VALIDATION OF NUMERICAL MODELS BY STATISTICS

Mathieu Couplet

EDF R&D, département PRISME

(Performance, Risque Industriel et Surveillance pour la Maintenance et l'Exploitation)

Contents

Contents

Terminology

According to The American Society of Mechanical Engineers (2006) :

> ***Verification*** *: the process of determining that a computational model accurately represents the underlying mathematical model and its solution.*

- Solve the equations right -

> ***Validation*** *: the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.*

- Solve the right equations -

> ***Calibration*** *: the process of adjusting physical modeling parameters in the computational model to improve agreement with experimental data.*

Remark. Possible distinction between **code verification** (unitary tests, check of the numerical scheme convergence, etc.) and **solution verification** (ensuring that the mesh is thin enough, check of the input data, etc.).

The American Institute of Aeronautics and Astronautics (1998) provides the following definition :

> ***Prediction*** *: use of a computational model to foretell the state of a physical system under conditions for which the computational model has not been validated.*

Concepts and notation

- SRQ = *System Reponse Quantity* (Oberkampf and Roy, 2010) : physical quantity of interest considered during the *validation*
- $\mathcal{U} \subset \mathbb{R}^L$ où $L \leq 4$ spatio-temporal domain, e.g. $\mathcal{U} = \mathcal{S} \times [0\,;T]$ where $\mathcal{S} \subset \mathbb{R}^3$ is a physical domain and T is a duration
- $x \in \mathcal{X} \subset \mathbb{R}^d$: d scalars on which the SRQ depends (features of the system of of its environment), that is *the experimental conditions*
- *Real* SRQ : $r(x,u)$
- *Measured* SRQ : $m(x,u)$; $m(x,u) - r(x,u)$ unknown measurement error
- *Theoretical* SRQ (mathematical model) : $v(x,u)$
- *Numerical* SRQ (numerical model) : $y(x,u)$
- Possible natural randomness : $R(\omega \in \Omega, x, u)$, hence $V(\omega \in \Omega, x, u)$...

Example of an ambitious approach

Standard of The American Society of Mechanical Engineers (2009)

$$v(x) - r(x) = \underbrace{[y(x+\delta_x) - m(x)]}_{\substack{\text{observed gap} \\ \text{d(x) between} \\ \text{measurement and simulation} \\ \text{(no uncertainty)}}} - \underbrace{[y(x+\delta_x) - v(x)]}_{\substack{\text{uncertain numerical} \\ \text{errors} \\ \varepsilon_y + \varepsilon_x}} + \underbrace{[m(x) - r(x)]}_{\substack{\text{uncertain measurement} \\ \text{errors} \\ \varepsilon_m}}$$

- $\varepsilon_y = y(x) - v(x)$: subject of the *verification* ;
- $\varepsilon_x = y(x+\delta_x) - y(x)$: uncertainty to propagate through y.

Principle : modelling the uncertainty about $(\varepsilon_y, \varepsilon_x, \varepsilon_m)$ *by a joint distribution in order to estimate a confidence interval of a given level* α *for* $v(x) - r(x)$.
Assumption of independent gaussian errors :

$$\left[d(x) - \beta\sqrt{{\sigma_y}^2 + {\sigma_x}^2 + {\sigma_m}^2}; \quad d(x) + \beta\sqrt{{\sigma_y}^2 + {\sigma_x}^2 + {\sigma_m}^2} \right]$$

where β is the quantile of order $1 - (\alpha/2)$ of $\mathcal{N}(0;1)$.

Calibration-validation : formalization

- Assumption of a verified numerical model : $Y \sim V$ (e.g. $y = v$).

- Model of $R(x)$ (by abuse of notation) :

$$Y : \underbrace{(x, \theta_y)}_{\text{numerical model inputs}} \mapsto \underbrace{Y(\omega \in \Omega, u \in \mathcal{U}, x, \theta_y)}_{\text{multivariate distribution}}$$

  where θ_y are uncertain parameters assumed invariant from an experiment to another one.

- N data (measurements) associated to conditions x^n, $1 \leq n \leq N$: $m^n(u \in \mathcal{U})$ realizations of $M(x^n, u \in \mathcal{U})$ such that $M(x^n, u \in \mathcal{U}) - R(x^n, u \in \mathcal{U})$ is the measurement error.

- The exact conditions x^n are generally unknown ! This leads to group the experiments which are supposed to share the same conditions (repetitions) and to associate each of the K groups with some best-estimate conditions $\hat{x}^k$ (arbitrary indeed) and a prior on the discrepancy between $\hat{x}^k$ and the true conditions.

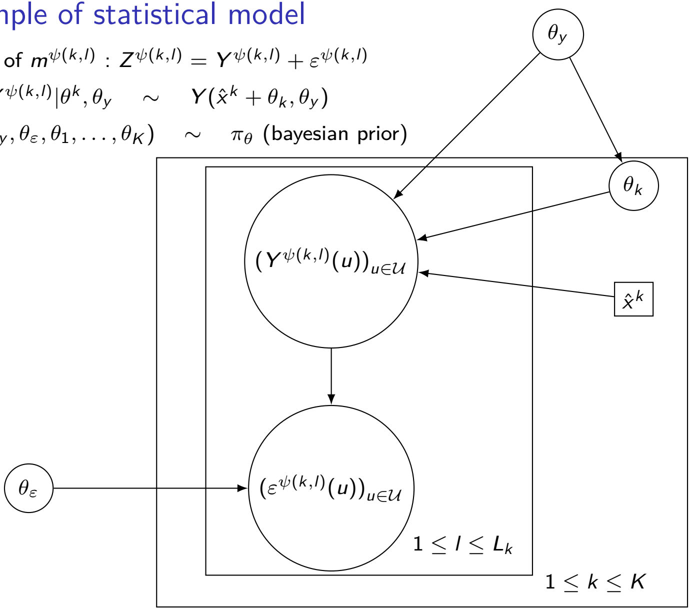
Calibration-validation : formalization

Example of statistical model

Model of $m^{\psi(k,l)}$: $Z^{\psi(k,l)} = Y^{\psi(k,l)} + \varepsilon^{\psi(k,l)}$

with $Y^{\psi(k,l)}|\theta^k, \theta_y \quad \sim \quad Y(\hat{x}^k + \theta_k, \theta_y)$

$\theta = (\theta_y, \theta_\varepsilon, \theta_1, \dots, \theta_K) \quad \sim \quad \pi_\theta$ (bayesian prior)



Pure validation

Calibration already done

- $\theta = \hat{\theta}$: pure validation.
- $\theta \in \Theta$ or $\theta \sim \pi_\theta$ (non-Dirac prior) : validation-calibration.

Pure validation in theory

Let $\mathcal{V} \in \mathcal{X}$ be the domain of validation :

$$\forall x \in \mathcal{V} \qquad \Big(R(x,\, u)\Big)_{u \in \mathcal{U}} \qquad \sim \qquad \Big(Y(\hat{x},\, u,\, \hat{\theta}_y)\Big)_{u \in \mathcal{U}}$$

Pure validation in practice

$$\forall k \in \{1,\, \cdots,\, K\} \qquad \underbrace{\Big(M_1(x^{\psi(k,1)}),\, \cdots,\, M_P(x^{\psi(k,1)})\Big)}_{L_k\text{-sample available}} \quad \sim \quad \underbrace{\Big(Z_1(\hat{x}^k,\, \hat{\theta}),\, \cdots,\, Z_P(\hat{x}^k,\, \hat{\theta})\Big)}_{\text{may incorporate a CPU-consuming black-box model...}}$$

Un jour, j'irai vivre en Théorie, car, en Théorie, tout se passe bien !

La différence entre théorie et pratique ? En théorie c'est pareil.

Pure validation

Pure validation in practice

$$\forall k \in \{1, \cdots, K\} \qquad \underbrace{\left(M_1(x^{\psi(k,1)}), \cdots, M_P(x^{\psi(k,1)})\right)}_{L_k\text{-sample available}} \sim \underbrace{\left(Z_1(\hat{x}^k, \hat{\theta}), \cdots, Z_P(\hat{x}^k, \hat{\theta})\right)}_{\text{may incorporate a CPU-consuming black-box model...}}$$

data (H_0)

Statistical tools

- Hypothesis statistical testing (e.g. p-value),
- Estimation of discrepancy/divergence between distributions (can be used as a test statistic),
- Detection of outliers,
- graphics (qqplot).

Pure validation

Pure validation in practice

$$\forall k \in \{1, \cdots, K\} \qquad \underbrace{\left(M_1(x^{\psi(k,1)}), \cdots, M_P(x^{\psi(k,1)})\right)}_{L_k\text{-sample available}} \sim \underbrace{\left(Z_1(\hat{x}^k, \hat{\theta}), \cdots, Z_P(\hat{x}^k, \hat{\theta})\right)}_{\text{may incorporate a CPU-consuming black-box model...}}$$

M $\qquad$ Z

Statistical tools

- Hypothesis statistical testing (e.g. p-value),
- Estimation of discrepancy/divergence between distributions (can be used as a test statistic),

$$\mathcal{D}(M, Z) \approx 0?$$

- Detection of outliers,
- graphics (qqplot).

Designing the experiments

How to choose the experimental conditions ?

Two interesting approaches

- Classical screening designs for physical experiments :
  - Objective : identifying the conditions x_i (components of x) whose variations have a significant impact on the SRQ.
  - A solution : fractorial factorial designs.
- Space filling designs
  - Objective : « covering » the domain $\mathcal{V}$.
  - A solution : optimized Latin Hypercube Designs.

Another approach

Carrying out a sensitivity analysis of the (non validated !) numerical model.

Designing the experiments

The following figure illustrates two-dimensional Simple Random Sampling (SRS), Latin Hypercube Sampling (LHS) and discrepancy-optimised LHS with $n = 10$.

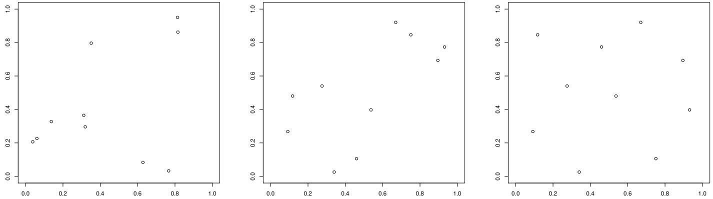


Figure: **Left** : SRS. **Middle** : LHS. **Right** : discrepancy-optimised LHS based on the previous LHS (design at the middle).

A simple situation

Hypothesis

Scalar deterministic SRQ r and unbiased measurements.

A well-known result

Let $(M^1, \ldots, M^n)$ be a sample of size $n > 1$ of iid variables such that $\mu_M = \mathbb{E}\left[M^k\right] < +\infty$ and $\mathbb{V}\left[M^k\right] < +\infty$ and

$$\overline{M_n} = \frac{1}{n}\sum_{k=1}^{n} M^k \text{ and } {S_n}^2 = \frac{1}{n}\sum_{k=1}^{n} \left(M^k\right)^2 - \overline{M_n}^{\,2},$$

then a confidence interval of level α of μ_M is

$$\mathcal{I}_n = [\overline{M_n} - u_{1-\alpha/2}\,\frac{S_n}{\sqrt{n}}\ ;\overline{M_n} + u_{1-\alpha/2}\,\frac{S_n}{\sqrt{n}}]$$

where $u_{1-\alpha/2}$ is the quantile of order $1 - \frac{\alpha}{2}$

- of the distribution $\mathcal{N}(0\,;1)$ and this confidence interval is asymptotic (large n)
- of the distribution $t(n-1)$ (Student distribution with $n-1$ degrees of freedom) if the distribution of the M_i is gaussian.

A simple situation

Hypothesis

n measurements $M^k(x)$, $1 \leq k \leq n$, of $r(x)$ such that $\mathbb{E}\left[M^k(x)\right] = r(x)$.

Indicators proposed by Oberkampf and Barone (2006)

- Since $\mathbb{P}(r(x) - y(x) \in \mathcal{I}_n(x)) = 1 - \alpha$: $\overline{M_n(x)} - y(x)$ and $|\mathcal{I}_n(x)| = 2\, u_{1-\alpha/2} \dfrac{S_n(x)}{\sqrt{n}}$.
- Estimation by regression in $\mathcal{V}$:
  - global indicators :

$$\text{Volume}(\mathcal{V})^{-1} \int_{\mathcal{V}} \left| \frac{\mathbb{E}\left[M(x)\right] - y(x)}{\mathbb{E}\left[M(x)\right]} \right| \mathrm{d}x \quad \text{et} \quad \text{Volume}(\mathcal{V})^{-1} \frac{u_{1-\alpha/2}}{\sqrt{n}} \int_{\mathcal{V}} \left| \frac{S_n(x)}{\hat{M}(x)} \right| ;$$

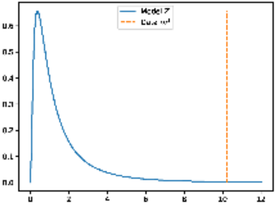
  - maximal error indicators :

$$\left| \frac{r(x^*) - y(x^*)}{r(x^*)} \right| = \left| \frac{\mathbb{E}\left[M(x^*)\right] - y(x^*)}{\mathbb{E}\left[M(x^*)\right]} \right| \quad \text{and} \quad \frac{u_{1-\alpha/2}}{\sqrt{n}} \left| \frac{S_n(x^*)}{\hat{M}(x^*)} \right|$$

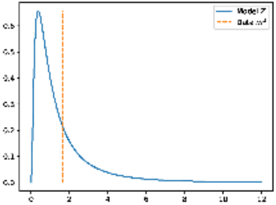
  where $x^* = \arg\max_{x \in \mathcal{V}} \left| \frac{r(x) - y(x)}{r(x)} \right|$.

Hypothesis testing (p-value)

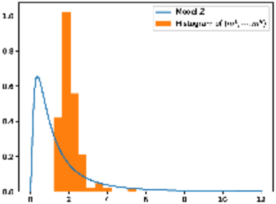
Different situations (monovariate case)



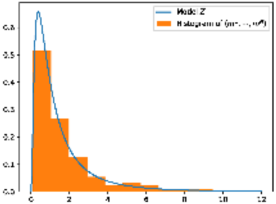
Wish to reject the model



No particular wish

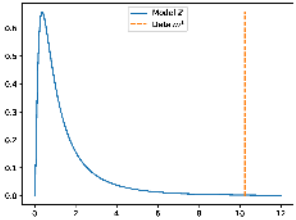


Strong wish to reject the model

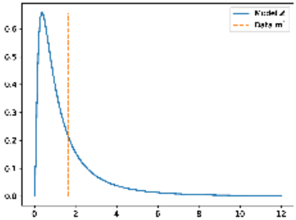


Wish to accept the model

Hypothesis testing (p-value)



Shallow data



Deep data

Let Ψ be a test statistic :

$$\begin{aligned} p\text{-}value &= \text{proba. to wrongly reject the model} \\ &\quad \text{(small value} \Rightarrow \text{wish of rejection)} \\ &= \mathbb{P}(\Psi(Z^1, \cdots, Z^N) \notin \underbrace{\mathcal{I}(m^1, \cdots, m^n)}_{\substack{\text{set of values} \\ \text{« less extreme »} \\ \text{than } \Psi(m^1, \cdots, m^N)}}) \end{aligned}$$

Left :

- $p\text{-}value = \mathbb{P}(Z > m^1)$ $(\Psi(Z) = Z)$
- The *p-value* corresponds to the order α of the quantile $q_Z^\alpha = m^1$: could be generalized to multivariate/functional data thanks to the notion of depth (López-Pintado and Romo, 2007).

Comparison of distributions

Examples of indicators

Let F_R and F_Y be two Cumulative Density Functions (CDF) ; the corresponding Probability Density Functions (PDF) are f_R and f_Y if they exist.

Area metric
[Ferson et al., 2008]

$$\|F_R - F_Y\|_{L^1} = \int_{\mathbb{R}} |F_R - F_Y|$$

- *Easy to estimate from empirical CDF*
- *Always well-defined and robust*
- *Same physical unit than r (scale variant)*

Relative entropy (divergence of Kullback-Leibler)

$$\mathcal{K}(F_R, F_Y) = \int_{\mathbb{R}} f_R \log(\frac{f_R}{f_Y})$$

$$\mathcal{K}(F_R, F_Y) = 0 \Rightarrow F_R = F_Y$$

- Not symmetric (notion of reference F_R)
- Issues : estimation of PDF and cases where $\mathcal{K}(F_R, F_Y) = +\infty$

Distance of Hellinger

$$\mathcal{H}(F_R, F_Y) = \int_{\mathbb{R}} (\sqrt{f_R} - \sqrt{f_Y})^2$$

$$\mathcal{H}(F_R, F_Y) = 0 \Rightarrow F_R = F_Y$$

- $\mathcal{H}(F_R, F_Y) \leq 1$
- Minor issue : estimation of PDF
- *Undimensional overlap indicator*

Comparison of distributions

Area metric (Ferson et al., 2008)

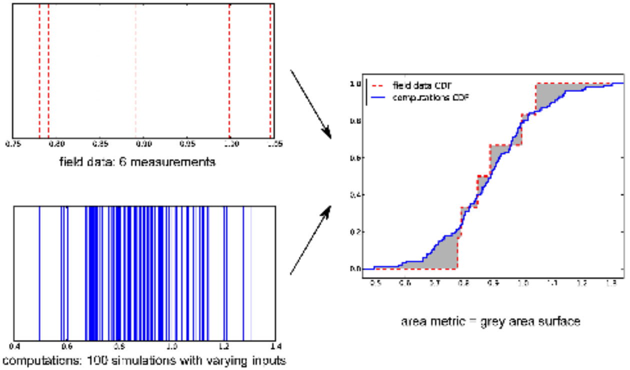


Illustration for $n = 6$ real responses and $m = 100$ simulations.

Comparison of distributions

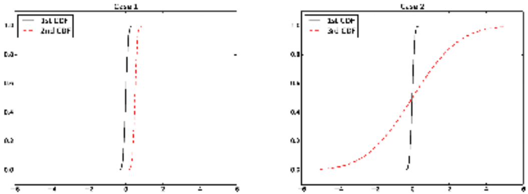
Area metric (Ferson et al., 2008)

- Ferson et al. (2008) propose a technique called *u-pooling* to aggregate several couples of samples to get a global indicator :
  - for different experimental conditions x
  - or for different times t^i, points u^i, *etc.*

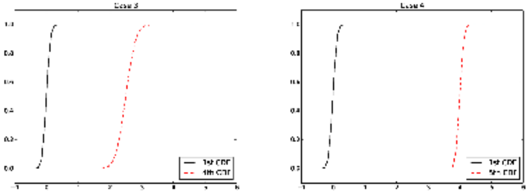
  however this technique is dangerous : multivariate two-sample test of Székely and Rizzo (2004) better for example.
- All indicators of distribution comparison are not equivalent.

Comparison of distributions

- Comparison with the Hellinger distance $\left(\mathcal{H}(F_R, F_Y) = \int_{\mathbb{R}} (\sqrt{f_R} - \sqrt{f_Y})^2\right)$:

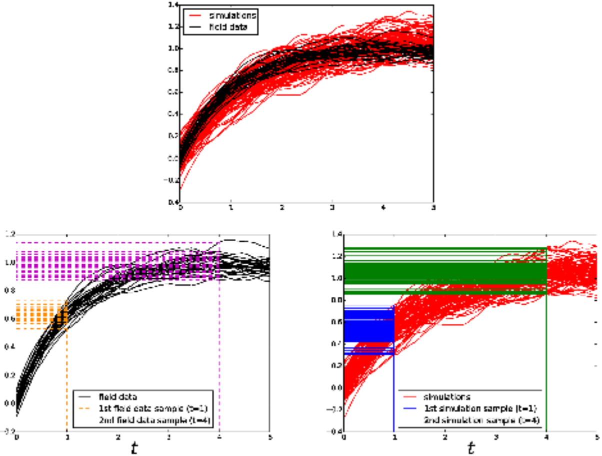


$$\underbrace{\mathcal{H}(\text{Case 1})}_{=1 \text{ (no overlapping)}} > \underbrace{\mathcal{H}(\text{Case 2})}_{\neq 1} \text{ and } \mathcal{A}(\text{Case 1}) < \mathcal{A}(\text{Case 2})$$



$$\mathcal{H}(\text{Case 3}) = \mathcal{H}(\text{Case 4}) = 1 \text{ and } \mathcal{A}(\text{Case 3}) < \mathcal{A}(\text{Case 4})$$

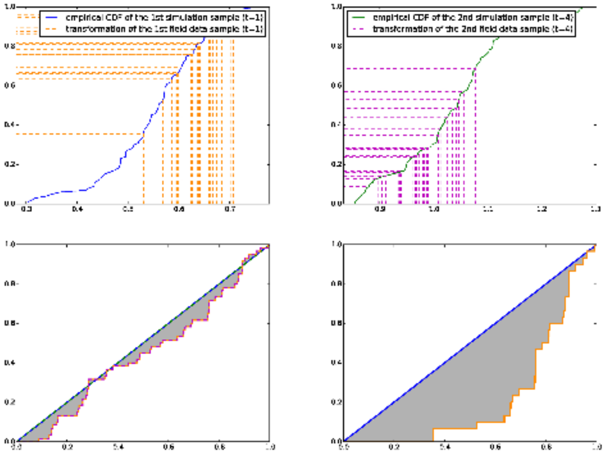
Illustration of the u-pooling procedure



Top : data $r^k(t^j)$, $1 \leq k \leq 30$, and $y^i(t^j)$, $1 \leq i \leq 100$, with $t^j = 0.05 \times j$, $1 \leq j \leq 100$.
Bottom : u-pooling only applied to $r^k(t^j)$ and $y^i(t^j)$ for $j \in \{20, 80\}$ ($t^{20} = 1$ and $t^{80} = 4$).

Comparison of distributions

Illustration of the u-pooling procedure



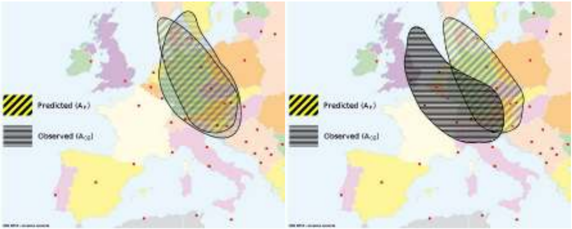
Top : transformation of the field samples according to the empirical CDF of the simulation samples ($G(x) = x$ for any $0 \leq x \leq 1$).

Bottom : u-pooled area metric (left) and area metric restricted to the transformed data at $t-1$ (right).

Contents

Figure of Merit in Space (FMS)

Prediction of the location of a plume of pollutant (concentration > threshold) :



IGN 2012 - Licence ouverte : produced from a map provided by the Institut National de l'information Géographique et forestière under the Etalab Open Licence (http://www.etalab.gouv.fr/licence-ouverte-open-licence)

$$FMS = \frac{\text{observed area} \bigcap \text{predicted area}}{\text{observed area} \bigcup \text{predicted area}}$$

Generalization (Warner et al., 2004)

Let $\mathcal{R}$ be a set of interest of the SRQ.

$$\mathcal{T} = \frac{\overline{\mathbb{1}_{(r,y)\in\mathcal{R}^2}(r,y)}}{\overline{\mathbb{1}_{r\in\mathcal{R} \text{ ou } y\in\mathcal{R}}(r,y)}} = \frac{\overline{\mathbb{1}_{(r,y)\in\mathcal{R}^2}(r,y)}}{\overline{\mathbb{1}_{(r,y)\in\mathcal{R}^2}(r,y)} + \overline{\mathbb{1}_{(r,y)\in\mathcal{R}\times\bar{\mathcal{R}}}(r,y)} + \overline{\mathbb{1}_{(r,y)\in\bar{\mathcal{R}}\times\mathcal{R}}(r,y)}}$$

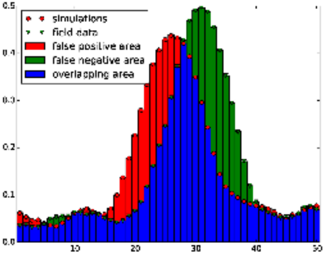
Indices of Tversky

Indices of Tversky : extensions (Warner et al., 2004)

$$T = \frac{A^{OV}\ (overlap)}{A^{OV} + A^{FN}\ (false\ negative) + A^{FP}\ (false\ positive)}$$

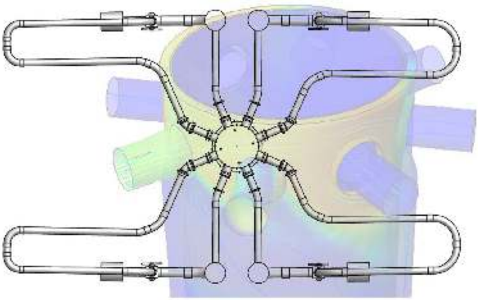
- Risk-weighted FMS : RWFMS $= \frac{A^{OV}\ (overlap)}{A^{OV} + \alpha A^{FN}\ (false\ negative) + \beta A^{FP}\ (false\ positive)}$
- Measure Of Effectiveness : MOE $= \left(\frac{A^{OV}}{A^{OB}}, \frac{A^{OV}}{A^{PR}}\right) = \left(1 - \frac{A^{FN}}{A^{OB}}, 1 - \frac{A^{FP}}{A^{PR}}\right)$

  avec $A^{OB} = A^{OV} + A^{FN}$ $(observed)$ et $A^{PR} = A^{OV} + A^{FP}$ $(predicted)$.
- Definitions not based on a set $\mathcal{R}$ of interest :
  - $A^{OV} = \overline{A_i^{OV}}$ where $A_i^{OV} = \min(r_i\ ; y_i)$
  - $A^{FN} = \overline{\max(r_i - A_i^{OV}\ ; 0)}$
  - $A^{FP} = \overline{\max(y_i - A_i^{OV}\ ; 0)}$



ROCOM

- Experimental data from project OCDE PKL2 (HZDR, Helmholtz Zentrum Dresden Rossendorf).



- Simulation performed with Code_Saturne by Arnaud Barthet (EDF).

ROCOM

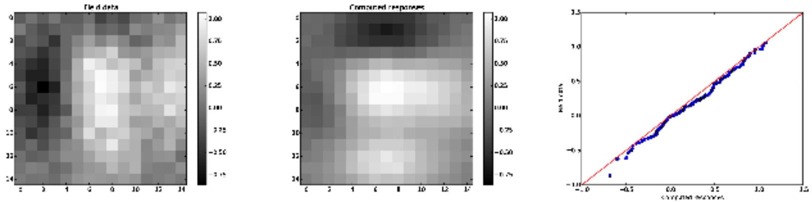
- Experimental data from project OCDE PKL2 (HZDR, Helmholtz Zentrum Dresden Rossendorf).


- Simulation performed with Code_Saturne by Arnaud Barthet (EDF).

Proposition

Because of the randomness of the physics, it may appear relevant only to check that the distribution of the fluid temperature in space and/or time is invariant (without consideration for the specific localizations of volume of fluid at a given temperature).

Graphical tool : *Q-Q plot*



Left : real data over a spatial domain ($r(u \in \mathcal{U})$). **Middle** : simulation ($y(u \in \mathcal{U})$). **Right** : corresponding Q-Q plot.

We look for a synthetic indicator rather than plotting dozens of Q-Q plots.

more

Principle of the analysis

- Complete spatio-temporal domain :

$$\mathcal{D} = ((\mathcal{S}_i \cup \mathcal{S}_e) \times [0\ ;\ T]) \cup (\mathcal{S}_c \times [T_0\ ;\ T])$$

  where

  - $\mathcal{S}_e$ is the downcomer external surface,
  - $\mathcal{S}_i$ is the downcomer internal surface,
  - et $\mathcal{S}_c$ is the core entrance surface.

- Definition of a subdomain $\mathcal{W} \subset \mathcal{D}$ and a random variable $U \sim \mathcal{U}(\mathcal{W})$.
- Analysis of scalar random variables (five repetitions of the physical experience, one numerical simulation) : $(r^1(U), r^2(U), r^3(U), r^4(U), r^5(U), y(U))$.

Indices de Tversky

- Definition based on a set $\mathcal{R}$ of interest ($\mathcal{R} \subset \mathbb{R}$) :
  - $A_{OV}(v, w) = \mathbb{E}\left[\mathbb{1}_{\mathcal{R}\times\mathcal{R}}(v(U), w(U))\right]$ (overlapping area) ;
  - $A_{FP}(v, w) = \mathbb{E}\left[\mathbb{1}_{\bar{\mathcal{R}}\times\mathcal{R}}(v(U), w(U))\right]$ (false positive area) ;
  - $A_{FN}(v, w) = \mathbb{E}\left[\mathbb{1}_{\mathcal{R}\times\bar{\mathcal{R}}}(v(U), w(U))\right]$ (false negative area) †.
- Let $b(u) = \min(v(u),\ w(u))$ for all $u \in \mathcal{W}$, second definition based on :
  - $A_{OV}(v, w) = \mathbb{E}\left[b(U)\right]$;
  - $A_{FP}(v, w) = \mathbb{E}\left[\max(w(U) - b(U),\ 0)\right]$ (the more w over-predicts v on $\mathcal{W}$, the greater) ;
  - $A_{FN}(v, w) = \mathbb{E}\left[\max(v(U) - b(U),\ 0)\right]$ (the more w under-predicts v on $\mathcal{W}$, the greater).
  - Index de Tversky :

$$\mathcal{T}(v, w) = \frac{A_{OV}(v, w)}{A_{OV}(v, w) + A_{FN}(v, w) + A_{FP}(v, w)}.$$

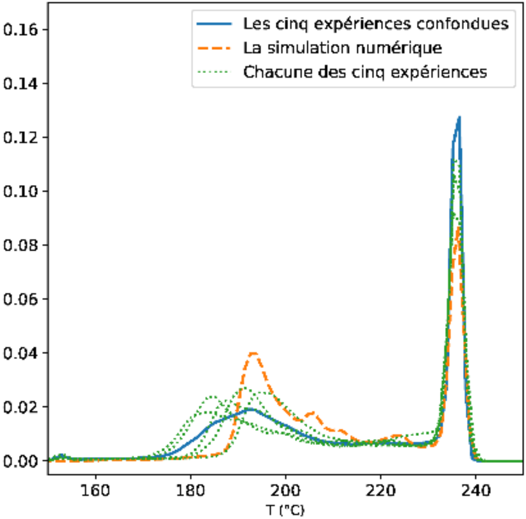
  - Over-prediction index :

$$\mathcal{T}_{+}(v, w) = 1 - \frac{A_{FP}(v, w)}{A_{OV}(v, w) + A_{FP}(v, w)}.$$

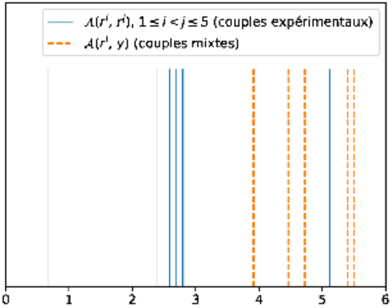
  - Under-prediction index :

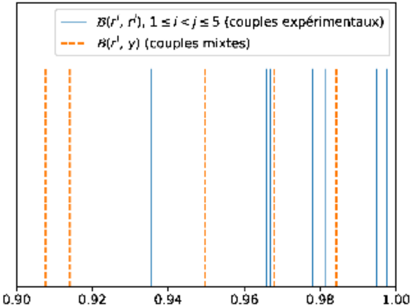
$$\mathcal{T}_{-}(v, w) = 1 - \frac{A_{FN}(v, w)}{A_{OV}(v, w) + A_{FN}(v, w)}.$$

Global analysis : downcomer

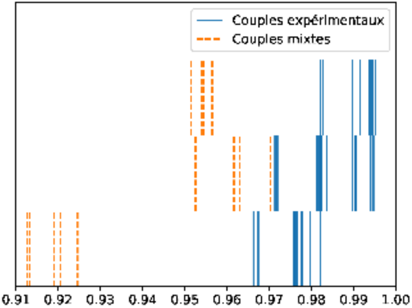


Global analysis : downcomer

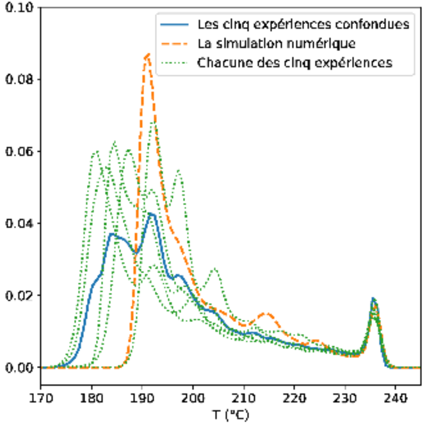




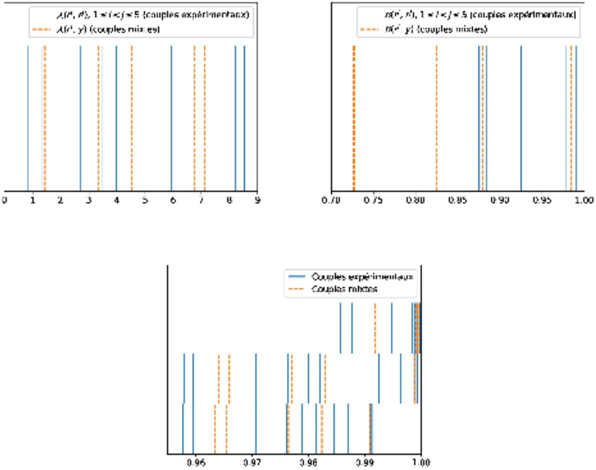
Global analysis : downcomer



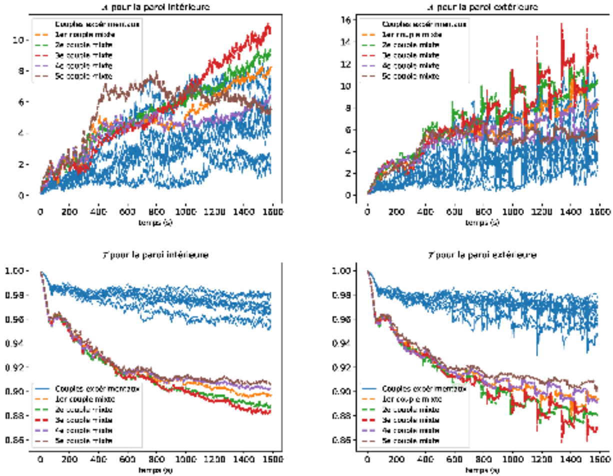
Global analysis : core entrance



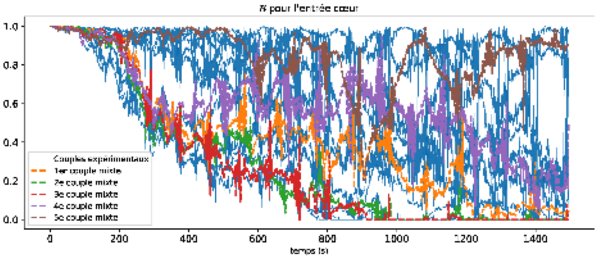
Global analysis : core entrance



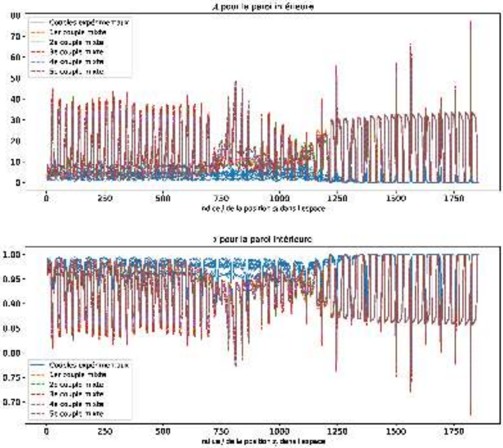
Analysis in time : downcomer



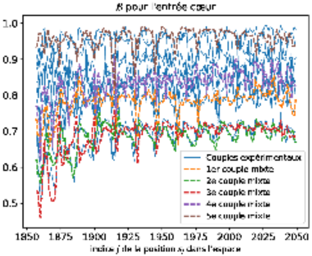
Analysis in time : core entrance

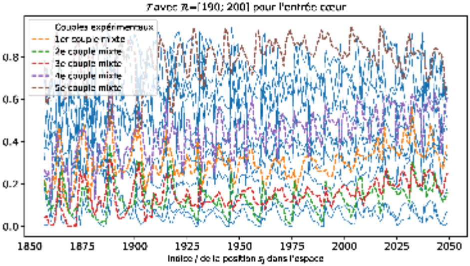


Analysis in space : downcomer (internal wall)



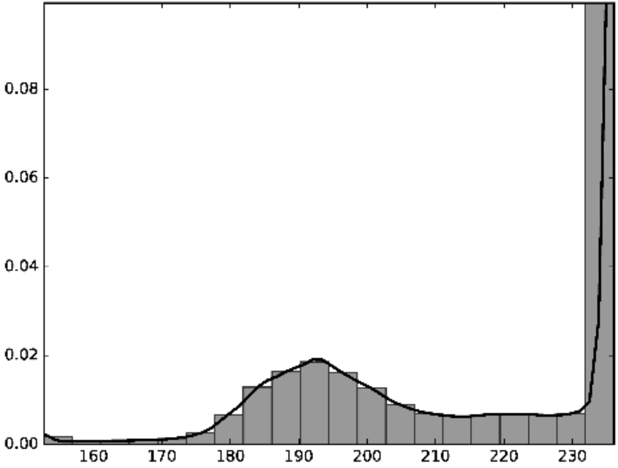
Analysis in space : core entrance





Remark : Bhattacharyya coefficient computation

So as to limit the CPU cost : estimation of PDF by histograms of fixed bins :



Contents

Validation of CATHARE2

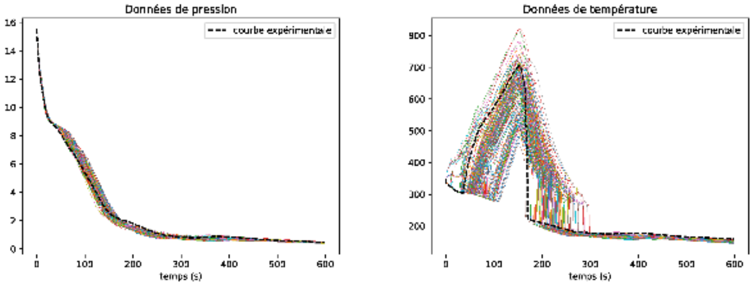


FIGURE: One experiment, $N = 208$ simulations.

Not a pure validation !

- SRQ r deterministic.
- Measurement errors neglected.
- Sampling of $Z(\hat{x}, \theta)$ according to $\theta \sim \pi_\theta$.

Statistical depth

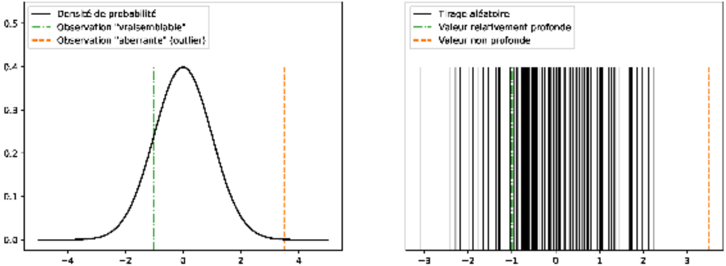
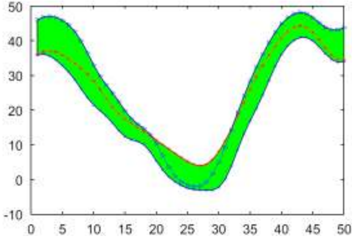


Figure: Illustration in the univariate case.

Band depth (López-Pintado and Romo, 2007)

- Functional data : $y^n(t)$ for $t \in [0\ ;\ T]$, $1 \leq n \leq N$
- Graph of y : $G(y) = \{(t, y(t)) : t \in [0\ ;\ T]\}$
- Band of $y^{n_1}, \ldots, y^{n_j}$:
  $B(y^{n_1}, \ldots, y^{n_j}) = \{(t, y) : t \in [0\ ;\ T], \min_{k=1,\ldots,j} y^{n_k}(t) \leq y \leq \max_{k=1,\ldots,j} y^{n_k}(t)\}$



A band defined by three curves (López-Pintado and Romo, 2007)

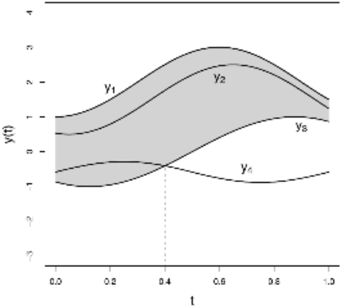
- Band depth of y ($J = 2$ or 3 in practice) :

$$S_{N,J}(y) = \sum_{j=2}^{J} \underbrace{\binom{N}{j}^{-1} \sum_{1 \leq n_1 < \ldots < n_j \leq N} \mathbb{1}(G(y) \subset B(y^{n_1}, \ldots, y^{n_j}))}_{\text{proportion of bands given by } j \text{ curves containing } G(y)}$$

Generalized/modified band depth (López-Pintado and Romo, 2007)

- Set where y is inside $B(y^{n_1}, \ldots, y^{n_j})$:
$A(y\,; y^{n_1}, \ldots, y^{n_j}) = \{t \in [0\,;T] : \min_{k=1,\ldots,j} y^{n_k}(t) \leq y \leq \max_{k=1,\ldots,K} y^{n_k}(t)\}$
- Generalized band depth of y ($J = 2$ or 3 in practice) :

$$GS_{N,\,J}(y) = \sum_{j=2}^{J} \binom{N}{j}^{-1} \sum_{1 \leq n_1 < \ldots < n_j \leq N} \underbrace{\lambda(A(y\,; y^{n_1}, \ldots, y^{n_j}))/T}_{\text{proportion of time during which } y \text{ is inside } B(y^{n_1}, \ldots, y^{n_j})}$$



(Sun and Genton, 2011)

Statistical depth

Tool : function fbplot of the R package fda.

Statistical depth

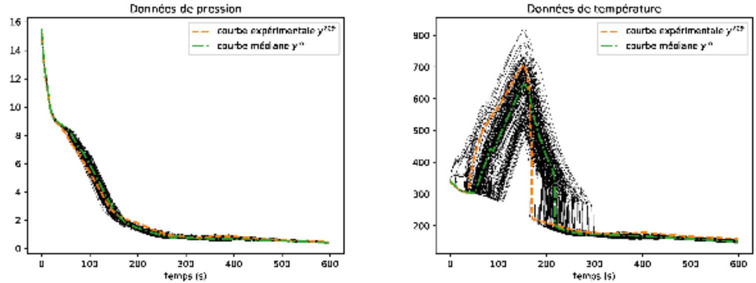
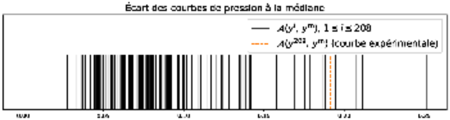
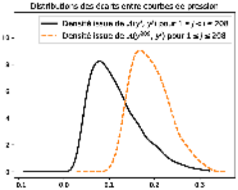
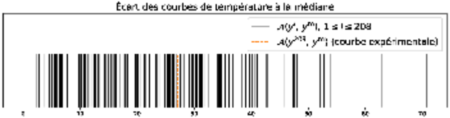


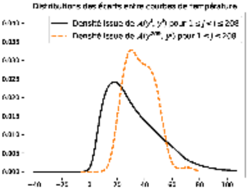
Figure: Median curves (maximal depth).

Area metric (over the time interval)

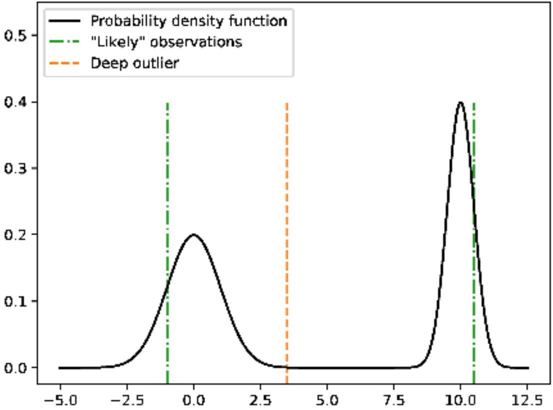








Deep outliers !



Idea : look whether the experimental curve y^{209} belongs to a high-density zone.

- Reduce the curves to few latent variables $C = (C_1, \ldots, C_K)$ (dimension reduction), e.g. by Principal Component Analysis.
- Estimate the multivariate density of the latent variables C, e.g. by gaussian mixture (Python module Pymixmod).

High-density zone

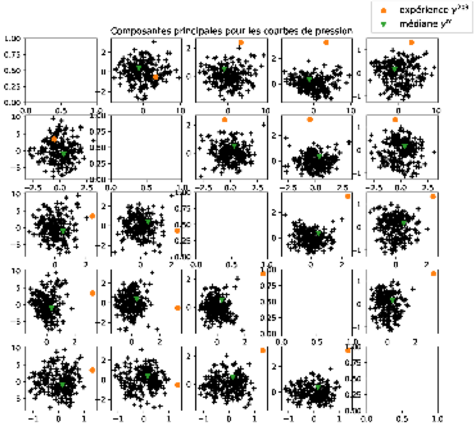


Figure: Curves of pressure : matrix of the scatter plots of the five first principal component.

High-density zone

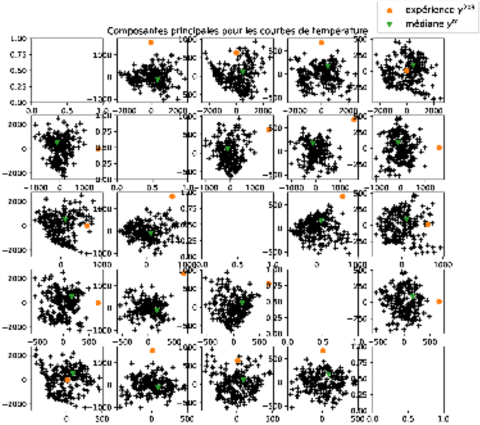


Figure: Curves of temperature : matrix of the scatter plots of the five first principal component.

High-density zone

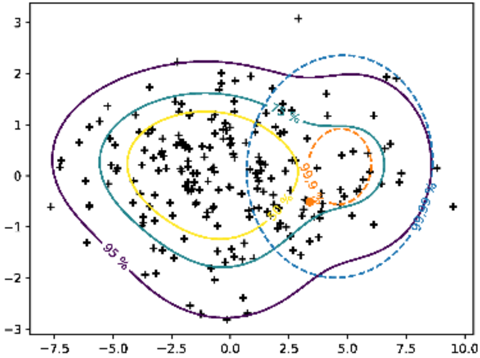
Minimum volume of a given confidence level

$$\mathcal{V}_\beta = \arg \min_{\substack{E \in \mathcal{B}(\mathbb{R}^K) \\ \mathbb{P}(C \in E) = \beta}} \lambda(E)$$

There is a density threshold d_β such that

$$\mathcal{V}_\beta = \{x \in \mathbb{R}^K \quad / \quad f_C(x) \geq d_\beta\}.$$

High-density zone



Points (c_2^n, c_1^n) pour $1 \leq n \leq N$ and some level curves of the marginal PDF of (C_2, C_1) (trais continus) and of the joint PDF of $C = (C_1, \ldots, C_K)$ (dashed lines) in the plane $C_3 = c_3^{209}$, $C_4 = c_4^{209}$, $C_5 = c_5^{209}$.

Tool : method Distribution.computeMinimumLevelSet of OpenTURNS.

Very small approximate p-value $\in [0,01\%; 0,1\%]$!

Contents

Perspectives

Bayesian model of Bayarri et al. (2007) (following Kennedy and O'Hagan (2001))

- Measurement error : $M(x) - r(x) \sim \mathcal{N}(0, 1/\lambda^m)$ with a unknown precision λ^m.
- Biased numerical model : $r(x) = y(x, \theta) + b_\theta(x)$
- Independent priors on θ (uniform) and b.
- Kriging of the unknown bias : $b(x) \sim GP(\mu(x), c(\delta x))$ where : †
  - *a systematic bias* : $\mathbb{E}[b(x)] = \mu(x) = \mu^b$;
  - $\mathrm{Cov}[b(x), b(x')] = c(x - x') = \frac{1}{\lambda^b} c^{gecf}_{\alpha^b,\beta^b}(x - x')$ with a smoothing/regularity parameter α^b and some unknown precision and scale parameters. λ^b and β^b.

†. c^{gecf} : *generalized exponential correlation function*

Bayes factor (Rebba et al., 2006; Damblin, 2015)

$\frac{\mathbb{P}(Z|m)}{\mathbb{P}(Z'|m)}$ with Z without bias and Z' with bias. Ask Merlin !

PERSPECTIVES

The multifidelity approach

- Objective : building a predictive model of the finest fidelity y^1 (e.g. fine mesh) from evaluations of y_1 together with some of a lower fidelity model y^0 (e.g. coarse mesh).
- Le Gratiet (2013) :

$$y^1(x) = \rho(x)\, y^0(x) + b(x) \text{ with } y^0(.),\, b(.) \sim \text{ GP}$$

and $\rho(x) = \theta^T f(x)$ with β unknown.

- Zertuche (2015) :

$$y^1(x) = \varphi(y^0(x)) + b(x) \text{ with } y^0(.),\, b(.) \sim \text{ GP}$$

and φ estimated by (one-dimensional) locally linear polynomial regression.

- Replace y^1 by r and y^0 by y.

Last words : evaluation, not validation

Chang and Hanna (2004) :

> *However, according to Oreske et al. (1994), verification and validation of numerical models of natural systems are impossible, because natural systems are never closed and because model solutions are always non-unique. The random nature of the process leads to a certain irreductible inherent uncertainty. Oreske et al. (1994) suggest that models can only be confirmed or evaluated by the demonstration of good agreement between several sets of observations and predictions. Following this guidance, the term evaluation is used instead of verification throughout this paper.*

RÉFÉRENCES

American Institute of Aeronautics and Astronautics (1998). Guide for the verification and validation of computational fluid dynamics simulations.

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007). A framework for validation of computer models. *Technometrics*, 49(2) :138–154.

Chang, J. C. and Hanna, S. R. (2004). Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, 87(1-3).

Damblin, G. (2015). *Contributions statistiques au calage et à la validation des codes de calcul*. PhD thesis, Université Paris-Saclay.

Ferson, S., Oberkampf, W. L., and Ginzburg, L. (2008). Model validation and predictive capability for the thermal challenge problem. *Computer Methods in Applied Mechanics and Engineering*, 197(29–32) :2408 – 2430. Validation Challenge Workshop.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 63(3) :425–464.

Le Gratiet, L. (2013). *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis, Université Paris-Diderot.

López-Pintado, S. and Romo, J. (2007). Depth-based inference for functional data. *Computational Statistics & Data Analysis*, 51 :4957–4968.

Oberkampf, W. L. and Barone, M. F. (2006). Measures of agreement between computation and experiment : Validation metrics. *Journal of Computational Physics*, 217(1) :5 – 36. Uncertainty Quantification in Simulation Science.

Oberkampf, W. L. and Roy, C. J. (2010). *Verification and Validation in Scientific Computing*. Cambridge University Press.

Rebba, R., Mahadevan, S., and Huang, S. (2006). Validation and error estimation of computational models. *Reliability Engineering and System Safety*, 91(10–11) :1390—1397.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2) :316—334.

Székely, G. J. and Rizzo, M. L. (2004). Testing for equal distributions in high dimension. *InterStat*.

The American Society of Mechanical Engineers (2006). Guide for verification and validation in computational solid mechanics.

The American Society of Mechanical Engineers (2009). Standard for verification and validation in computational fluid dynamics and heat transfer. ISBN : 9780791832097.

Warner, S., Platt, N., and Heagy, J. F. (2004). User-oriented two-dimensional measure of effectiveness for the evaluation of transport and dispersion models. *Journal of Applied Meteorology*, 43(1) :58–73.

...rtuche, F. (2015). *Validation de simulateurs multi-fidélité pour les études d'incertitudes dans les codes de calcul*.