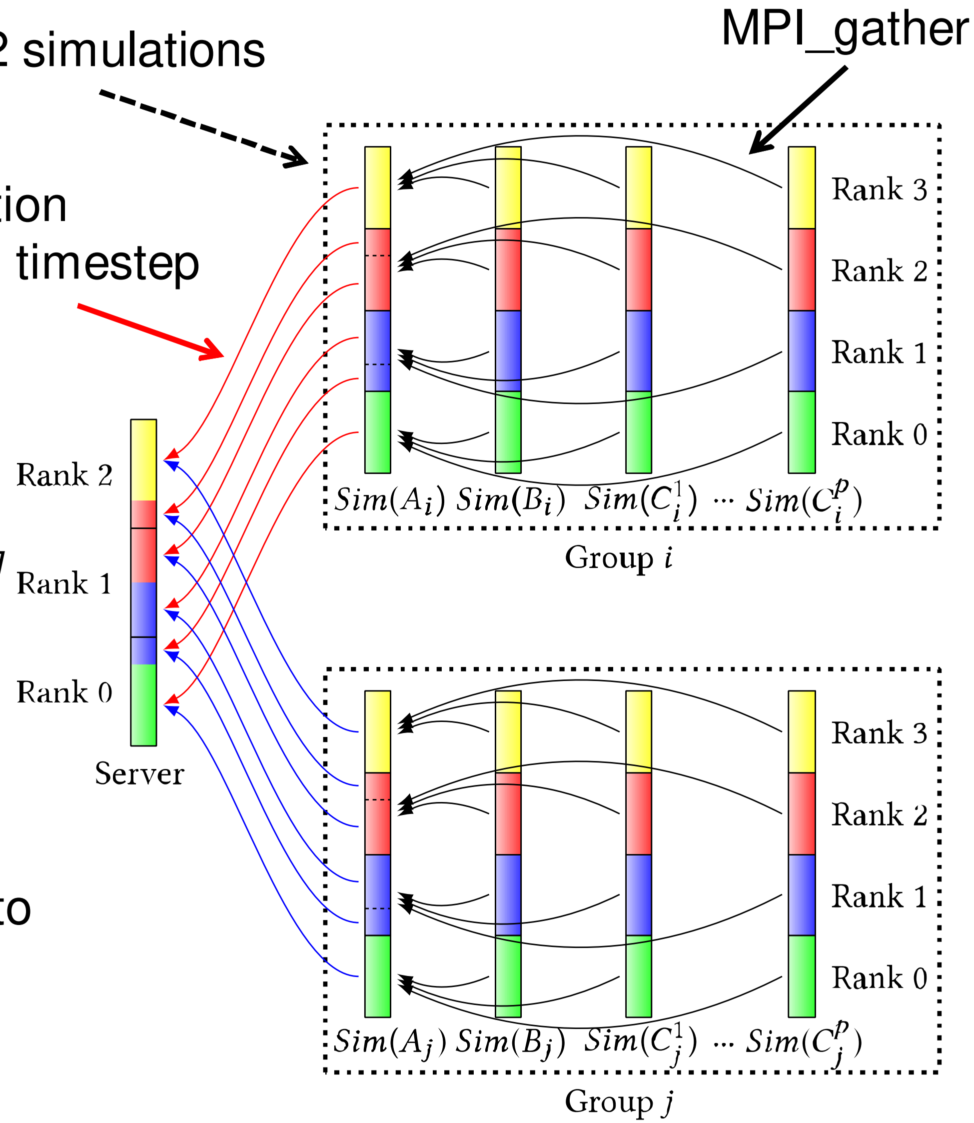
Estimators of first order and total Sobol' Indices:

$$S_k = \frac{\mathrm{Cov}(Y^B, Y^{C_k})}{\sqrt{\mathrm{Var}(Y^B)}\sqrt{\mathrm{Var}(Y^{C_k})}}$$

$$S_{Tk} = 1 - \frac{\mathrm{Cov}(Y^A, Y^{C_k})}{\sqrt{\mathrm{Var}(Y^A)}\sqrt{\mathrm{Var}(Y^{C_k})}}$$

Implementing the iterative estimation of Sobol' Indices

Dynamics parallel connection
Data redistribution at each timestep



<u>Melissa server:</u>

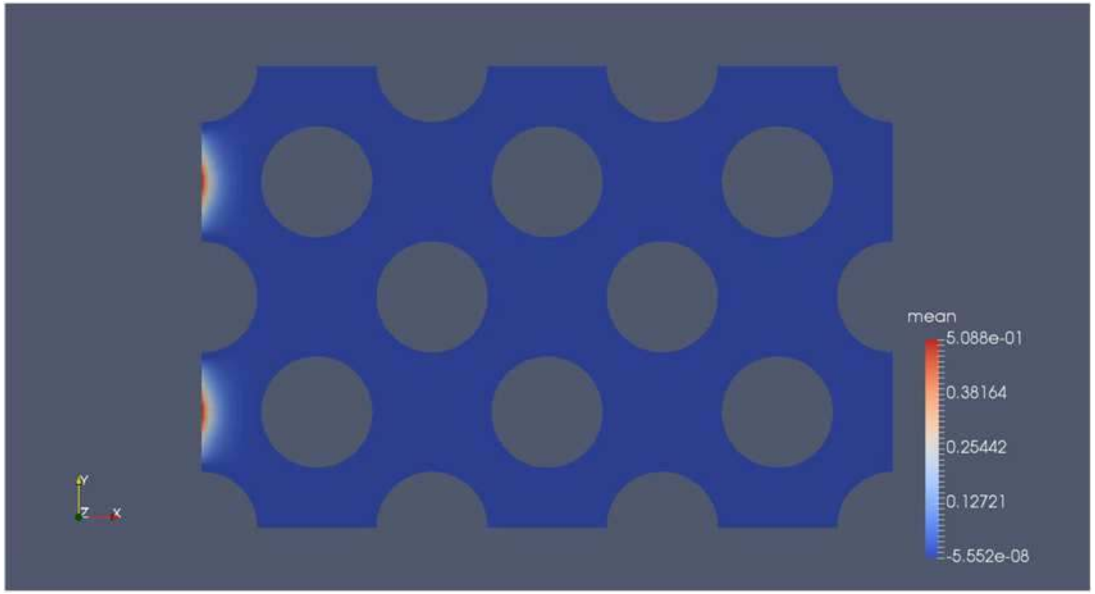
- Receive data (any order)
- Update Sobol' indices using variance & **covariance update formula** *[Pébay 2008]* General formula for Cov(A,B):

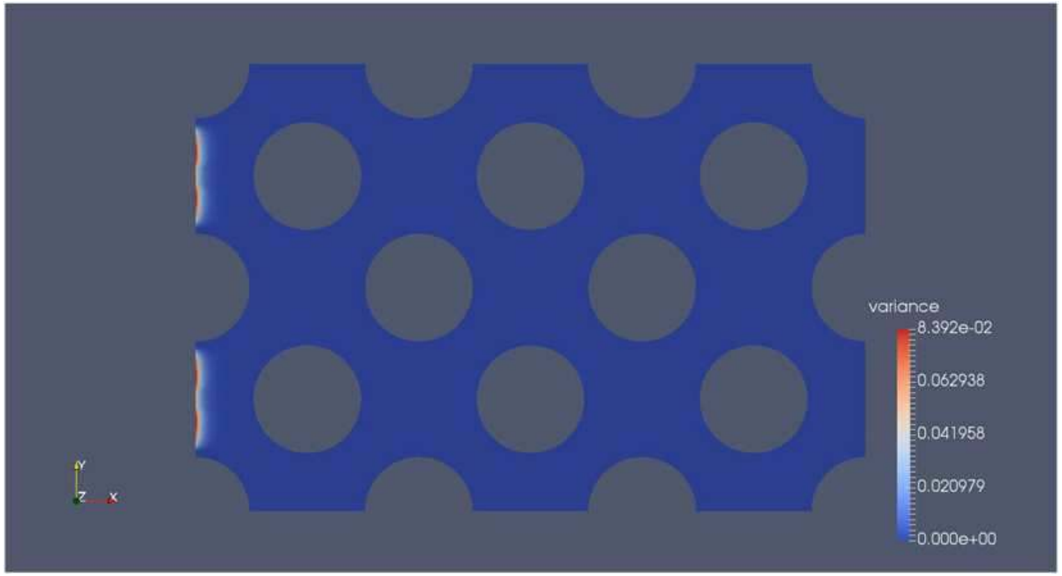
$$C_n = C_{n-1} + \frac{n-1}{n}(a - \mu_{n-1}^A)(b - \mu_{n-1}^B)$$

with $n = 1, \ldots, N$ and $C_0 = 0, \mu_0^A = 0, \mu_0^B = 0$

- Use of asymptotic confidence intervals to control their precision
- Discard data

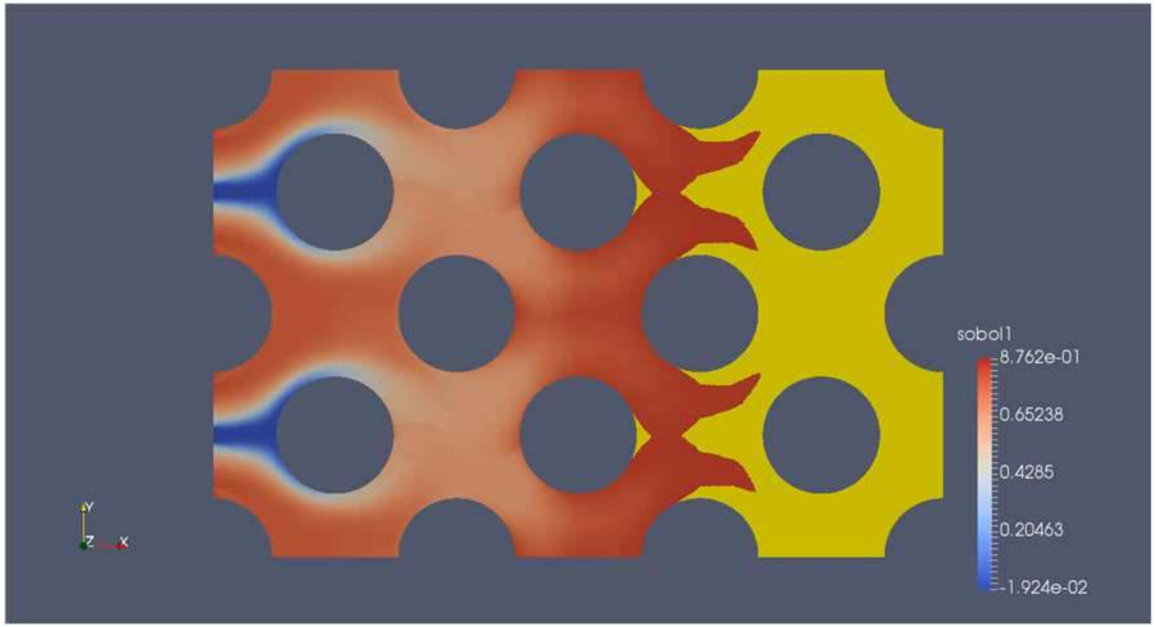
Uncertainty analysis results: Mean and variance of the temperature field



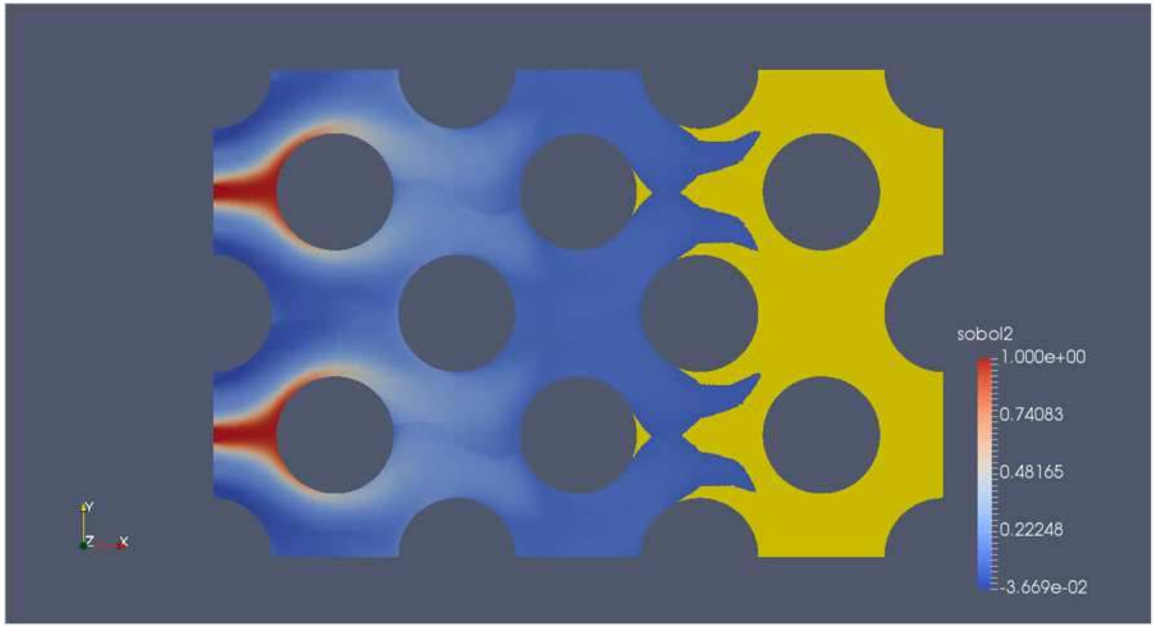


Sensitivity analysis results: First-order Sobol' indices

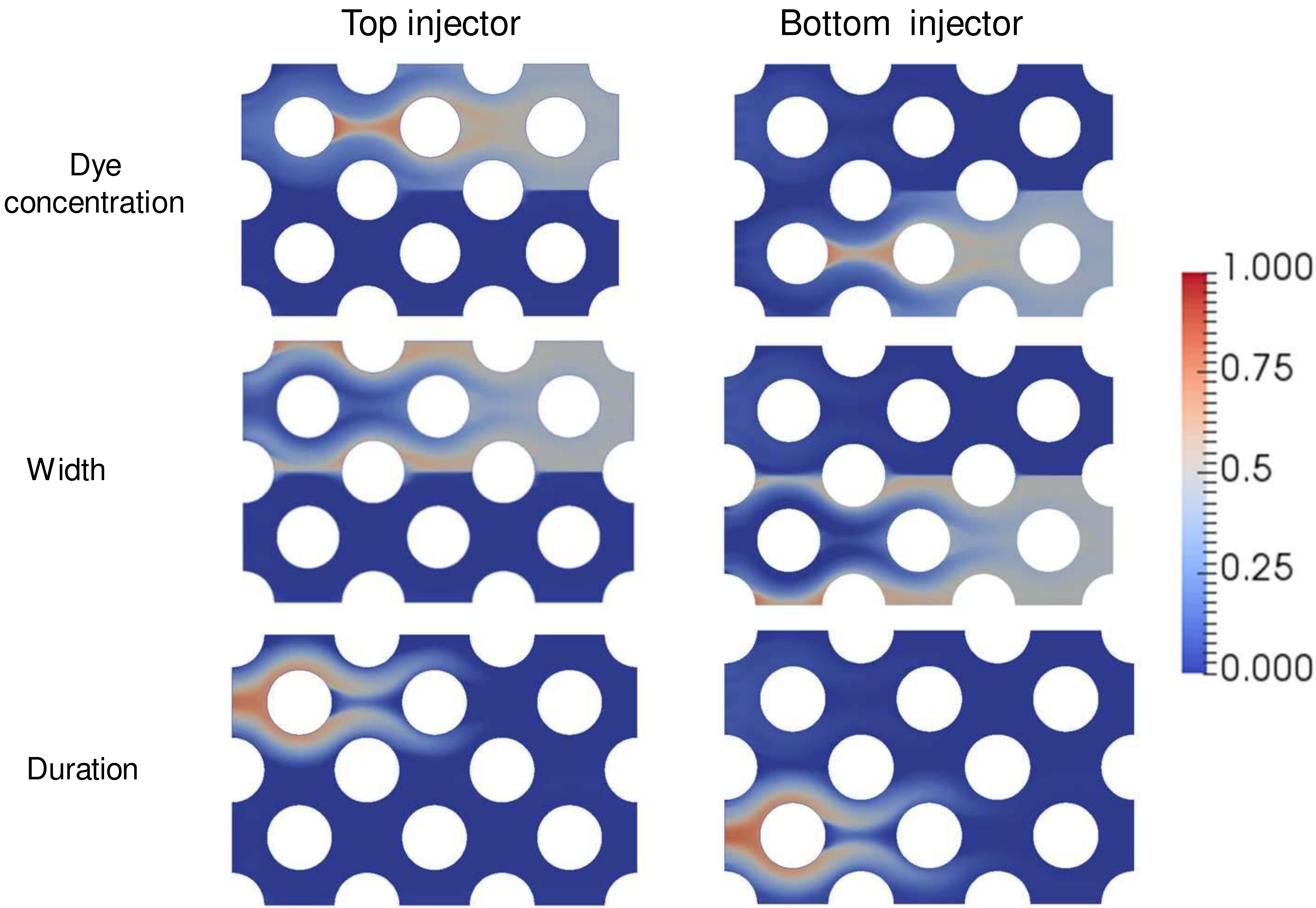
Injection width



Injection duration

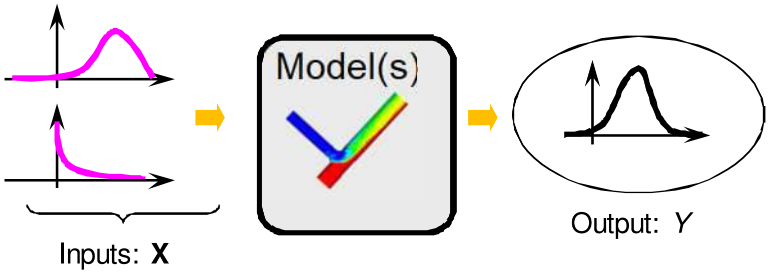


First-order Sobol' indices at timestep 80



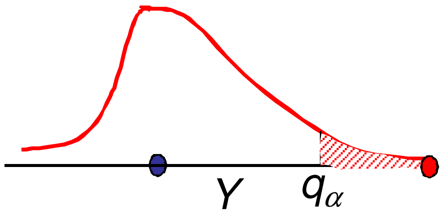
Part 2: Uncertainty propagation

Uncertainty propagation



Quantities of interest:

- Quantile of order α of Y



$$q^{\alpha} = \inf\{y : \mathrm{P}(Y \leq y) \geq \alpha\}$$

- Quantile function $Q(\alpha)$, $\alpha \in]0,1[$

Quantile estimation

$$\text{For } \alpha \in]0,1[, q^{\alpha} = \inf\{t \in \mathbb{R}, F_Y(t) \geq \alpha\}$$

- In the applications under study (quantile estimation of outputs of expensive numerical simulation code), we consider not so extreme α values:
  - $\alpha \in [0.01 , 0.99]$
  - N (total number of simulations) $\in [100 , 1000]$
  - The sample is denoted $(y^{(1)}, \dots, y^{(N)})$

- Empirical (Monte-Carlo) estimator (with an i.i.d. sample):

$$\hat{q}^{\alpha N} = \inf\{t \in \mathbb{R}, \hat{F}_Y^N(t) \geq \alpha\} \text{ where } \hat{F}_Y^N(t) = 1/N \sum_{n=1}^{N} 1_{\{y^{(n)} \leq t\}}$$

with 1_A the indicator function of the set A

Iterative α-quantile estimation: Robbins-Monro algorithm (RM)

$$q_{n+1} = q_n - \frac{C}{n^\gamma}(1_{y^{(n+1)} \le q_n} - \alpha)$$

$$\text{with } n = 1,2,\ldots, N \text{ and } q_1 = y^{(1)}$$

- Important hypotheses for the asymptotic convergence of the RM estimator:

$$\frac{C}{n^\gamma} \text{ decreases to } 0,\ \sum_{n\ge1} \frac{C}{n^\gamma} = \infty \text{ and } \sum_{n\ge1}\left(\frac{C}{n^\gamma}\right)^2 < \infty$$

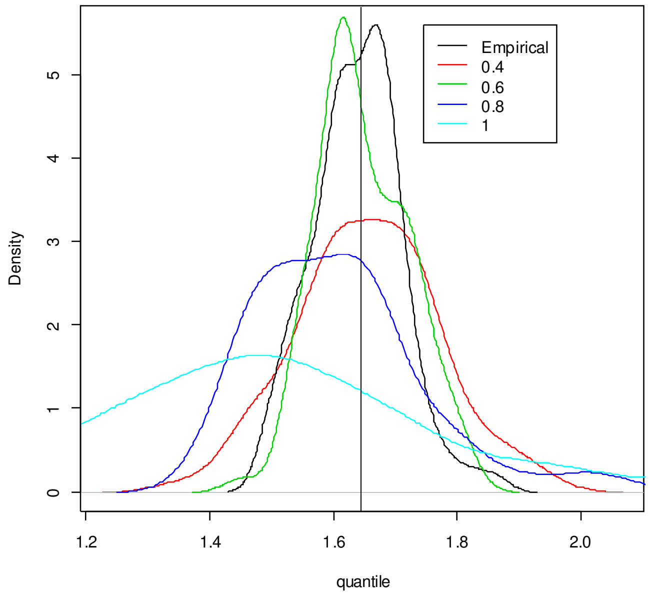
=> OK for $C > 0$ and γ€]0.5;1]

- RM averaging version is known to be more efficient then basic RM (it minimizes its asymptotic variance)

- However, when N is not large (our case):
  - averaging version does not work well,
  - there are important tuning issues for the constants C and γ

Following numerical tests: $C = 1$, $N = 1000$ and $\alpha = 0.95$

Mixing issues due to γ



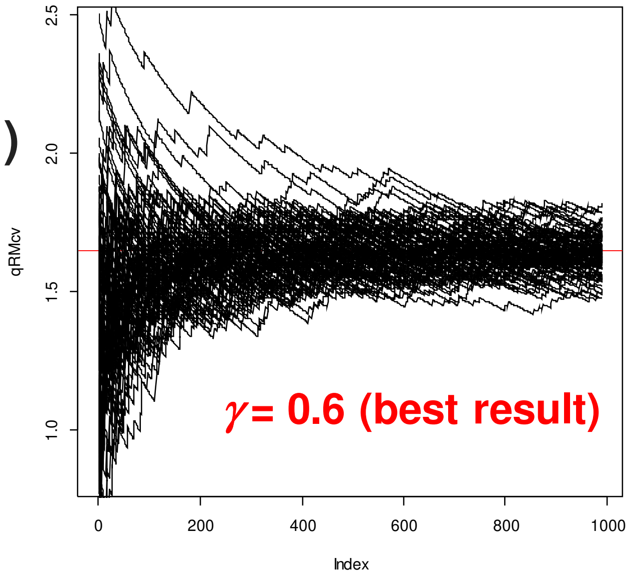
$\alpha = 0.95$

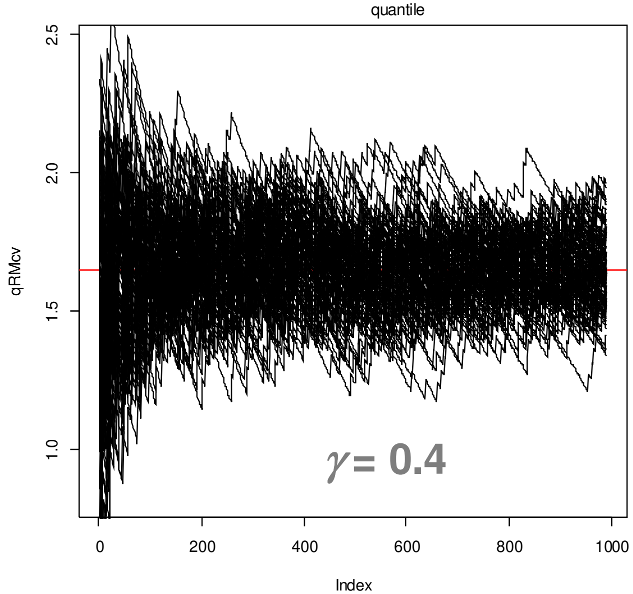
Gaussian pdf: $Y \sim \mathcal{N}(0,1)$

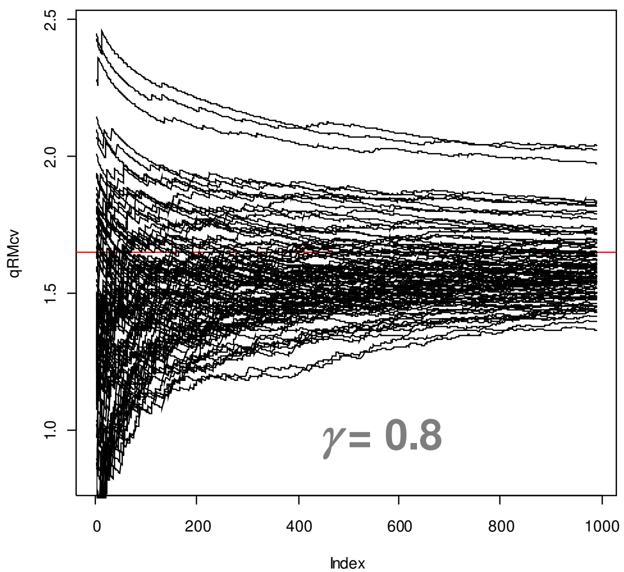
$C = 1$

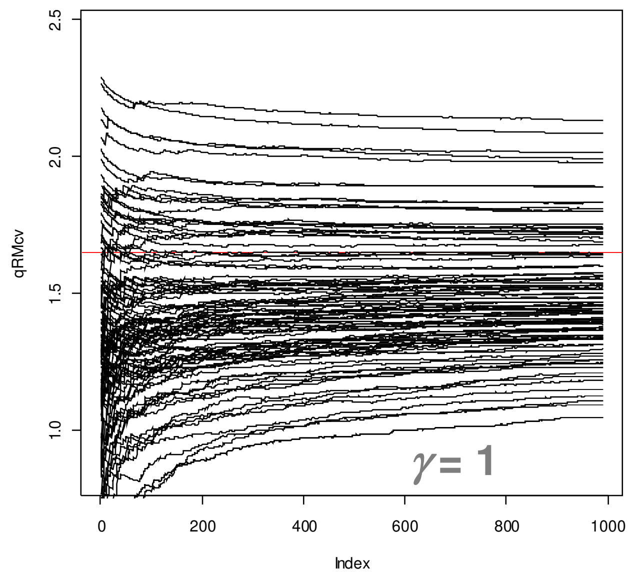
$R = 1000$

(repetitions of the RM algo)



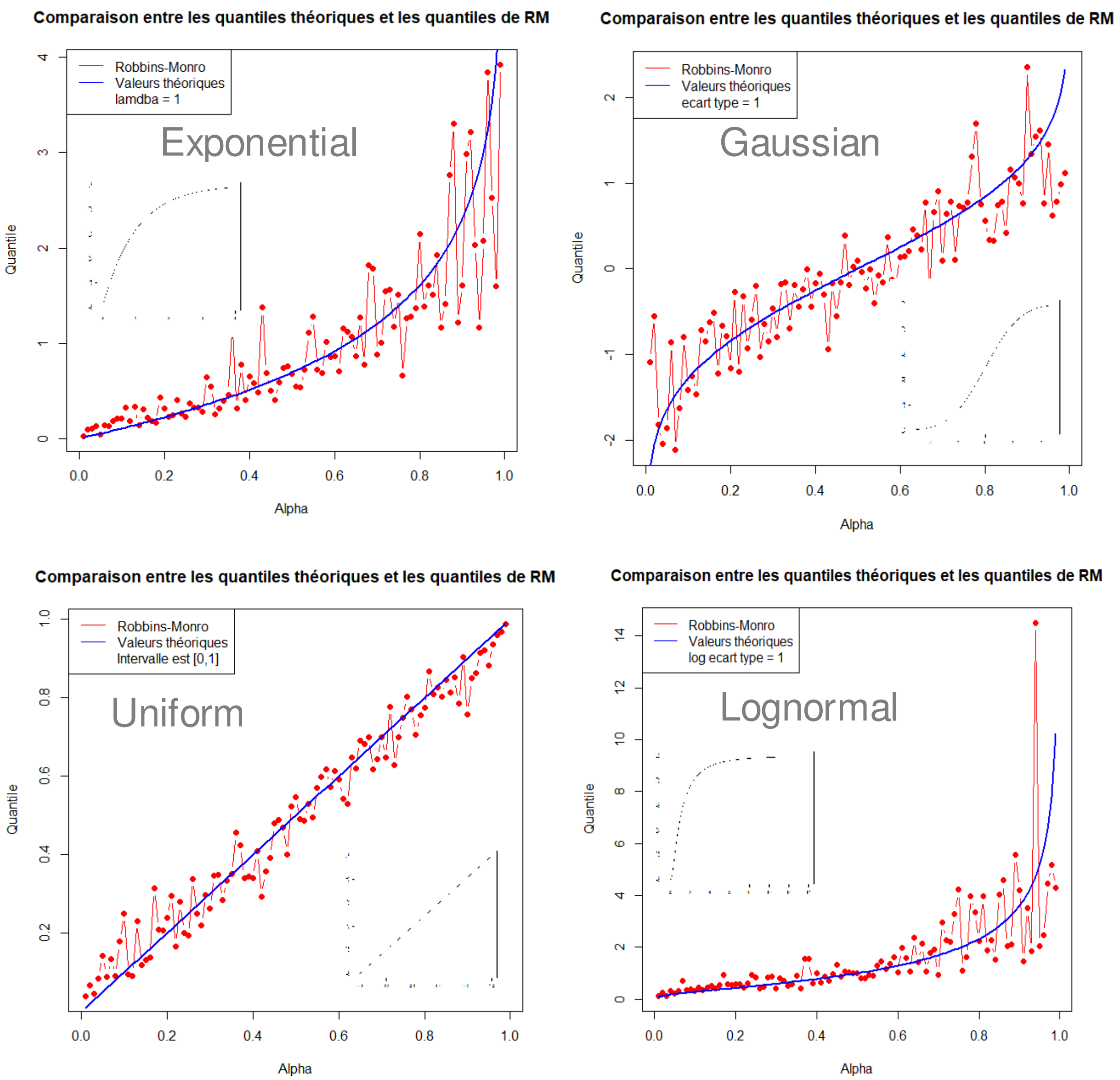






For other distributions of Y, best results are obtained with different γ values:
$\gamma \sim 1$ for uniform pdf; $\gamma \sim 0.6$ for exponential pdf; etc.

Issues due to the distrib. fct behaviour at α



N = 1000

γ = 1 and C = 1

Convergence difficulties appear at quantile levels corresponding to slow variations zone of the underlying distribution fct

Not shown: optimal γ values differ for different α values

Choice of a moving γ

$$q_{n+1} = q_n - \frac{1}{n^{\gamma(n)}}(1_{y^{(n+1)} \leq q_n} - \alpha)$$

- **A simple first idea:**

Linear evolution of γ (n) between 0.1 and 1 along the iterations of RM

$$\gamma(n) = 0.1 + 0.9\frac{n-1}{N-1}$$

$\Rightarrow$ **Algorithm that we call « Sequential RM »**

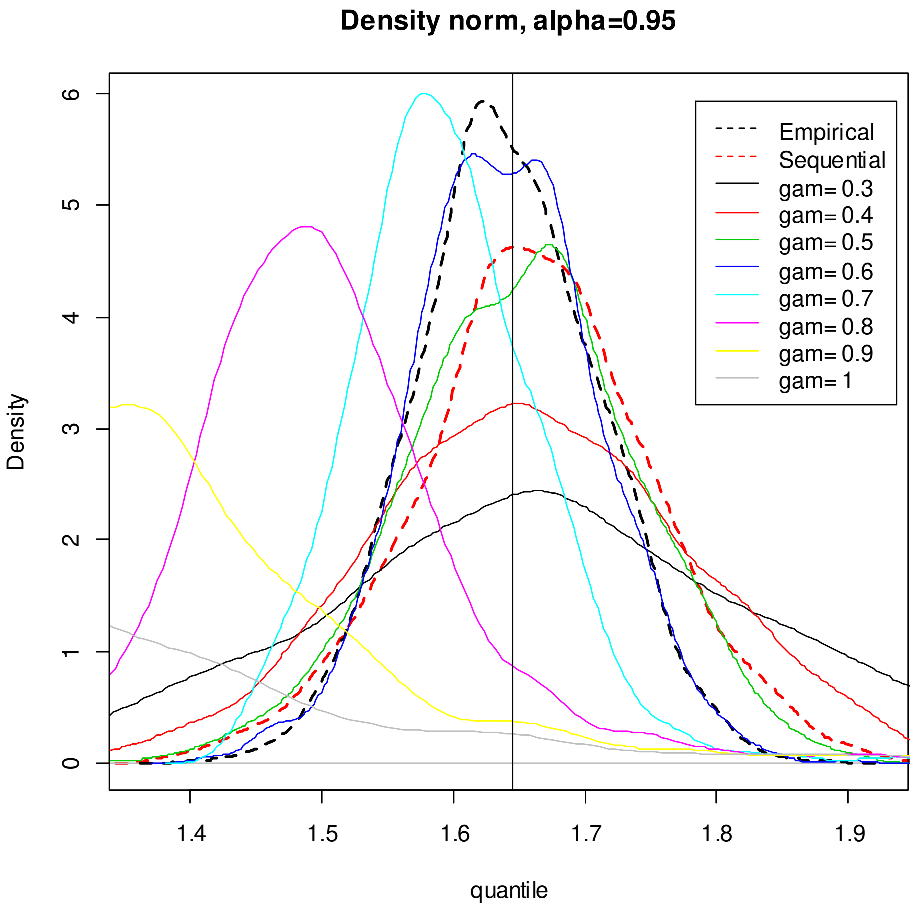
Asymptotical convergence is guaranteed if $\gamma(n) \leq 0.5$ for a finite number of iterations

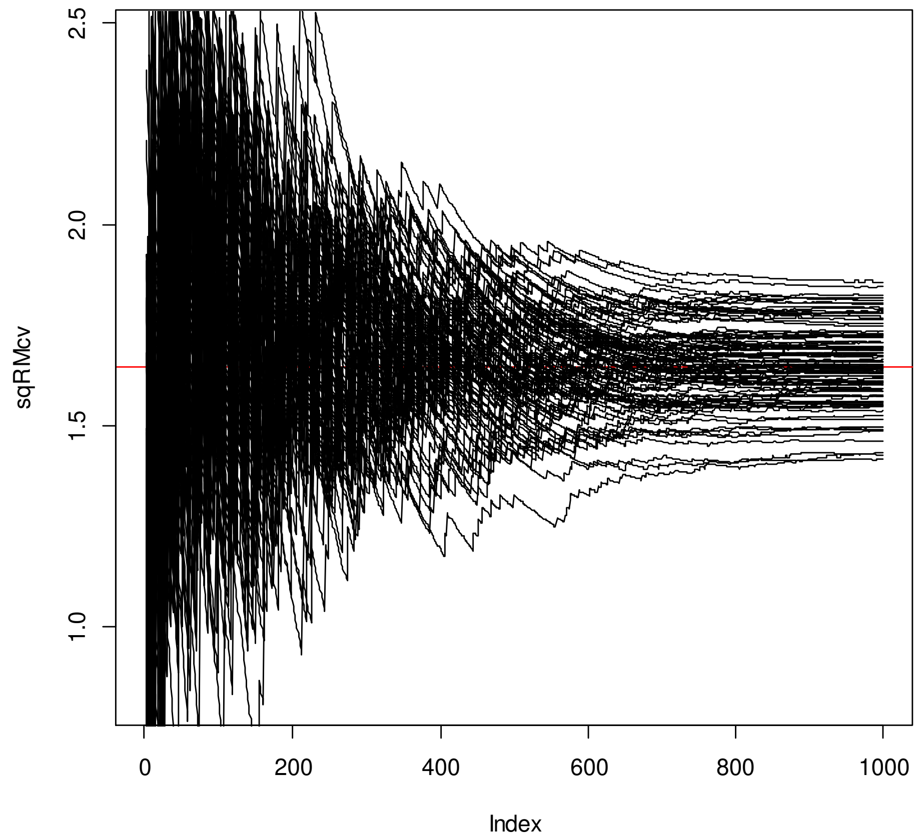
$$\frac{C}{n^{\gamma(n)}} \text{ decreases to } 0, \sum_{n\geq 1}\frac{C}{n^{\gamma(n)}} = \infty \text{ and } \sum_{n\geq 1}\left(\frac{C}{n^{\gamma(n)}}\right)^2 < \infty$$

Example on the Gaussian distribution N(0,1)

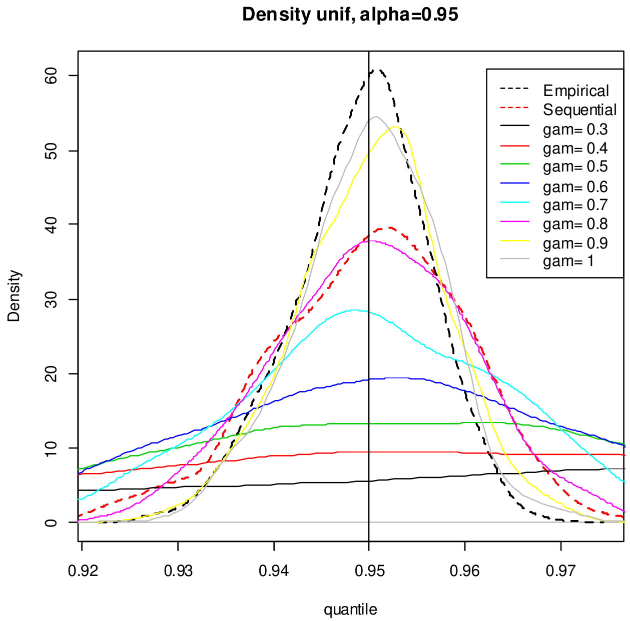
$C = 1$, $N = 1000$ and $\alpha = 0.95$

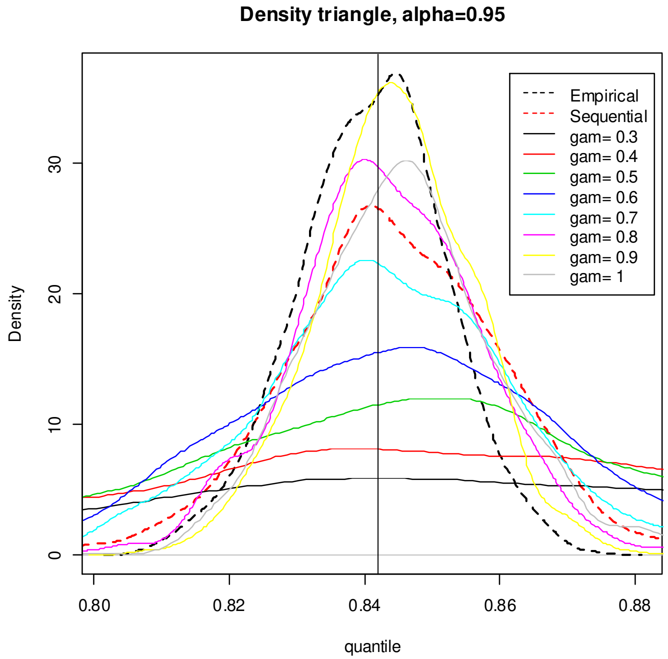
$R = 1000$ (repetitions of the RM algo)



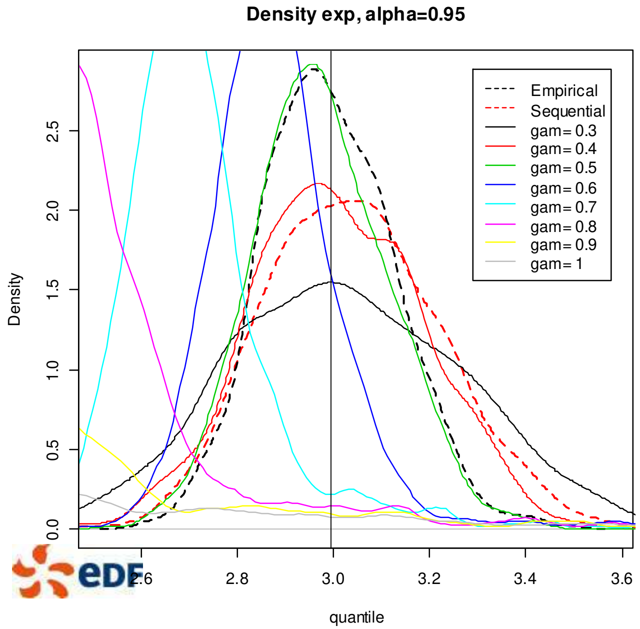


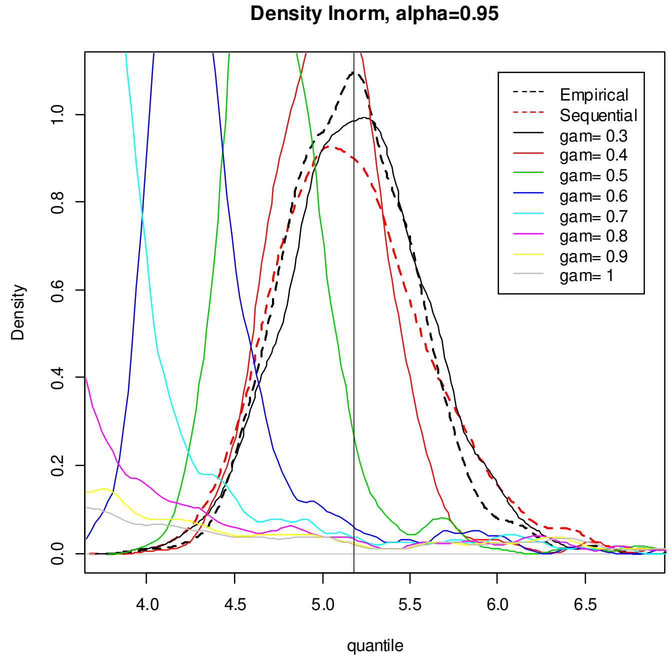
Numerical tests with other distributions for Y





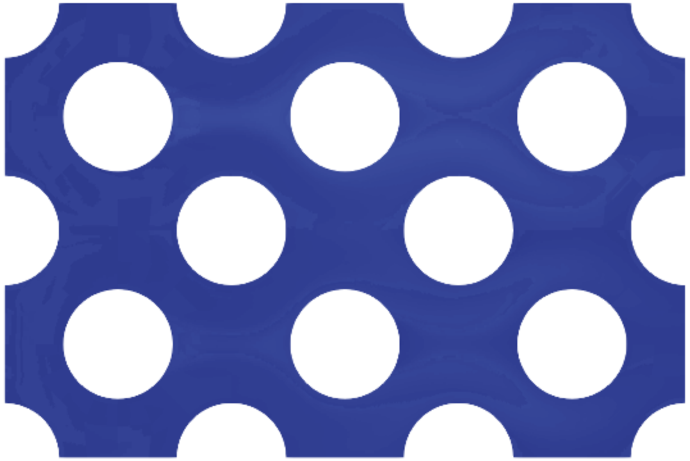
$$C = 1$$
$$N = 1000$$
$$\alpha = 0.95$$
$$R = 1000$$



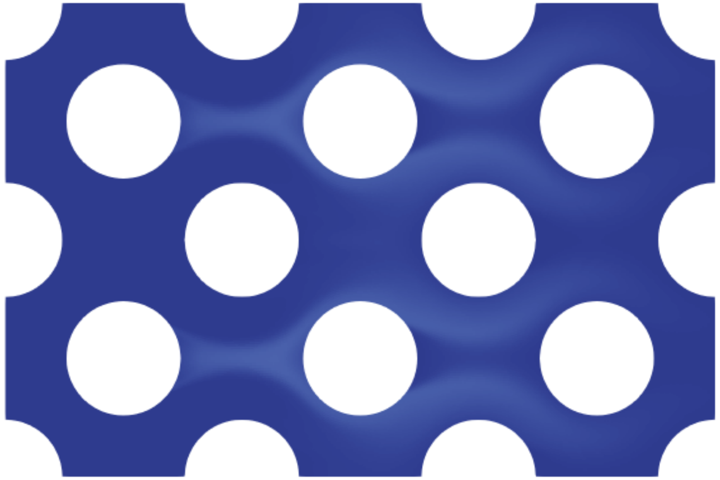


Results on the CFD application (1/2)

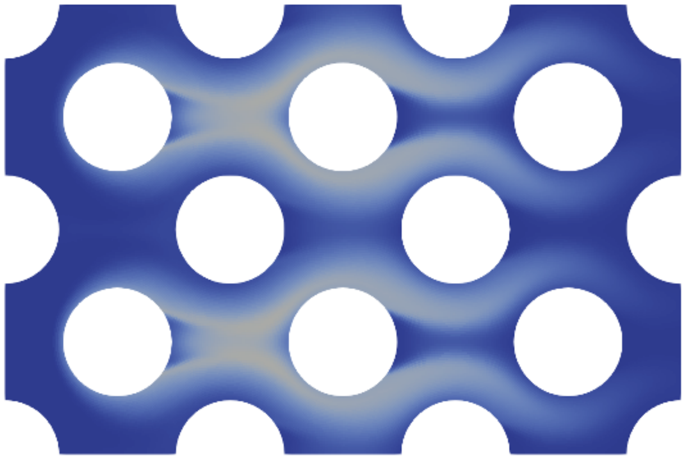
80th time-step over 100



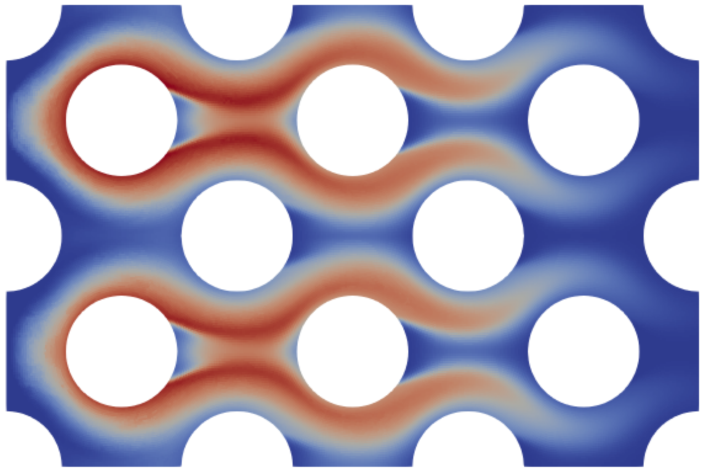
5%-quantile



25%-quantile



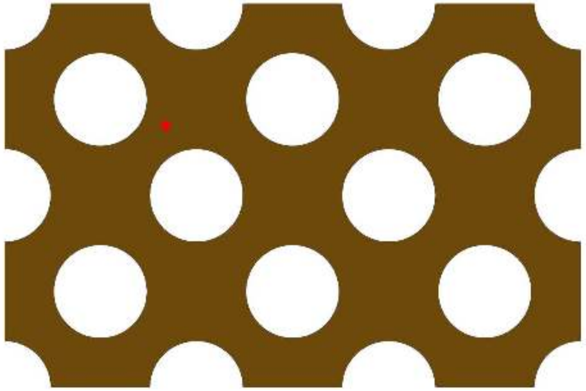
75%-quantile

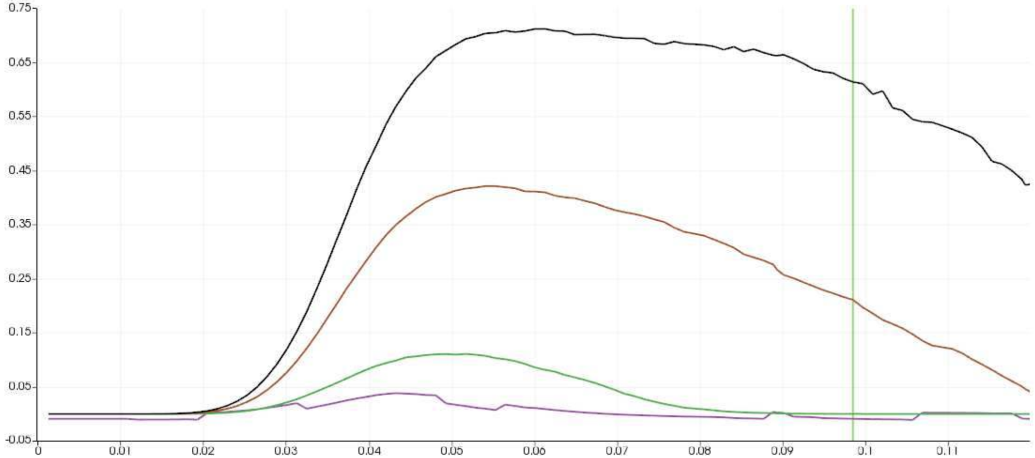


95%-quantile

Results on the CFD application (2/2)

Temporal evolution of the quantile at one spatial location





Quantiles of order $\alpha = 0.05, 0.25, 0.75$ and 0.95

Another thermal-hydraulic test case

PWR scenario:

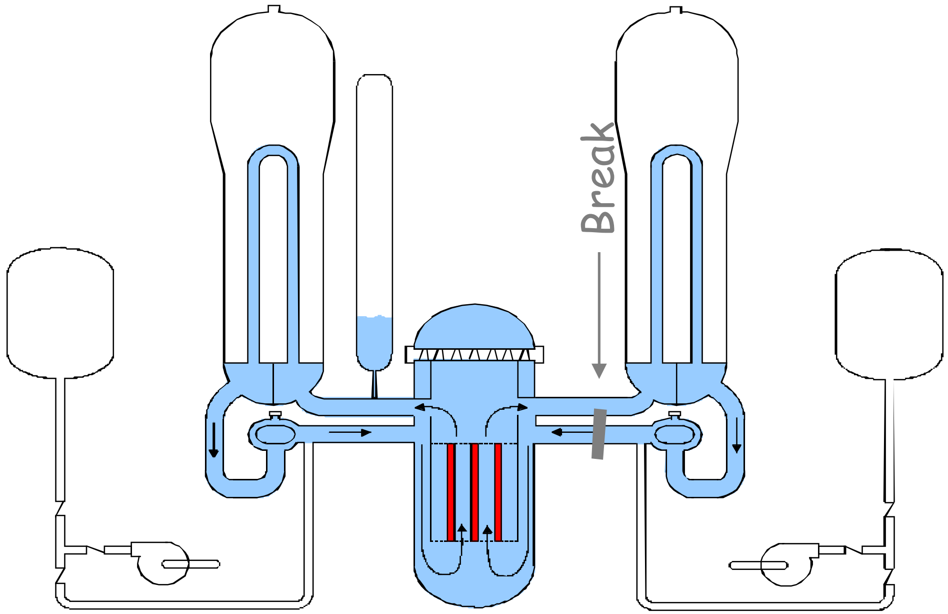
Loss of primary coolant accident due to a break in cold leg

Variable of Interest :

Second peak of cladding temperature (PCT) = scalar output

***p* (~ 100) input random variables :**

Critical flowrates, initial/boundary conditions, phys. eq. coef., …

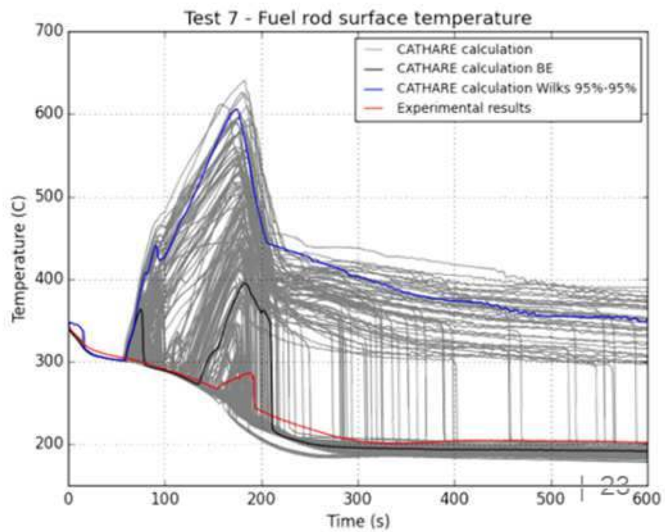


<u>Modelled using **CATHARE code**</u>:

- Models complex thermal-hydraulic phenomena
- Uncertain inputs

$\Rightarrow$ Exploration with Monte Carlo methods

- Large CPU cost for one code run (> 1 hour)

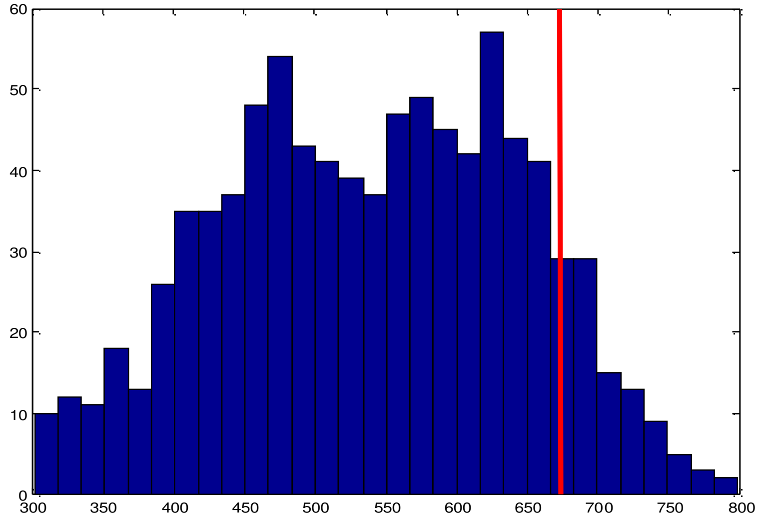


Quantity of Interest (QoI) :

High-order quantile (e.g. 90%, 95%...)

Classical uncertainty propagation results

N **= 889 (Monte-Carlo sample, applying the pdf of the inputs) –** ***p*** **= 96 variables**

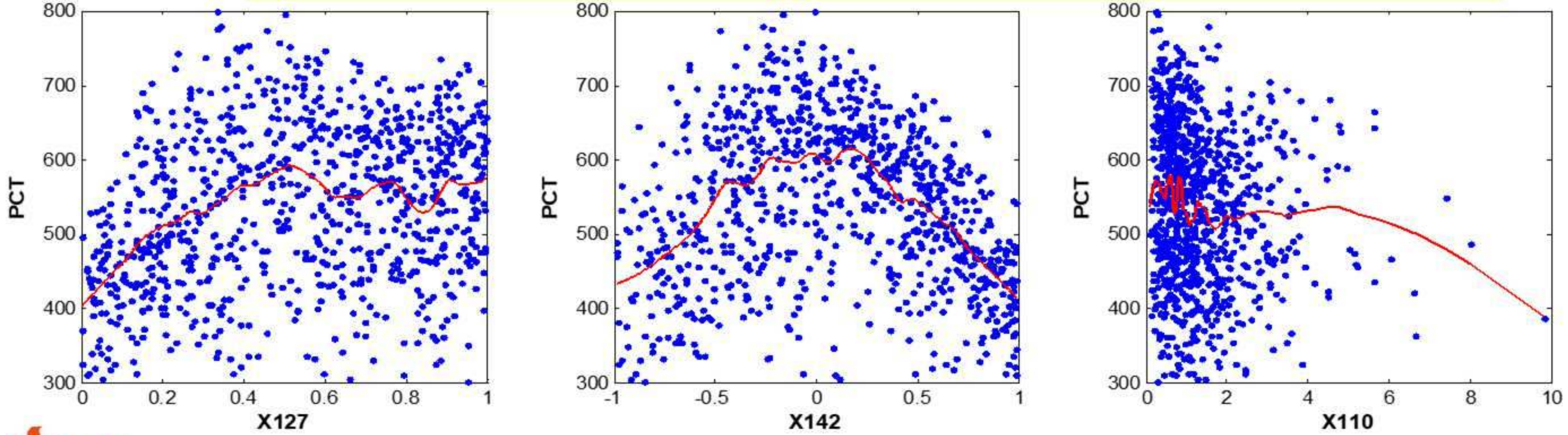


Analysis of the 889 PCT outputs (in °C)

Empirical quantile 90%: $q^{0.9}$ ~ 673

Empirical quantile 95%: $q^{0.95}$ ~ 703

Scatter plots with 1-D local polynomials for trends

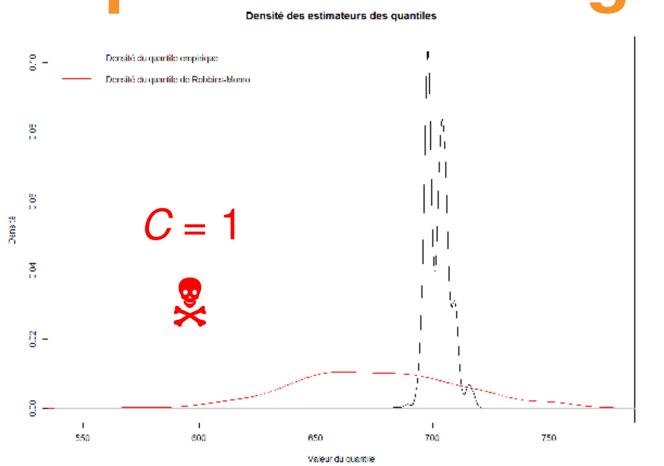


[Marrel et al., 2020]

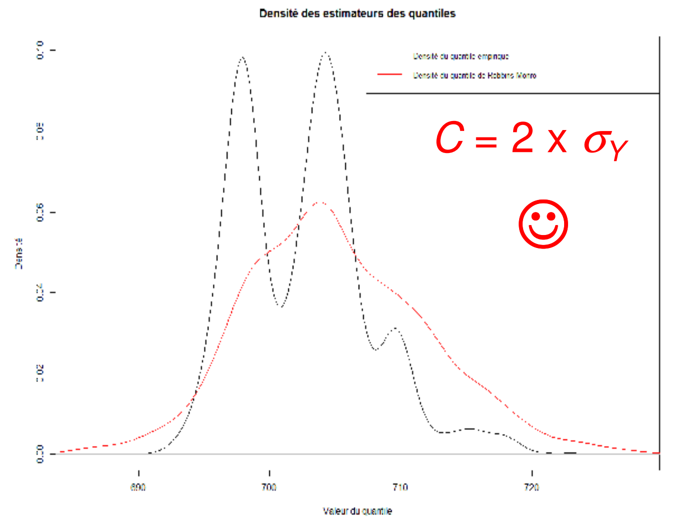
Thermalhydraulic application - Sequential RM algo

$N = 889$ - $\alpha = 0.95$ - γ = linear profile - $R = 100$

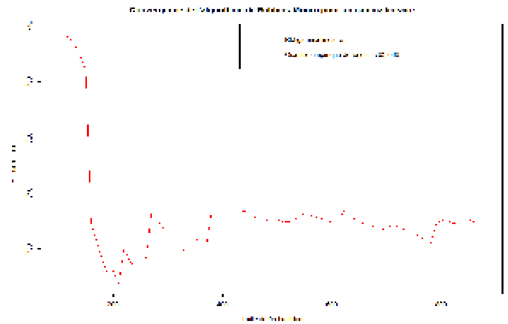
We imagine that we receive the output Y values sequentially (on-the-fly)
(as we have access to the full sample, we can repeat the RM algo bu using R bootstrap samples)



Indeed, at the beginning of the RM algo, perturbations for quantile updating have to be of the order of the Y dispersion



However, this dispersion is unkown in practice



Work (under progress): adaptive tuning of C

Goal: approximation of the quantile function of Y by estimating, at each iteration, all the α–quantiles (α being finely discretized between 0.05 and 0.95)

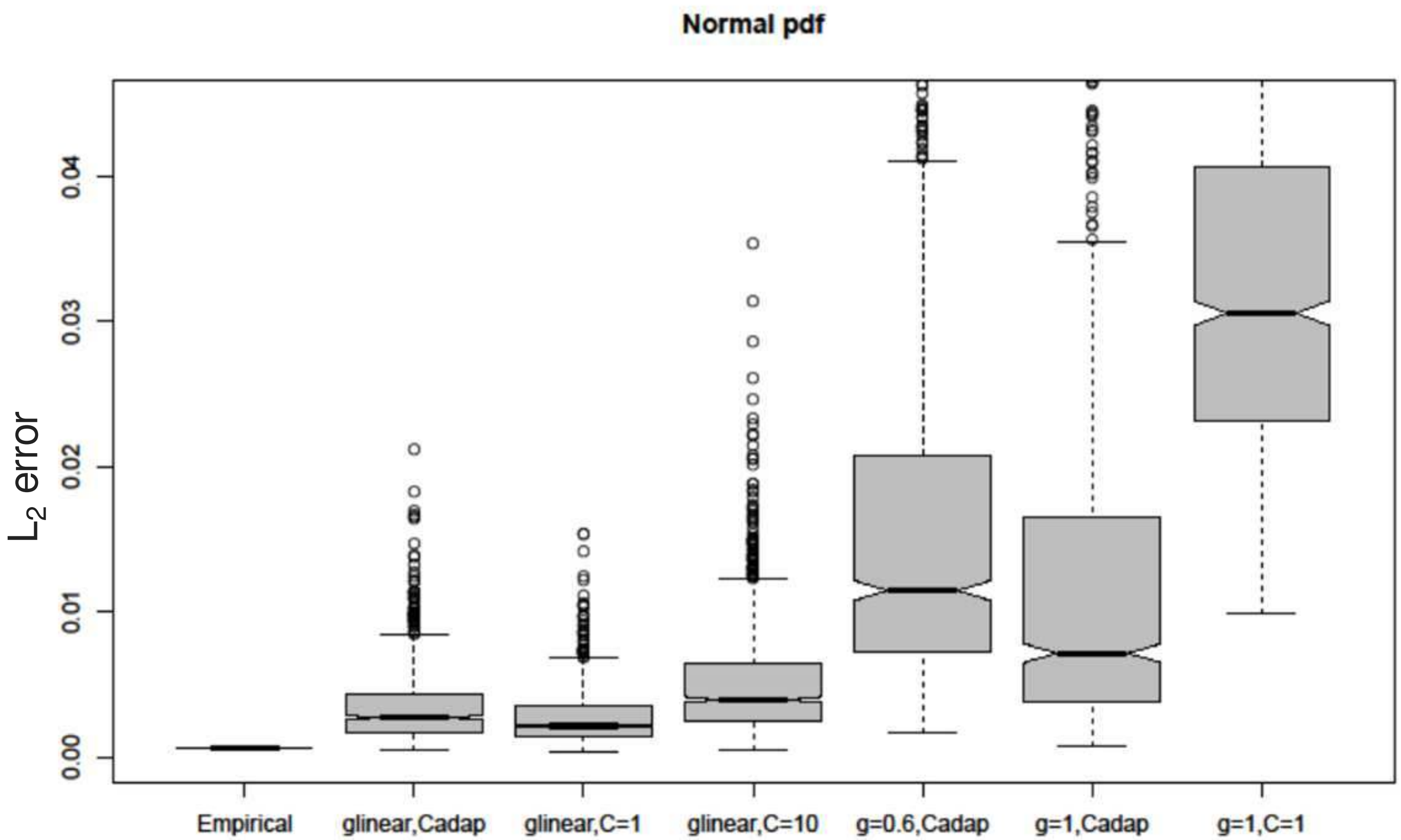
$$q_{n+1}^{\alpha} = q_n^{\alpha} - \frac{C(n)}{n^{\gamma(n)}}\left(1_{y^{(n+1)} \le q_n^{\alpha}} - \alpha\right) \text{ for } n = 1,\dots,N, \forall\, \alpha \in \{\alpha_{\min},\dots,\alpha_{\max}\}$$

$$\gamma(n) = 0.5 + 0.5\frac{n-1}{N-1} \text{ and } C(n) = \left|q_n^{\alpha_{\max}} - q_n^{\alpha_{\min}}\right|$$

$$q_1^{\alpha} = y^{(1)} \text{ and } C(1) = \left|y^{(2)} - y^{(1)}\right|$$

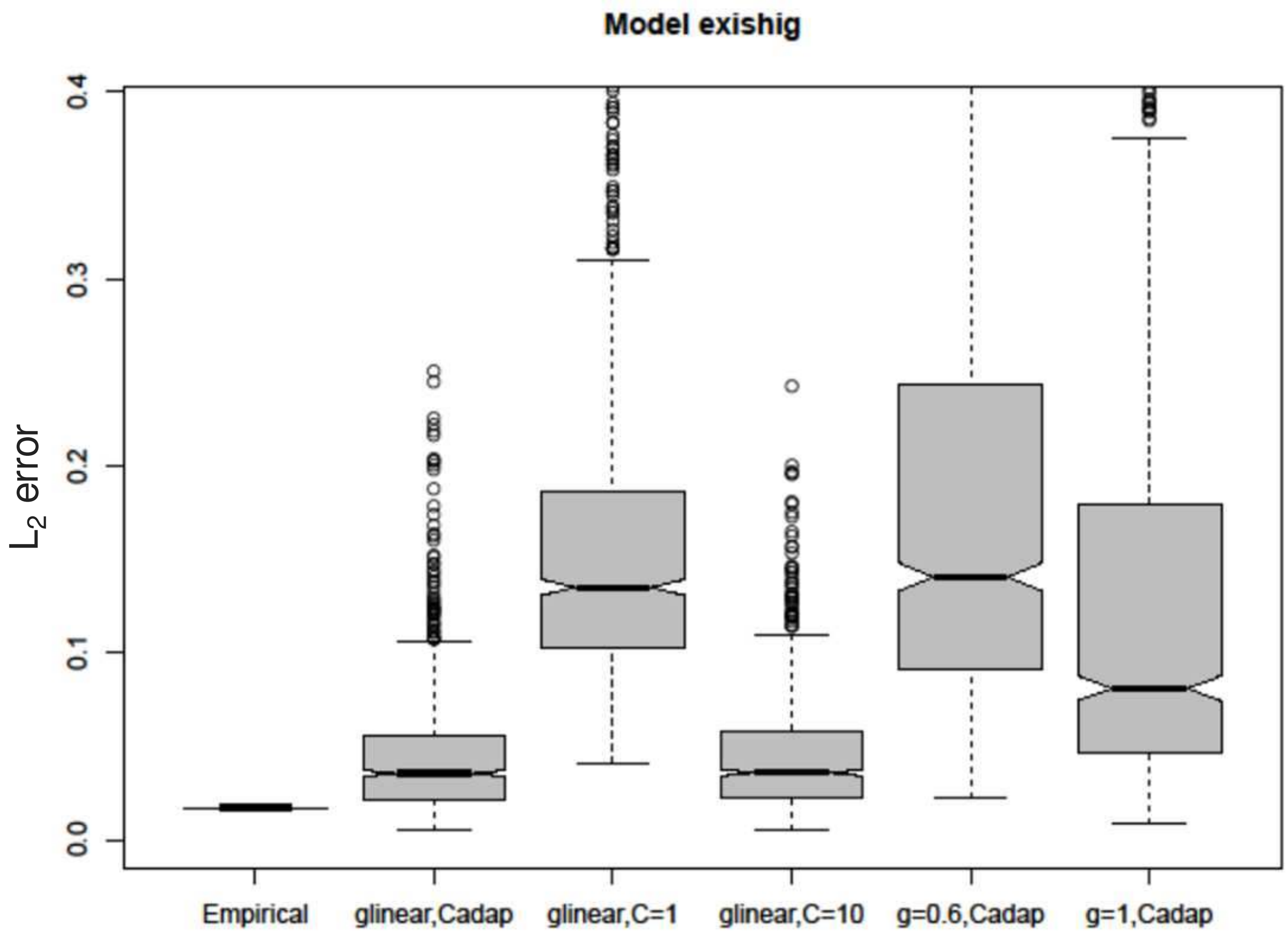
Numerical tests (1/3) – Gaussian distribution N(0,1)

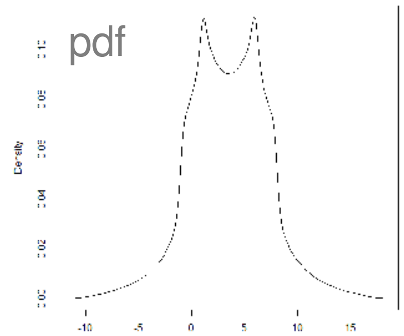
- $N = 1000$; $\alpha = \{0.05, 0.06, \ldots, 0.94, 0.95\}$; $R = 1000$ repetitions of the estimation process
- We give the distribution of an error metric (L_2-distance between the exact quantile function and the estimated one)

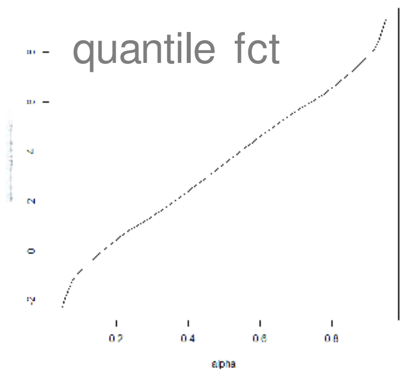


Numerical tests (2/3) – Ishigami function

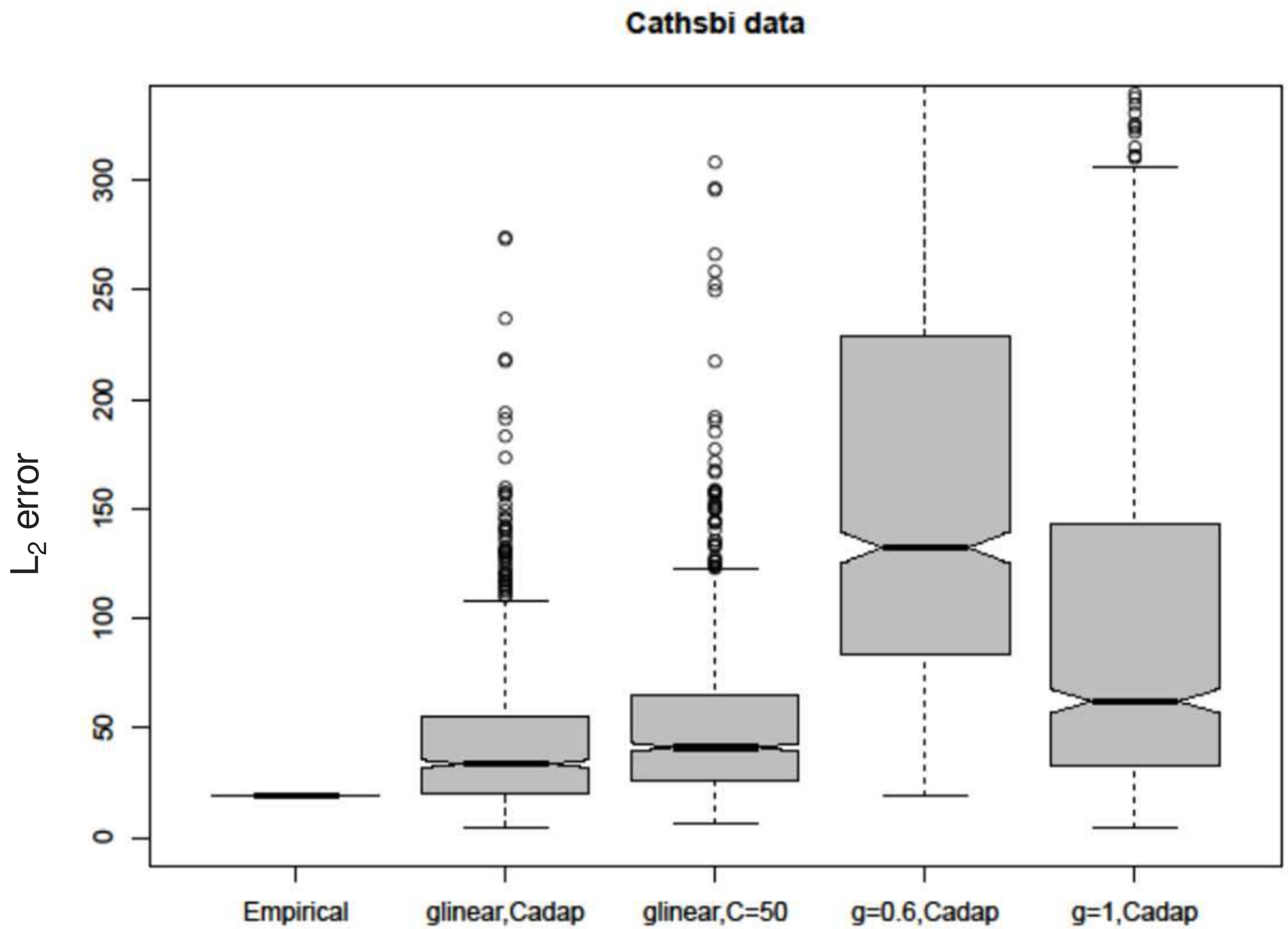
$Y = \sin(X_1) + 7\sin(X_2)^2 + 0{,}1X_3^4\sin(X_1)$ with $X_i \sim U[-\pi,\pi]\ \forall i = 1{,}2{,}3$







Numerical tests (3/3) – Thermalhydraulic application



Conclusion

In transit quantile estimation and sensitivity analysis with one-pass statistics

- No intermediate files – software properties: elastic, fault tolerant, adaptive
- Ubiquitous spatio-temporal statistics (Sobol' indices and quantile function)

Software: https://melissa-sa.github.io/ - OpenTURNS module under development

Current works and perspectives:

- Improving the robustness of the RM algo and tests on real applications
- Giving access to confidence intervals on estimates (no needs to specify N)
- Iterative dimension reduction and metamodeling

References:

- B. Bercu, ETICS 2019 lectures, www.gdr-mascotnum.fr/etics.html
- T. Terraz, A. Ribes, Y. Fournier, B. Iooss and B. Raffin. Large scale in transit global sensitivity analysis avoiding intermediate files, *Conference SC17,* November 2017
- A. Ribes, T. Terraz, B. Iooss, Y. Fournier and B. Raffin, Large scale in transit computation of quantiles for ensemble runs, *Preprint,* *https://hal.inria.fr/hal-02016828*
- B. Iooss, Estimation itérative en propagation d'incertitudes : réglage robuste de l'algorithme de Robbins-Monro, In preparation

Thanks for your attention

http://www.gdr-mascotnum.fr/

MASCOT 2020 Meeting

The MASCOT NUM 2020 meeting is organized from May 4th to May 7th, 2020, by Peter Challenor (univ. of Exeter - UK), Céline Helbert (Ecole Centrale de Lyon) and Clémentine Prieur (univ. Grenoble Alpes) in Aussois (France)

Call of PhD students abstracts (deadline 31/01/2020)

Summer school ETICS2020 (CEA-EDF-ENS), October, 4-9, Ile d'Oléron, France

http://www.gdr-mascotnum.fr/etics.html

Prof. Josselin Garnier (Ecole Polytechnique, France)

Prof. Anne-Laure Fougères (Université Claude Bernard Lyon 1) - Extreme value

Prof. Robert B. Gramacy (Virginia Tech) - Advances in Gaussian process modelling