

CLUSTERING METHODS FOR DECISION-MAKING

APPLICATION TO NATURAL & TECHNOLOGICAL (NATECH) RISKS

Y. RICHET



DECISIONS FACING NATECH RISKS



RISKS MANAGEMENT

NATURAL AND/OR TECHNOLOGICAL (NATECH)



- Crisis situations
 - High **uncertainties** are intrinsic to accidental situations
 - **Anticipation** : numerical simulations are a mandatory to predict potential consequences and protect population
 - Strong **time constraints**
- *Operational* decisions
 - Need for **clear and concise** information
 - Communication of **evaluation products** to **decision makers** *should* account for uncertainties

Crisis : **quick** decisions under **uncertainties**

L'Aquila earthquake scientists win appeal

Six seismologists and official had been convicted over reassurance issued to residents before fatal quake in 2009



Orages en Corse : Météo-France invoque une situation « difficilement prévisible »

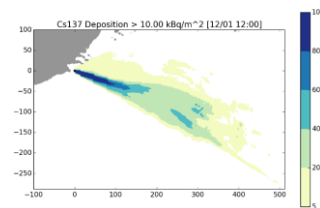
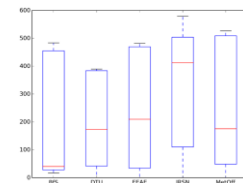
L'activation tardive de la vigilance orange, alors que les rafales ont atteint 200 km/h par endroits, illustre la difficulté à traduire les probabilités de phénomènes météo en un système d'alerte crédible.

Le Monde avec AFP
Publié le 18 août 2022 à 17h34, modifié le 24 juillet 2023 à 18h09 - Lecture 3 min.

UNCERTAINTY ANALYSIS & INTERPRETABILITY

Uncertainty models for quantitative analysis or decision-making

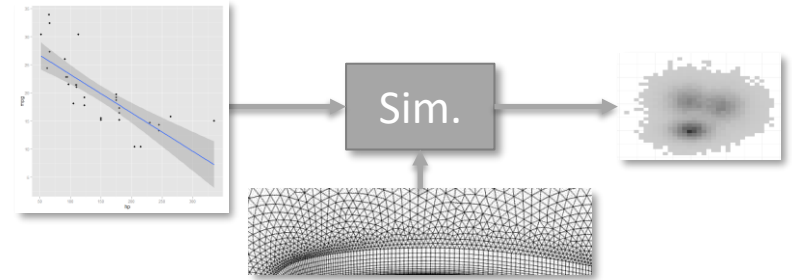
- Quantification of input / model uncertainties
 - Sensitivity analyses/indices (Sobol, Shapley, HSIC, ...)
 - "Envelope" of trust, probability of threshold exceedance...
 - Confidence level in evaluations
 - **Identification of representatives/prototypes**
- ✓ Should be decision-oriented, incl. practical information (e.g. population, agriculture...)
 - ✓ If possible, avoid interpretation bias



Simplicity / Interpretability

UNCERTAINTIES QUANTIZATION

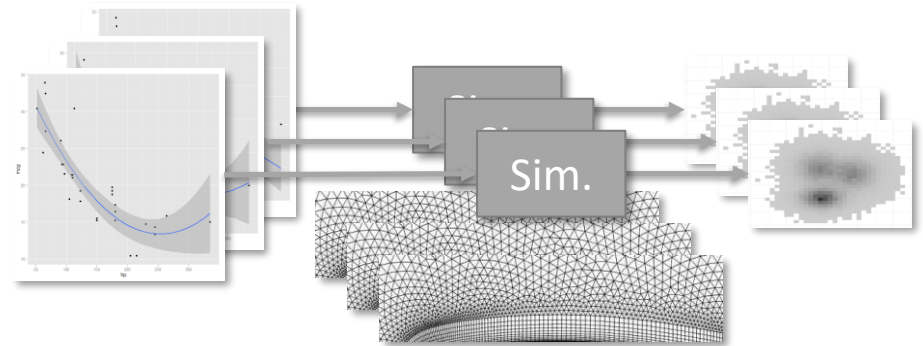
Simulations:



... a suitable support for propagation of uncertainties

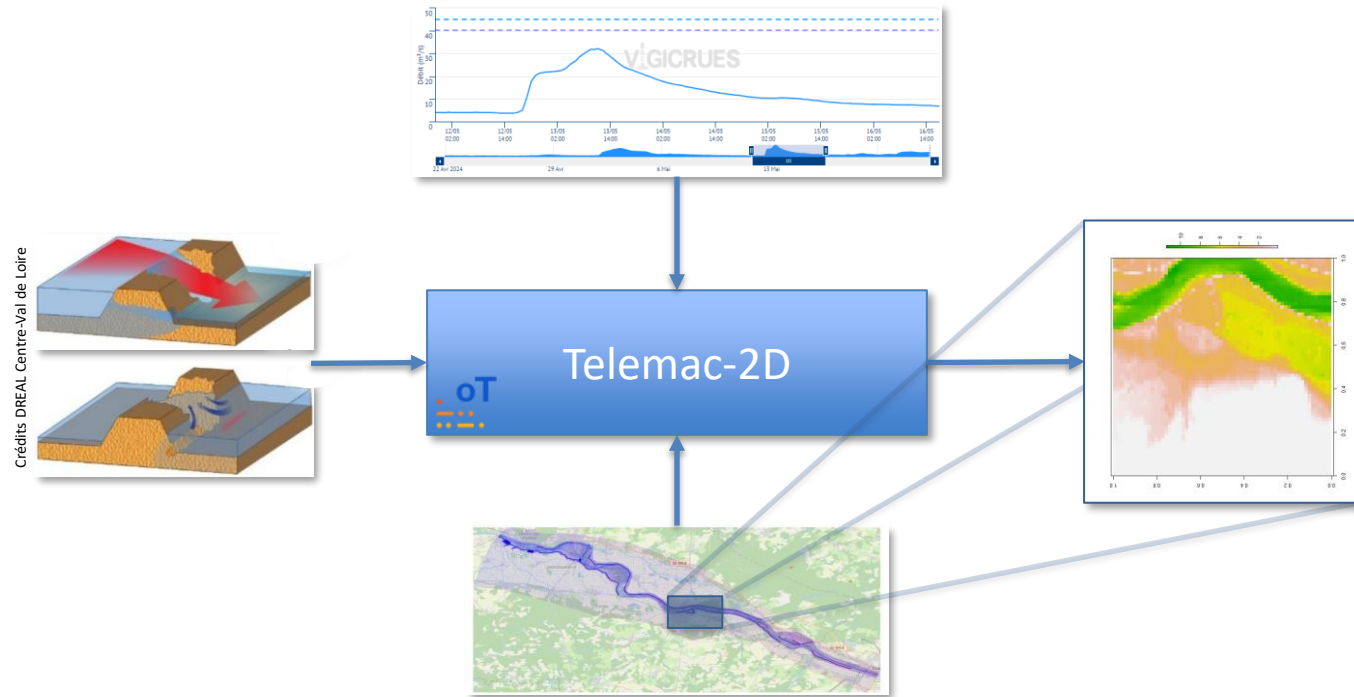
Ex. :

- Monte Carlo / random sampling
- Sets
- Quantiles / delta
- [Fuzzy logic]
- [Experimental Calibration]

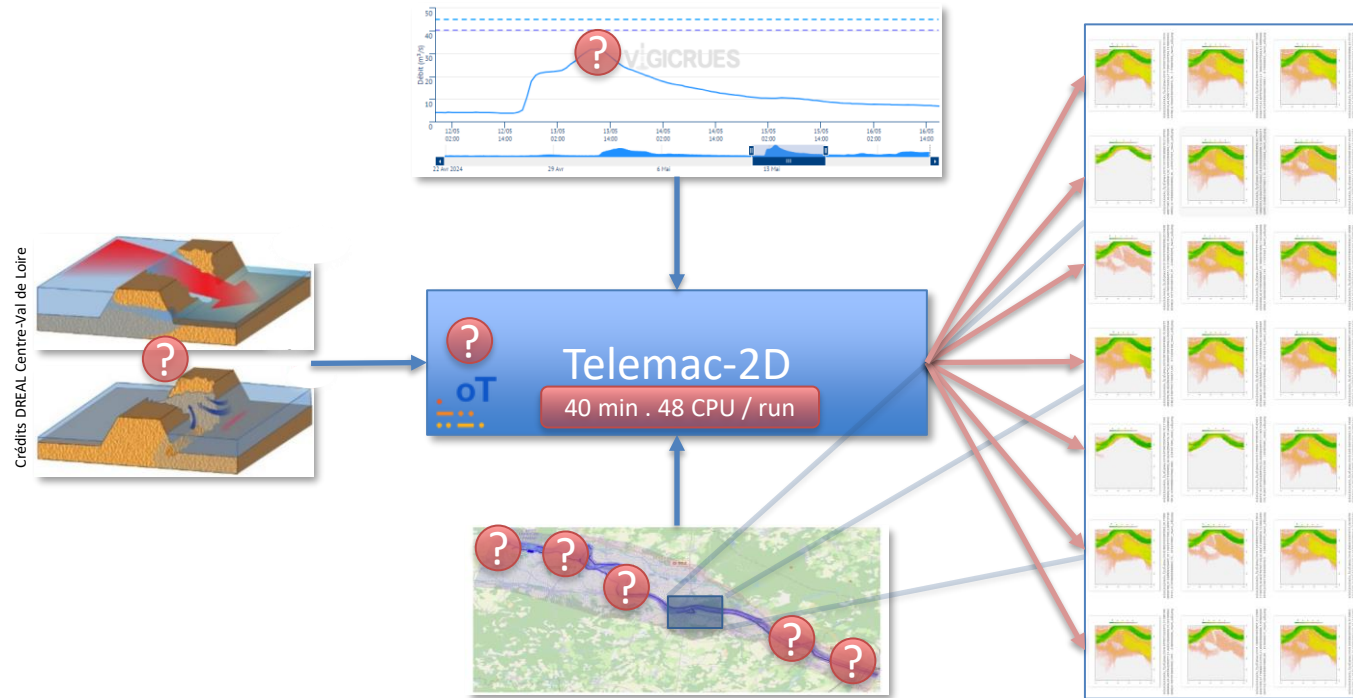


MODELLING HIGH-DIMENSIONAL DATA

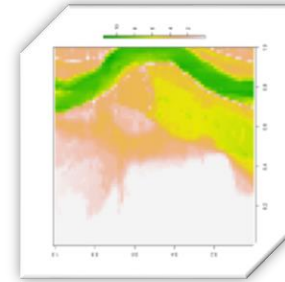
[EX.] FLOODING



[Ex.] FLOODING



HOW TO INCLUDE UNCERTAINTIES IN DECISION MAKING ?



Challenge 1: high-dimension input / output

- Spatio-temporal physical fields
- Interactions / correlations between variables
- Use of appropriate **dimension reduction** methods

Challenge 2: Computational cost of physical models

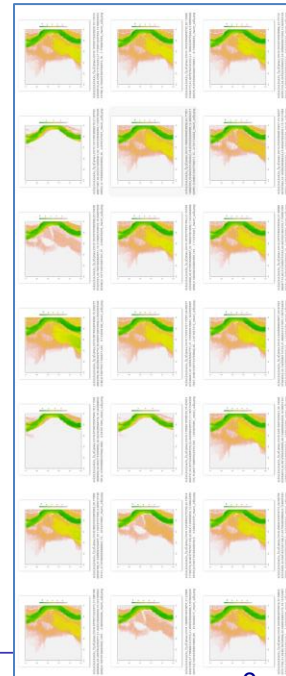
- Use of **metamodel** as surrogate model

Challenge 3: interpretability of output

- Postage stamp ? Too many maps , Probability maps ? Complex interpretation
- Scenario-based approach: “best estimate” vs. “worst case”
- Use of **clustering** methods

Challenge 4: representativity for extreme (& hopefully rare) events

- Use controlled/weighted **sampling**



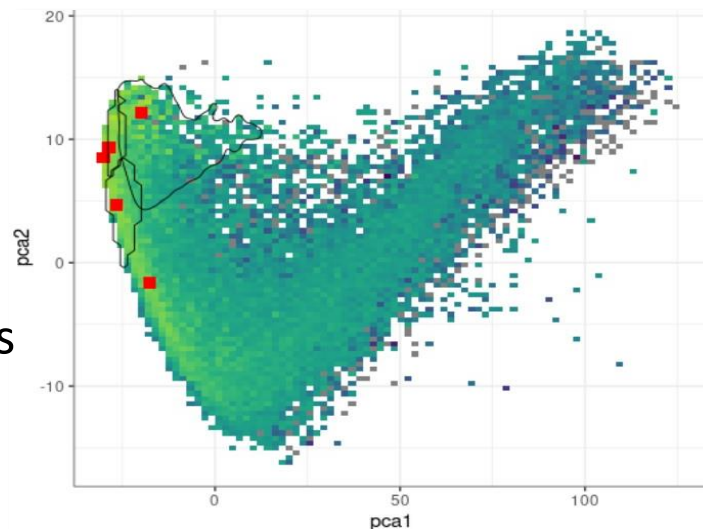
1. HIGH-DIM. INPUT/OUTPUT

Complex (*ie. non scalar*) numerical results:

- Physical quantities: spatial (lat,lon), temporal (t)
- Non-linear output / operational consequences

... supported by a **dimension reduction**:

- Unsupervised / ~weighted / supervised
- Non-significant "latent" space
- ...but keep some desirable math. properties



Credits: Charlie Sire PhD

2. COMPUTATIONAL COST

Parametric modelling to assess physical behaviour:

- *High Performance Computing to compute...*
- *...a numerical Design of Experiments to fit...*
(but still curse of dimensionality)
- *...a response surface*

= Training of a surrogate / **metamodel**
(ex. *Gaussian Process Regression*)

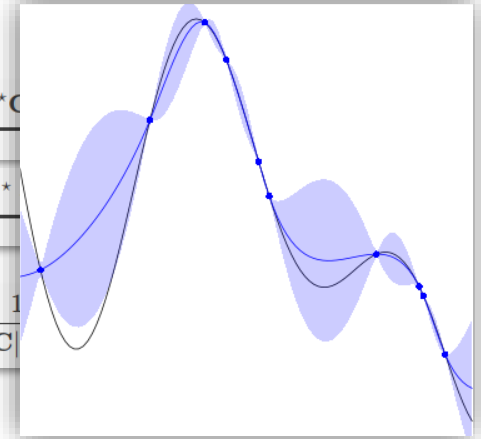
$$\mathbf{y}(\mathbf{x}_i) = \underbrace{\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}}_{\text{trend}} + \underbrace{\zeta(\mathbf{x}_i)}_{\text{smooth GP}}$$

$$C_\zeta(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}, \sigma^2) = C_\zeta(\mathbf{h}; \boldsymbol{\theta}, \sigma^2) = \sigma^2 \prod_{\ell=1}^d \kappa(h_\ell / \theta_\ell)$$

$$\mathbb{E}[\mathbf{y}^* | \mathbf{y}] = \underbrace{\mathbf{F}^* \hat{\boldsymbol{\beta}}}_{\text{trend}} + \underbrace{\mathbf{C}^* \mathbf{C}^{-1} \mathbf{y}}_{\text{smooth GP}}$$

$$\text{Cov}[\mathbf{y}^* | \mathbf{y}] = \underbrace{[\mathbf{F}^* - \hat{\mathbf{F}}^*] \text{Cov}(\hat{\boldsymbol{\beta}}) [\mathbf{F}^* - \hat{\mathbf{F}}^*]^\top}_{\text{trend}} + \underbrace{\mathbf{C}^* (\mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{F} \mathbf{F}^\top \mathbf{C}^{-1}) \mathbf{C}^*}_{\text{smooth GP}}$$

$$L(\boldsymbol{\psi}, \boldsymbol{\beta}; \mathbf{y}) = \frac{1}{[2\pi]^{n/2}} \frac{1}{|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{y} - \mathbf{F} \boldsymbol{\beta})^\top \mathbf{C}^{-1} (\mathbf{y} - \mathbf{F} \boldsymbol{\beta})\right)$$



Credits: <https://libkriging.readthedocs.io>

2.' METAMODELLING

Parametric modelling to assess physical behaviour:
metamodel safety requirements

- **Compliance** with the state of the art in metamodel:
learning set features, metamodel properties,
targeted fitting error, training-testing-validation,
- **Robustness** of performance between the *learning* domain and the domain of *use* of the metamodel,
- **Explainability** that provide *understanding* of the phenomena and *completeness* of the information provided between input data and output results,
- **Transparency** to ensure that all information regarding the metamodel is accessible.

3. INTERPRETABILITY BY SPARSE SAMPLING

Objective: sparse & synthetic sampling

- Some **prototypes** / centroids (on projected output)
- Weighted / probabilistic classes (by input measure)
- "Real" <-> "Latent" space projection

Algorithm 2 Prototype Maps Algorithm

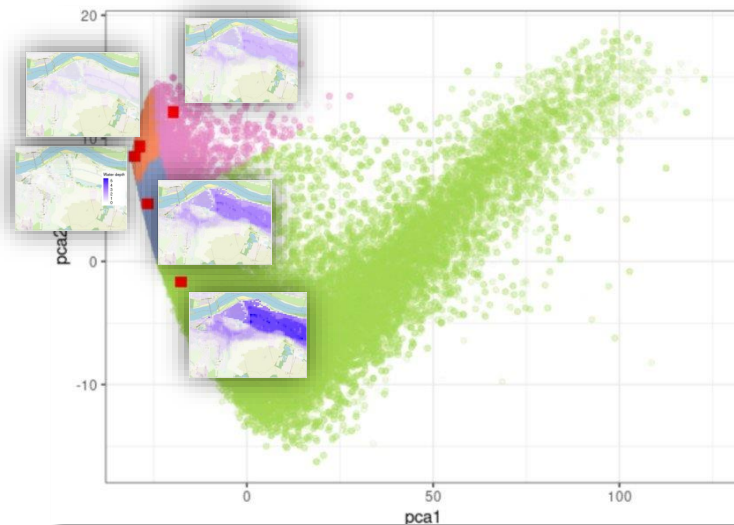
Require: $(y(x^i))_{i=1,\dots,n_{\text{train}}}$, f_X , g , minDistance , ℓ

- 1: Sample $(\tilde{x}^k)_{k \in \{1,\dots,n_{\text{maps}}\}}$ i.i.d. of density function g
- 2: Compute $(\hat{y}(\tilde{x}^k))_{1 \leq k \leq n_{\text{maps}}}$ from $(y(x^i))_{i=1,\dots,n_{\text{train}}}$ (GP & FPCA)
- 3: Compute $(\frac{f_X(\tilde{x}^k)}{g(\tilde{x}^k)})_{1 \leq k \leq n_{\text{maps}}}$
- 4: Initialize $\Gamma_\ell^{[0]} \leftarrow \{\gamma_0^{[0]}, \dots, \gamma_\ell^{[0]}\} \in \mathcal{Y}$
- 5: **while** $\|\Gamma_\ell^{[k+1]} - \Gamma_\ell^{[k]}\| > \text{minDistance}$ **do**
 $\gamma_j^{[k+1]} \leftarrow \hat{E}_{n_{\text{maps}}}(\Gamma_\ell^{[k]}, j, \hat{y})$, $j = 1, \dots, \ell$
 $k \leftarrow k + 1$
- 6: **end while**
- 7: $\hat{\Gamma}_\ell^* = \Gamma_\ell^{[k]}$
- 8: Compute $\hat{P}_{\hat{n}}(\hat{\Gamma}_\ell^*, j, \hat{y})$, $j = 1, \dots, \ell$

Output: $\hat{\Gamma}_\ell^*$ and $\hat{P}_{\hat{n}}(\hat{\Gamma}_\ell^*, j, \hat{y})$

Algorithm 1 Lloyd's algorithm

- $$\Gamma_\ell^{[0]} \leftarrow \{\gamma_1^{[0]}, \dots, \gamma_\ell^{[0]}\}, k \leftarrow 0$$
- 1: **while** stopping criterion not met **do**
 $\gamma_j^{[k+1]} \leftarrow \mathbb{E}[Y(X) \mid Y(X) \in C_j^{\Gamma_\ell^{[k]}}]$, $j \in \{1, \dots, \ell\}$.
 $k \leftarrow k + 1$
 - 2: **end while**



Credits: Charlie Sire PhD

3. INTERPRETABILITY BY SPARSE SAMPLING

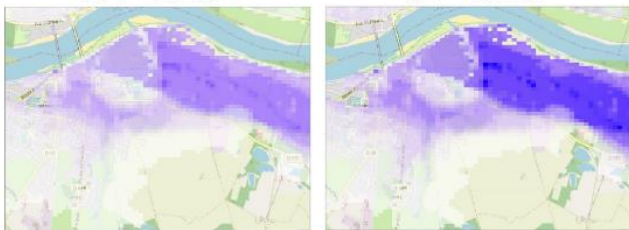


(a) $P = 9.9 \times 10^{-1}$



(b) $P = 3.5 \times 10^{-3}$

(c) $P = 1.5 \times 10^{-3}$



(d) $P = 8.9 \times 10^{-4}$

(e) $P = 5.8 \times 10^{-4}$

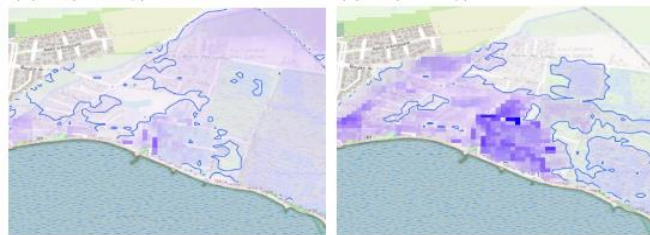


(a) $\hat{P}_n(\hat{\Gamma}_\ell^*, 1, \hat{y}) = 9.7 \times 10^{-1}$, 1 in 1.03



(b) $\hat{P}_n(\hat{\Gamma}_\ell^*, 2, \hat{y}) = 2.0 \times 10^{-2}$, 1 in 50

(c) $\hat{P}_n(\hat{\Gamma}_\ell^*, 3, \hat{y}) = 5.2 \times 10^{-3}$, 1 in 192



(d) $\hat{P}_n(\hat{\Gamma}_\ell^*, 4, \hat{y}) = 6.2 \times 10^{-3}$, 1 in 161

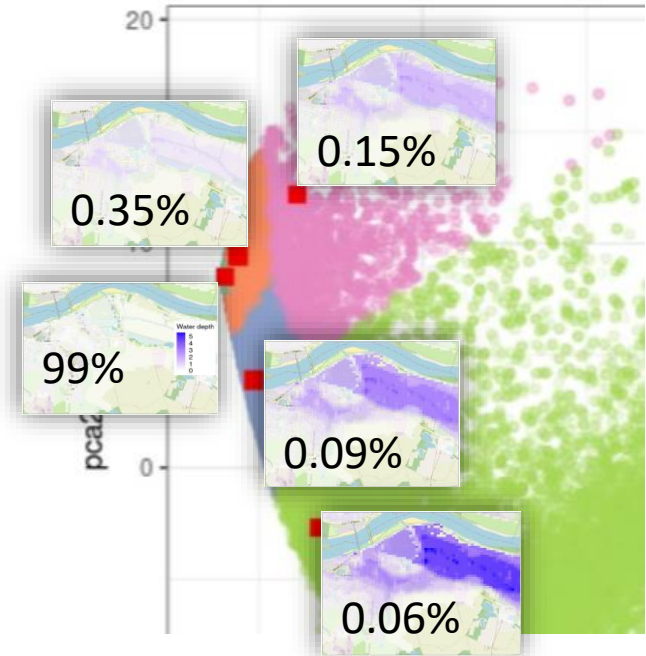
(e) $\hat{P}_n(\hat{\Gamma}_\ell^*, 5, \hat{y}) = 1.4 \times 10^{-3}$, 1 in 714

Credits: Charlie Sire PhD

4. RARE/EXTREME EVENTS SUPPORT

Objective: keep extreme events available to support decision

- Robust classes centroids
- Robust probability of classes
- Backward "Latent" -> "Real" space projection on input



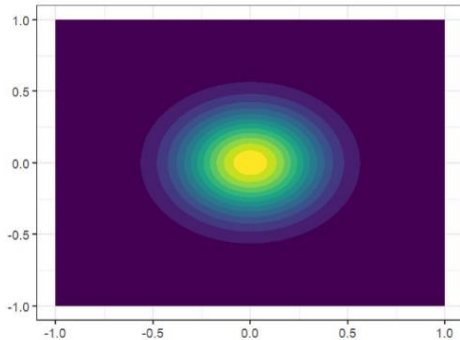
Credits: Charlie Sire PhD

4. RARE/EXTREME EVENTS SUPPORT

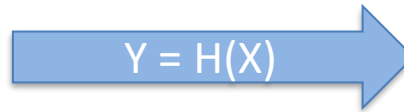
Objective : robust classes (centroid & probability/weight)

Example: analytical 2-dim. model

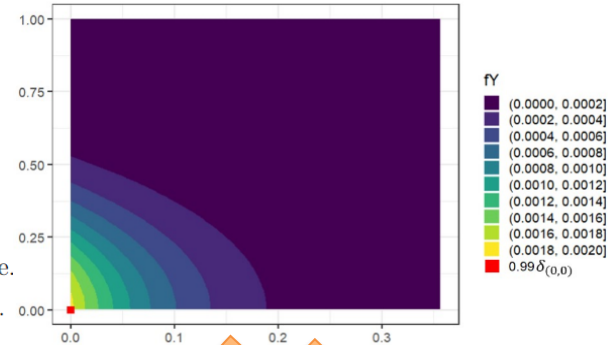
Credits: Charlie Sire PhD



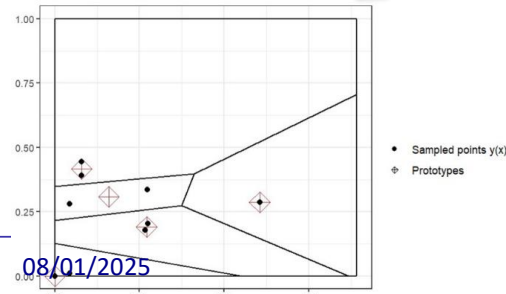
$$\begin{cases} X_i \sim \mathcal{N}_t(0, 0.25^2, -1, 1), i = 1, 2 \\ X_1 \text{ and } X_2 \text{ independent} \end{cases}$$



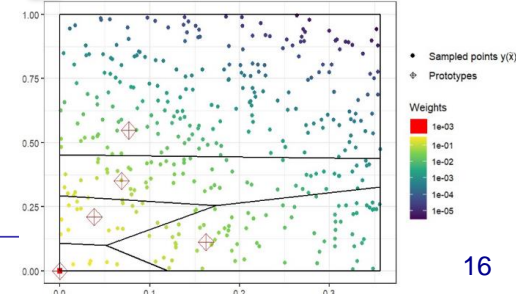
$$H(x) = \begin{cases} (0, 0) & \text{if } |x_1| \leq \alpha \\ (|x_1| - \alpha, |x_2|) & \text{otherwise.} \end{cases}$$
$$P(H(X) = (0, 0)) = 0.99.$$



without importance sampling:



with importance sampling:



SOFTWARE

← → ↻ 🏠 🔒 🔑 https://github.com/charliesire/FunQuant

README

adapted with the `rlikrking` R package. `Funquant` assists users in the fine-tuning of its hyperparameters for a quantization task, by providing a set of relevant performance metrics.

Installation

`Funquant` can be installed from GitHub, for the very latest version:

```
# If not already installed, install package
remotes::install_github("charliesire/FunQuant")
```

Illustrative example

We consider $X = (X_1, X_2) \in \mathbb{R}^2$ a random input

$$\begin{cases} X_1 \sim \mathcal{N}_t(0, 0.25^2, -1, 1) \\ X_1 \text{ and } X_2 \text{ independent} \end{cases}$$

where $\mathcal{N}_t(\mu, \sigma^2, a, b)$ is the Gaussian distribution truncated between a and b .

The density function of X , denoted f_X , is represented by



tranvivioldie / GpOutput2D

Code Issues Pull requests Actions Projects Security Insights

Files

main

Go to file

GpOutput2D.Rcheck

GpOutput2D

R

- Campbell2D.R
- Fpca2d_OrthoNormalBsplines.R
- Fpca2d_Wavelets.R
- Fpca2d_utils.R
- MeanPoe.R
- MeanPoe_utils.R
- OrthoNormalBsplines2D.R
- OrthoNormalBsplines2D_utils.R
- Wavelet2D.R
- gp_Fpca2d.R
- km_Fpca2d.R
- predict.R
- predict_utils.R
- utils.R

man

.Rbuildignore

DESCRIPTION

GpOutput2D / GpOutput2D / R /

tranvivioldie help update

Name

- ..
- Campbell2D.R
- Fpca2d_OrthoNormalBsplines.R
- Fpca2d_Wavelets.R
- Fpca2d_utils.R
- MeanPoe.R
- MeanPoe_utils.R
- OrthoNormalBsplines2D.R
- OrthoNormalBsplines2D_utils.R
- Wavelet2D.R
- gp_Fpca2d.R
- km_Fpca2d.R
- predict.R
- predict_utils.R
- utils.R

← → ↻ 🏠 🔒 🔑 https://libkriging.readthedocs.io/en/latest/functions/Kriging.html

libKriging

0.9.0

Search docs

Installation

Usage

API

Contructors

Kriging

Description

Usage

Arguments

Details

Value

Examples

Kriging: update

<code>y</code>	Numeric vector of response values.
<code>X</code>	Numeric matrix of input design.
<code>kernel</code>	Character defining the covariance model: "gauss", "exp", "matern3_2", "matern5_2".
<code>regmodel</code>	Universal Kriging linear trend: "constant", "linear", "interactive", "quadratic".
<code>normalize</code>	Logical. If <code>TRUE</code> both the input matrix <code>x</code> and the response <code>y</code> in normalized to take value in <code>[0,1]</code> .
<code>optim</code>	Character giving the Optimization method used to fit hyper-parameters. Possible values are "Nelder-Mead", "BFGS", "L-BFGS-B", "CG", "COBYQA", "CONMIN", "FMINBFGS", "FMINCG", "FMINCBO", "FMINCO", "FMINCON", "FMINLBFGS", "FMINLBFGS2", "FMINLBFGS3", "FMINLBFGS4", "FMINLBFGS5", "FMINLBFGS6", "FMINLBFGS7", "FMINLBFGS8", "FMINLBFGS9", "FMINLBFGS10", "FMINLBFGS11", "FMINLBFGS12", "FMINLBFGS13", "FMINLBFGS14", "FMINLBFGS15", "FMINLBFGS16", "FMINLBFGS17", "FMINLBFGS18", "FMINLBFGS19", "FMINLBFGS20", "FMINLBFGS21", "FMINLBFGS22", "FMINLBFGS23", "FMINLBFGS24", "FMINLBFGS25", "FMINLBFGS26", "FMINLBFGS27", "FMINLBFGS28", "FMINLBFGS29", "FMINLBFGS30", "FMINLBFGS31", "FMINLBFGS32", "FMINLBFGS33", "FMINLBFGS34", "FMINLBFGS35", "FMINLBFGS36", "FMINLBFGS37", "FMINLBFGS38", "FMINLBFGS39", "FMINLBFGS40", "FMINLBFGS41", "FMINLBFGS42", "FMINLBFGS43", "FMINLBFGS44", "FMINLBFGS45", "FMINLBFGS46", "FMINLBFGS47", "FMINLBFGS48", "FMINLBFGS49", "FMINLBFGS50", "FMINLBFGS51", "FMINLBFGS52", "FMINLBFGS53", "FMINLBFGS54", "FMINLBFGS55", "FMINLBFGS56", "FMINLBFGS57", "FMINLBFGS58", "FMINLBFGS59", "FMINLBFGS60", "FMINLBFGS61", "FMINLBFGS62", "FMINLBFGS63", "FMINLBFGS64", "FMINLBFGS65", "FMINLBFGS66", "FMINLBFGS67", "FMINLBFGS68", "FMINLBFGS69", "FMINLBFGS70", "FMINLBFGS71", "FMINLBFGS72", "FMINLBFGS73", "FMINLBFGS74", "FMINLBFGS75", "FMINLBFGS76", "FMINLBFGS77", "FMINLBFGS78", "FMINLBFGS79", "FMINLBFGS80", "FMINLBFGS81", "FMINLBFGS82", "FMINLBFGS83", "FMINLBFGS84", "FMINLBFGS85", "FMINLBFGS86", "FMINLBFGS87", "FMINLBFGS88", "FMINLBFGS89", "FMINLBFGS90", "FMINLBFGS91", "FMINLBFGS92", "FMINLBFGS93", "FMINLBFGS94", "FMINLBFGS95", "FMINLBFGS96", "FMINLBFGS97", "FMINLBFGS98", "FMINLBFGS99", "FMINLBFGS100".
<code>objective</code>	Character giving the objective function to optimize. Possible values are: "LL" for the Log-Likelihood, "RMSE" for the Root Mean Square Error, "MAE" for the Mean Absolute Error, "MSE" for the Mean Squared Error, "MAPE" for the Mean Absolute Percentage Error, "RMSECV" for the Root Mean Square Error of Cross-Validation, "MAECV" for the Mean Absolute Error of Cross-Validation, "MSECV" for the Mean Squared Error of Cross-Validation, "MAPECV" for the Mean Absolute Percentage Error of Cross-Validation.
<code>parameters</code>	Initial values for the hyper-parameters. When provided this must be named list with elements: "variance", "correlation", "trend", "kernel", "regmodel", "normalize", "optim".

Details

The hyper-parameters (variance and vector of correlation ranges) are estimated thanks to the optimization of a criterion given by `objective`, using the method given in `optim`.

Value

← → ↻ 🏠 🔒 🔑 https://funz.github.io

Funz Docs Demos

Search...

Standardize your simulation as a function.

Funz is a parametric & parallel remote execution engine to wrap scientific computing as a function. Then, just use it as an objective function for any standard algorithm (eg. optimization). Funz is available for **Windows, Mac OS, Linux**, through command line (**Bash, Cmd.exe, Python, R**).

TL;DR

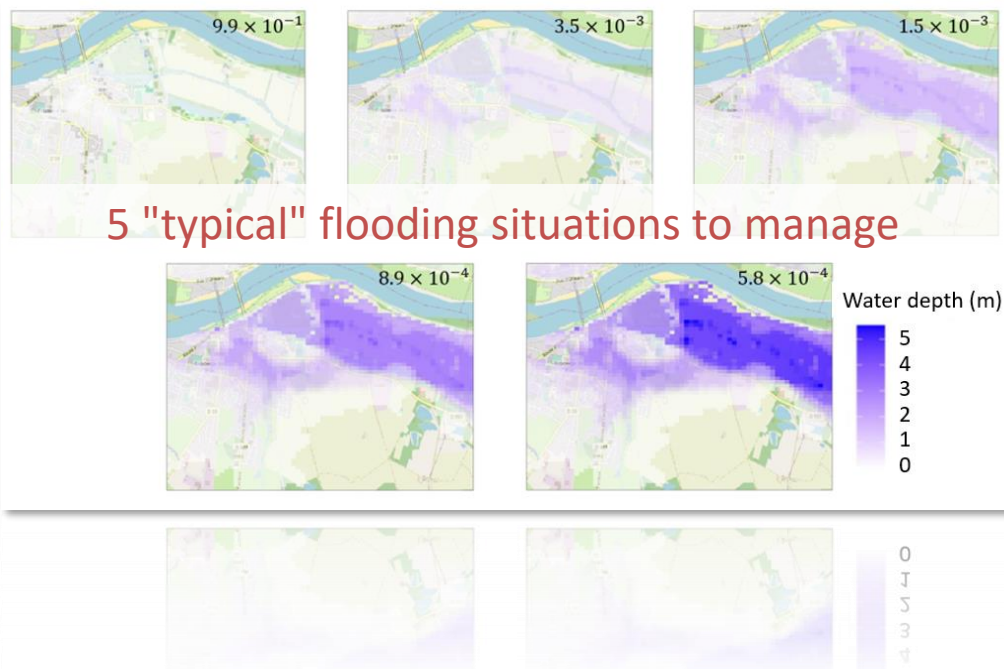
Parametric scientific computing ?

1. Take a "parametrized" input file (contains `$x`, `$y`, ... replacing some numerical values),
2. Define values taken by these parameter (or let a dedicated algorithm do),
3. Run calculations on remote resources,
4. Parse results & analyse.

... INTERPRETING FOR DECISION-MAKING

[Ex.] FLOODING RISK

1000 simulations of
the Loire flooding
(~40 kh.CPU)



Soutenance de thèse
Doctorat de l'université de Toulouse

Vers l'utilisation d'ensembles météorologiques pour la dispersion à courte distance de radionucléides en cas de rejets accidentels dans l'atmosphère

Youness El-Ouartassy

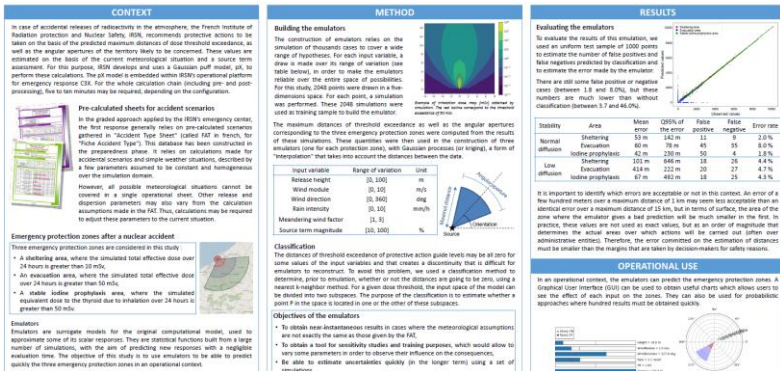
Membres du jury

V. Thourert
E. Blayo et Y. Roustan
L. Souhac, M. Rochoux et Y. Richet
M. Plu et I. Korsakissok
L. Descamps

Présidente
Rapporteurs
Examineurs
Directeurs de thèse
Co-encadrant

08 décembre 2023

Centre National de Recherches Météorologiques, Université de Toulouse, Météo-France, CNRS, Toulouse, France
Institut de Radioprotection et de Sûreté Nucléaire, PSE-SANTE/SESUC/BMCA, Fontenay-aux-Roses, France



Quantization methods for the visualization of the flooding risk

Charlie SIRE^{1,2,3}

Supervisors: R. LE RICHE³, D. RULLIERE³, J. ROHMER², L. PHEULPIN¹,
Y. RICHET¹

¹IRSN

²BRGM

³Mines Saint-Etienne and CNRS, LIMOS



WHAT'S NEXT ?

Major challenges remain in

- application field:
 - Communicating **uncertainties to decision** makers
 - Integrating **operational constraints** early in expertise
- ... and (still) in math. modeling tools:
 - Draw a path toward **standards in high-dim projections** (ex. similar to metamodel requirements)
 - Integrate **intrinsic physical properties** in metamodels (ex. "Physics Informed **", so mitigate **interpret-/explain-ability** tenet)