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

1. Generalities
2. First practical application: ROCOM
3. Second practical application: loss of coolant accident (LOCA)
4. Perspectives
Contents

1. Generalities

2. First practical application: ROCOM

3. Second practical application: loss of coolant accident (LOCA)

4. Perspectives
**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 Response Quantity* (Oberkampf and Roy, 2010): physical quantity of interest considered during the validation

- \( U \subset \mathbb{R}^L \) où \( L \leq 4 \) spatio-temporal domain, e.g. \( U = S \times [0 ; T] \) where \( S \subset \mathbb{R}^3 \) 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 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**: \( R(\omega \in \varOmega, x, u) \), hence \( V(\omega \in \varOmega, x, u) \)...
Example of an ambitious approach

Standard of The American Society of Mechanical Engineers (2009)

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

- \( \varepsilon_y = y(x) - v(x) \): 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 \( \alpha \) for \( v(x) - r(x) \).

Assumption of independent gaussian errors:

\[
\begin{align*}
\left[ d(x) - \beta \sqrt{\sigma_y^2 + \sigma_x^2 + \sigma_m^2} \right], & \quad d(x) + \beta \sqrt{\sigma_y^2 + \sigma_x^2 + \sigma_m^2}
\end{align*}
\]

where \( \beta \) 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):

\[
Y : (x, \theta_y) \mapsto Y(\omega \in \Omega, u \in U, x, \theta_y)
\]

where \( \theta_y \) are uncertain parameters assumed invariant from an experiment to another one.

- \( N \) data (measurements) associated to conditions \( x^n \), \( 1 \leq n \leq N \): \( m^n(u \in U) \) realizations of \( M(x^n, u \in U) \) such that \( M(x^n, u \in U) - R(x^n, u \in U) \) 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.
Example of statistical model

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

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

$\theta = (\theta_y, \theta_\varepsilon, \theta_1, \ldots, \theta_K) \sim \pi_\theta$ (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:

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

Pure validation in practice

$$\forall k \in \{1, \ldots, K\} \quad (M_1(x^{\psi(k,1)}), \ldots, M_P(x^{\psi(k,1)})) \sim (Z_1(\hat{x}^k, \hat{\theta}), \ldots, Z_P(\hat{x}^k, \hat{\theta}))$$

$L_k$-sample available

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 in practice

\[
\forall k \in \{1, \ldots, K\} \quad \left( M_1(x_{\psi(k,1)}), \ldots, M_P(x_{\psi(k,1)}) \right) \sim \left( Z_1(\hat{x}^k, \hat{\theta}), \ldots, Z_P(\hat{x}^k, \hat{\theta}) \right)
\]

\[L_k\text{-sample available}\]

\[\text{may incorporate a CPU-consuming black-box model...}(H_0)\]

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 in practice

\[ \forall k \in \{1, \ldots, K\} \quad \left( M_1(x^{\psi(k,1)}), \ldots, M_P(x^{\psi(k,1)}) \right) \sim \left( Z_1(\hat{x}^k, \hat{\theta}), \ldots, Z_P(\hat{x}^k, \hat{\theta}) \right) \]

L_k-sample available

may incorporate a CPU-consuming black-box model...

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: factorial 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).
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

$$\mu_M = \mathbb{E}[M^k] < +\infty \text{ and } \nu[M^k] < +\infty \text{ and}$$

$$\bar{M}_n = \frac{1}{n} \sum_{k=1}^{n} M^k \text{ and } S_n^2 = \frac{1}{n} \sum_{k=1}^{n} (M^k)^2 - \bar{M}_n^2,$$

then a confidence interval of level $\alpha$ of $\mu_M$ is

$$I_n = [\bar{M}_n - u_{1-\alpha/2} \frac{S_n}{\sqrt{n}} ; \bar{M}_n + u_{1-\alpha/2} \frac{S_n}{\sqrt{n}}]$$

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 $\mathbb{E}[M^k(x)] = r(x)$.

Indicators proposed by Oberkampf and Barone (2006)

- Since $\mathbb{P}(r(x) - y(x) \in \mathcal{I}_n(x)) = 1 - \alpha : \bar{M}_n(x) - y(x)$ and $|\mathcal{I}_n(x)| = 2 u_{1-\alpha/2} \frac{S_n(x)}{\sqrt{n}}$.

- Estimation by regression in $\mathcal{V}$:
  - global indicators:
    
    $\text{Volume}(\mathcal{V})^{-1} \int_{\mathcal{V}} \left| \frac{\mathbb{E}[M(x)] - y(x)}{\mathbb{E}[M(x)']} \right| \, dx$ et $\text{Volume}(\mathcal{V})^{-1} \frac{u_{1-\alpha/2}}{\sqrt{n}} \int_{\mathcal{V}} \left| \frac{S_n(x)}{\hat{M}(x)} \right|$ ;

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

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

Validation of numerical models by statistics 04/06/2018 14/52
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

Validation of numerical models by statistics

04/06/2018

15/52
HYPOTHESIS TESTING (P-VALUE)

Let $\Psi$ be a test statistic:

\[
p-value = \text{proba. to wrongly reject the model} \\
\quad \text{(small value }\Rightarrow\text{ wish of rejection)} \\
= \mathbb{P}(\Psi(Z^1, \ldots, Z^N) \notin I(m^1, \ldots, m^n))
\]

\(\text{set of values « less extreme »}
\text{than } \psi(m^1, \ldots, m^N)\)

Left:

- \(p-value = \mathbb{P}(Z > m^1) (\Psi(Z) = Z)\)
- The \(p\)-value corresponds to the order \(\alpha\) of the quantile \(q_\alpha^Z = m^1\): 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.

<table>
<thead>
<tr>
<th>Area metric [Ferson et al., 2008]</th>
<th>Relative entropy (divergence of Kullback-Leibler)</th>
<th>Distance of Hellinger</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|F_R - F_Y|<em>{L^1} = \int</em>{\mathbb{R}}</td>
<td>F_R - F_Y</td>
<td>$</td>
</tr>
<tr>
<td>Easy to estimate from empirical CDF</td>
<td>$\mathcal{K}(F_R, F_Y) = 0 \Rightarrow F_R = F_Y$</td>
<td>$\mathcal{H}(F_R, F_Y) = 0 \Rightarrow F_R = F_Y$</td>
</tr>
<tr>
<td>Always well-defined and robust</td>
<td>Not symmetric (notion of reference $F_R$)</td>
<td>$\mathcal{H}(F_R, F_Y) \leq 1$</td>
</tr>
<tr>
<td>Same physical unit than $r$ (scale variant)</td>
<td>Issues: estimation of PDF and cases where $\mathcal{K}(F_R, F_Y) = +\infty$</td>
<td>Minor issue: estimation of PDF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Undimensional overlap indicator</td>
</tr>
</tbody>
</table>
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 ($\mathcal{H}(F_R, F_Y) = \int_{\mathbb{R}} (\sqrt{f_R} - \sqrt{f_Y})^2$):

$$\mathcal{H}(\text{Case 1}) > \mathcal{H}(\text{Case 2}) \quad \text{and} \quad \mathcal{A}(\text{Case 1}) < \mathcal{A}(\text{Case 2})$$

\[= 1 \text{ (no overlapping)} \neq 1\]

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

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 \( t^{80} - 4 \)).
Comparision 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

1 Generalities

2 First practical application : ROCOM

3 Second practical application : loss of coolant accident (LOCA)

4 Perspectives
**Indices of Tversky**

**Figure of Merit in Space (FMS)**

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

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

**Generalization** (Warner et al., 2004)

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

\[ \mathcal{T} = \frac{\mathbb{1}_{(r,y) \in \mathcal{R}^2(r,y)}}{\mathbb{1}_{r \in \mathcal{R} \text{ ou } y \in \mathcal{R}(r,y)}} = \frac{\mathbb{1}_{(r,y) \in \mathcal{R}^2(r,y)}}{\mathbb{1}_{(r,y) \in \mathcal{R}^2(r,y)} + \mathbb{1}_{(r,y) \in \mathcal{R} \times \bar{\mathcal{R}}(r,y)} + \mathbb{1}_{(r,y) \in \bar{\mathcal{R}} \times \mathcal{R}(r,y)}} \]
Indices of Tversky

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

\[ T = \frac{A_{OV} (\text{overlap})}{A_{OV} + A_{FN} (\text{false negative}) + A_{FP} (\text{false positive})} \]

- Risk-weighted FMS : RWFMS = \[ \frac{A_{OV} (\text{overlap})}{A_{OV} | \alpha \ A_{FN} (\text{false negative}) | \beta \ A_{FP} (\text{false positive})} \]

- Measure Of Effectiveness : MOE = \left( \frac{A_{OV}}{A_{OB}} \right) \left( \frac{A_{OV}}{A_{PR}} \right) - \left( 1 - \frac{A_{FN}}{A_{OB}} \right) \left( 1 - \frac{A_{FP}}{A_{PR}} \right)

avec \( A_{OB} = A_{OV} + A_{FN} (\text{observed}) \) et \( A_{PR} = A_{OV} + A_{FP} (\text{predicted}) \).

- Definitions not based on a set \( \mathcal{R} \) of interest :

\[ A_{OV} = \overline{A_{i}^{OV}} \text{ where} \]
\[ A_{i}^{OV} = \min(r_{i} ; y_{i}) \]
\[ A_{FN} = \max(r_{i} - A_{i}^{OV} ; 0) \]
\[ A_{FP} = \max(y_{i} - A_{i}^{OV} ; 0) \]
Experimental data from project OCDE PKL2 (HZDR, Helmholtz Zentrum Dresden Rossendorf).

Simulation performed with Code_Saturne by Arnaud Barthet (EDF).
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:
  
  \[ \mathcal{D} = ((S_i \cup S_e) \times [0; T]) \cup (S_c \times [T_0; T]) \]

  where
  - \( S_e \) is the downcomer external surface,
  - \( S_i \) is the downcomer internal surface,
  - \( S_c \) is the core entrance surface.

- Definition of a subdomain \( \mathcal{W} \subset \mathcal{D} \) and a random variable \( U \sim U(\mathcal{W}) \).

- Analysis of scalar random variables (five repetitions of the physical experience, one numerical simulation): \( (r^1(U), r^2(U), r^3(U), r^4(U), r^5(U), y(U)) \).
Indice de Tversky

- Definición basada en un conjunto \( \mathcal{R} \) de interés \( (\mathcal{R} \subset \mathbb{R}) \):
  - \( A_{OV}(v, w) = \mathbb{E} \left[ 1_{\mathcal{R} \times \mathcal{R}}(v(U), w(U)) \right] \) (área de sobreposición);
  - \( A_{FP}(v, w) = \mathbb{E} \left[ 1_{\overline{\mathcal{R}} \times \mathcal{R}}(v(U), w(U)) \right] \) (área de falsos positivos);
  - \( A_{FN}(v, w) = \mathbb{E} \left[ 1_{\mathcal{R} \times \overline{\mathcal{R}}}(v(U), w(U)) \right] \) (área de falsos negativos).

- Sea \( b(u) = \min(v(u), w(u)) \) para todo \( u \in \mathcal{W} \), segunda definición basada en:
  - \( A_{OV}(v, w) = \mathbb{E} [b(U)] \);
  - \( A_{FP}(v, w) = \mathbb{E} [\max(w(U) - b(U), 0)] \) (más \( w \) sobrepredecir \( v \) en \( \mathcal{W} \), más grande);
  - \( A_{FN}(v, w) = \mathbb{E} [\max(v(U) - b(U), 0)] \) (más \( w \) underpredictar \( v \) en \( \mathcal{W} \), más grande).

- Índice de Tversky:
  \[ T(v, w) = \frac{A_{OV}(v, w)}{A_{OV}(v, w) + A_{FN}(v, w) + A_{FP}(v, w)} \]

- Índice de sobrepredicción:
  \[ T_+(v, w) = 1 - \frac{A_{FP}(v, w)}{A_{OV}(v, w) + A_{FP}(v, w)} \]

- Índice de underpredicción:
  \[ T_-(v, w) = 1 - \frac{A_{FN}(v, w)}{A_{OV}(v, w) + A_{FN}(v, w)} \]
GLOBAL ANALYSIS: DOWNCOMER

The diagram illustrates the comparison between different experimental conditions and a numerical simulation. The legend includes the following:

- Blue line: Les cinq expériences confondues (The five combined experiences)
- Orange dashed line: La simulation numérique (The numerical simulation)
- Green dotted line: Chacune des cinq expériences (Each of the five experiences)

The x-axis represents temperature (°C) ranging from 160 to 240, and the y-axis represents a value ranging from 0.00 to 0.16.
Global analysis : downcomer

\[ \mathcal{A}(r^i, r^j), 1 \leq i < j \leq 5 \text{ (couples expérimentaux)} \]
\[ \mathcal{A}(r^i, y) \text{ (couples mixtes)} \]

\[ \mathcal{B}(r^i, r^j), 1 \leq i < j \leq 5 \text{ (couples expérimentaux)} \]
\[ \mathcal{B}(r^i, y) \text{ (couples mixtes)} \]
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

1 Generalities

2 First practical application: ROCOM

3 Second practical application: loss of coolant accident (LOCA)

4 Perspectives
Validation of CATIARE2

**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.
**Statistical depth**

**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):

\[
S_{N, J}(y) = \sum_{j=2}^{J} \binom{N}{j}^{-1} \sum_{1 \leq n_1 < \ldots < n_j \leq N} \mathbb{1}(G(y) \subset B(y^{n_1}, \ldots, y^{n_j}))
\]

proportion of bands given by \( j \) curves containing \( G(y) \)
Statistical depth

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)
  \[ GS_N, J(y) = \sum_{j=2}^{J} \left( \frac{N}{j} \right)^{-1} \sum_{1 \leq n_1 < \ldots < n_j \leq N} \frac{\lambda(A(y ; y^{n_1}, \ldots, y^{n_j}))/T}{\text{proportion of time during which } y \text{ is inside } B(y^{n_1}, \ldots, y^{n_j})} \]

(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

\[ V_\beta = \arg\min_{E \in \mathcal{B}(\mathbb{R}^K)} \lambda(E) \]

\[ \mathbb{P}(C \in E) = \beta \]

There is a density threshold \( d_\beta \) such that

\[ V_\beta = \{ x \in \mathbb{R}^K \mid f_C(x) \geq d_\beta \} \]
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

1 Generalities

2 First practical application: ROCOM

3 Second practical application: loss of coolant accident (LOCA)

4 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 \( \lambda^m \).
- Biased numerical model: \( r(x) = y(x, \theta) + b_\theta(x) \)
- Independent priors on \( \theta \) (uniform) and \( b \).
- Kriging of the unknown bias: \( b(x) \sim GP(\mu(x), c(\delta x)) \) where:
  - a systematic bias: \( \mathbb{E}[b(x)] = \mu(x) = \mu^b \);
  - \( \text{Cov}[b(x), b(x')] = c(x - x') = \frac{1}{\lambda^b} c_{\alpha^b, \beta^b}^\text{gecf}(x - x') \) with a smoothing/regularity parameter \( \alpha^b \) and some unknown precision and scale parameters. \( \lambda^b \) and \( \beta^b \).

\( \dagger \). \( c_{\alpha^b, \beta^b}^\text{gecf} \): generalized exponential correlation function

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

\[
\frac{P(Z|m)}{P(Z'|m)} \text{ with } Z \text{ without bias and } Z' \text{ with bias. Ask Merlin!}
\]
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 $\beta$ unknown.

- **Zertuche (2015)**:

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

  and $\varphi$ estimated by (one-dimensional) locally linear polynomial regression.

- **Replace $y^1$ by $r$ and $y^0$ by $y$.**
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 irreducible 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


