

CLUSTERING METHODS FOR DECISION-MAKING

Application to Flood Risks and Radiological Emergencies

I. Korsakissok, Y. Richet

Institut de Radioprotection et de Sûreté Nucléaire

Crédits F. Dubray / Ouest France 2021



Crédits IRSN

DECISIONS FACING NATECH RISKS



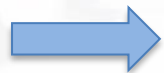


RISKS MANAGEMENT

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 »

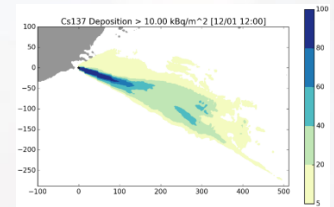
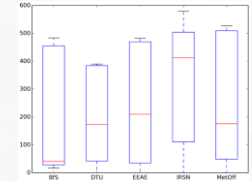
L'activation tardive de la vigilance orange, alors que les rafales ont atteint 200 km/h par endroits, illustre la difficulté à traduire les probabilités de phénomènes météo en un système d'alerte crédible.

La Monde avec AFP
Publié le 18 août 2022 à 21h34, modifié le 24 juillet 2023 à 19h09 - Lecture 3 min.

UNCERTAINTY ANALYSIS & INTERPRETABILITY

Uncertainty models for quantitative analysis or decision-making

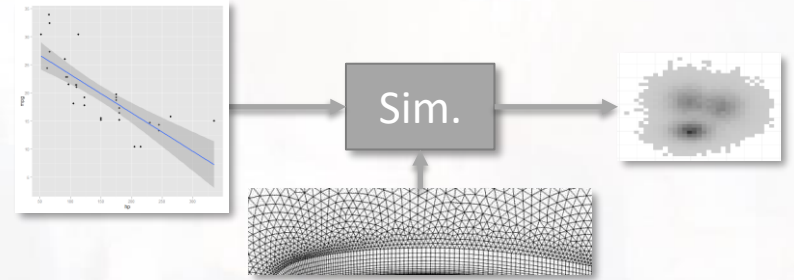
- Quantification of input / model uncertainties
 - Sensitivity analyses/indices (Sobol, Shapley, HSIC, ...)
 - "Envelope" of trust, probability of threshold exceedance...
 - Confidence level in evaluations
 - **Identification of representatives/prototypes**
- ✓ Should be decision-oriented, incl. practical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



Simplicity / Interpretability

UNCERTAINTIES QUANTIZATION

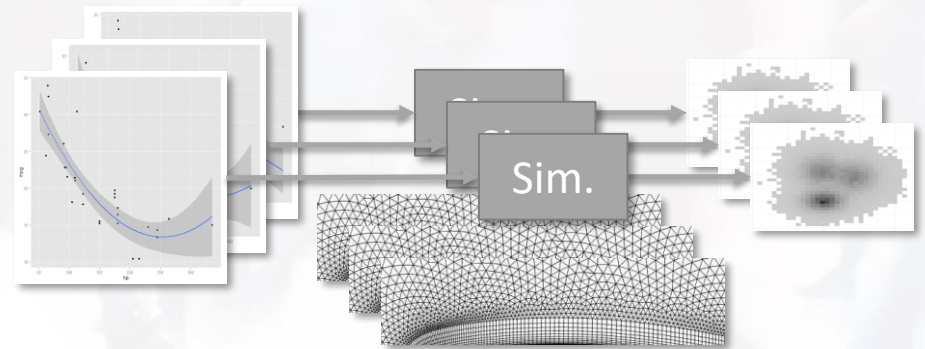
Simulations:



... a suitable support for propagation of uncertainties

Ex. :

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

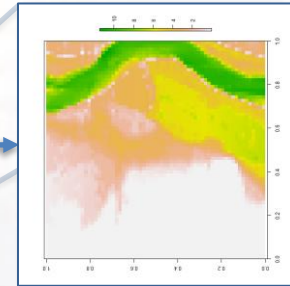
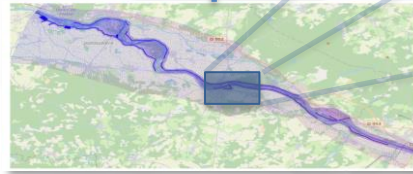
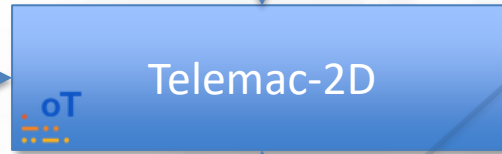
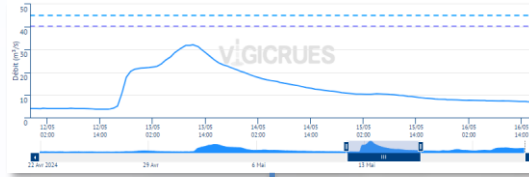
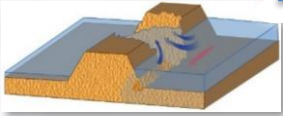
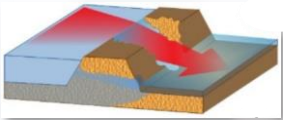


MODELLING HIGH-DIMENSIONAL DATA



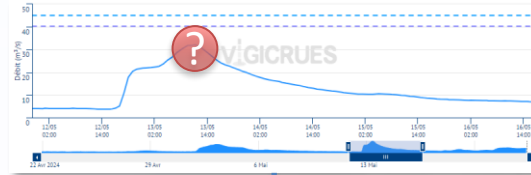
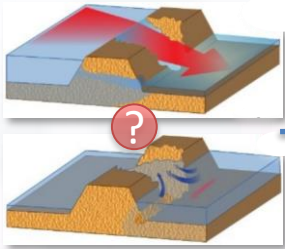
[EX.] FLOODING


Crédits DREAL Centre-Val de Loire

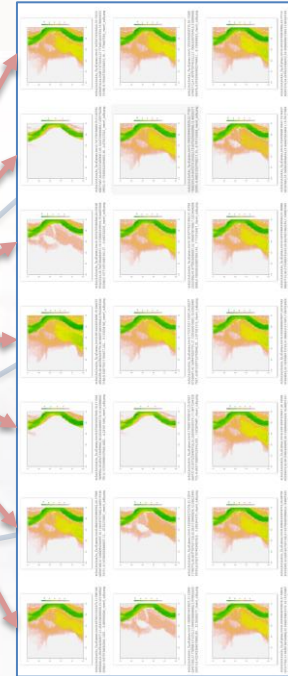
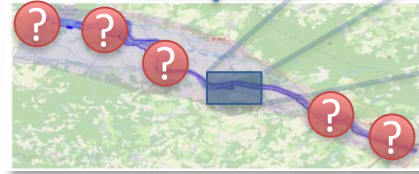


[EX.] FLOODING

Crédits DREAL Centre-Val de Loire



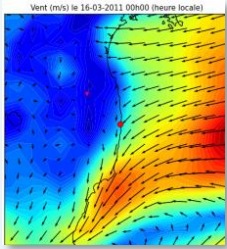
 **Telemac-2D**
40 min . 48 CPU / run



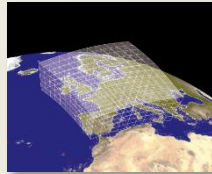
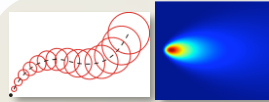
[EX.] RADIOLOGICAL EMERGENCY RESPONSE

Meteorology

- Observations
- 3D forecasts

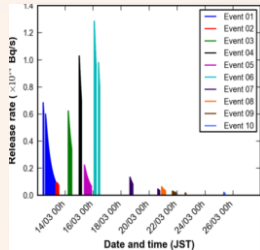


With *Météo France*



Dispersion models

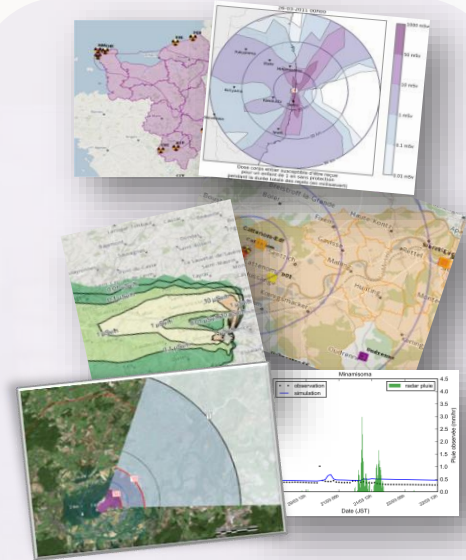
- Gaussian puff (pX)
- Eulerian (IdX)



With *operator*

Release

- Diagnosis
- Prognosis



Radiological consequences

- Air concentrations, dose...
- Maps, time evolution
- Zones of threshold exceedance

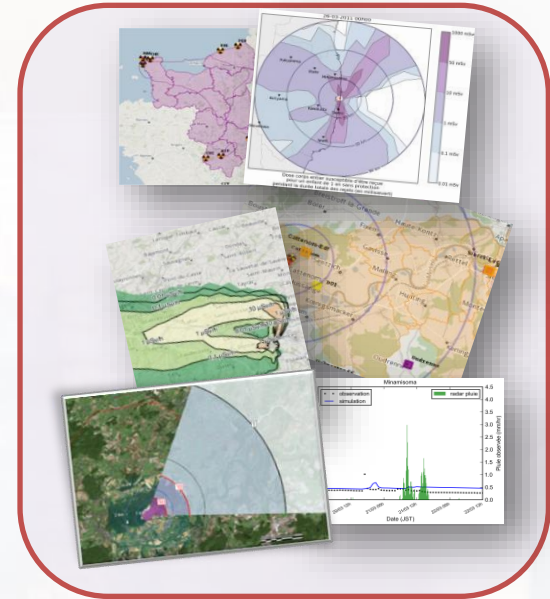
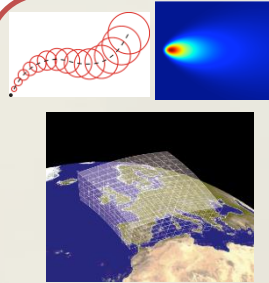
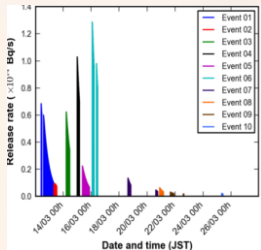
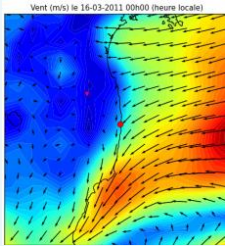
[Ex.] RADIOLOGICAL EMERGENCY RESPONSE

Epistemic
Stochastic
Modelisation

Physical and
numerical errors
User's decisions

Epistemic
Human errors, ambiguities
Lack of information
Modelisation

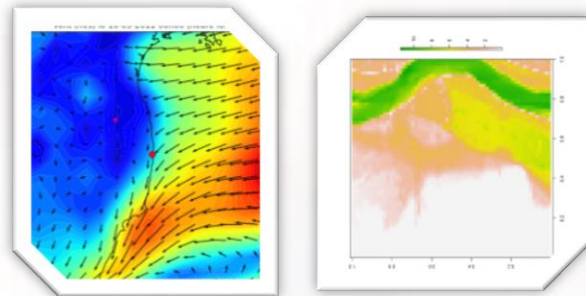
Choice of indicators and
output visualization
Interpretation
Decision making



HOW TO INCLUDE UNCERTAINTIES IN DECISION MAKING ?

Challenge 1: high-dimension inputs / outputs

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



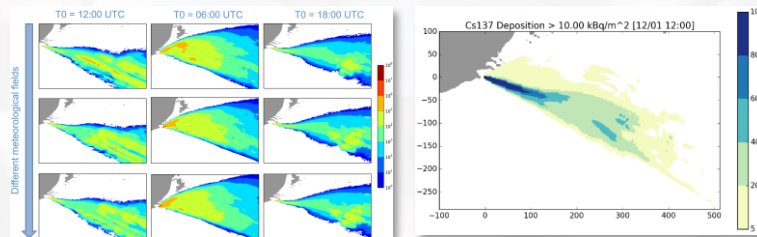
Challenge 2: Computational cost of physical models

- Use of **meta-models** or **surrogate models**

Challenge 3: interpretability of outputs

- Postage stamp ? Too many maps
- Probability maps ? Complex interpretation
- Scenario-based approach: “best estimate” vs. “worst case”

➡ Use of **clustering methods**



SURROGATE MODELLING

Parametric modelling to assess physical behaviour:

- High Performance Computing
- Numerical design of experiments (... but still curse of dimensionality)
- Agnostic response surface

... Training of a **surrogate model**
(*Gaussian Process Regression*)

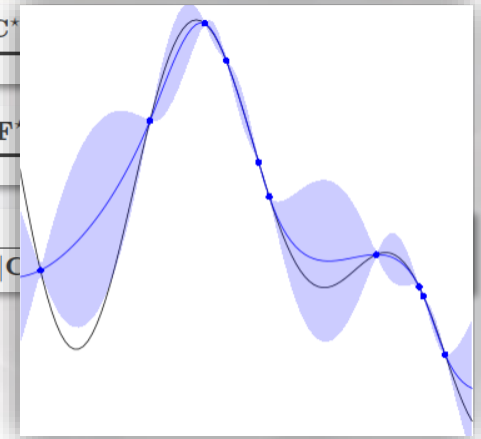
$$\mathbf{y}(\mathbf{x}_i) = \underbrace{\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}}_{\text{trend}} + \underbrace{\zeta(\mathbf{x}_i)}_{\text{smooth GP}}$$

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

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

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

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



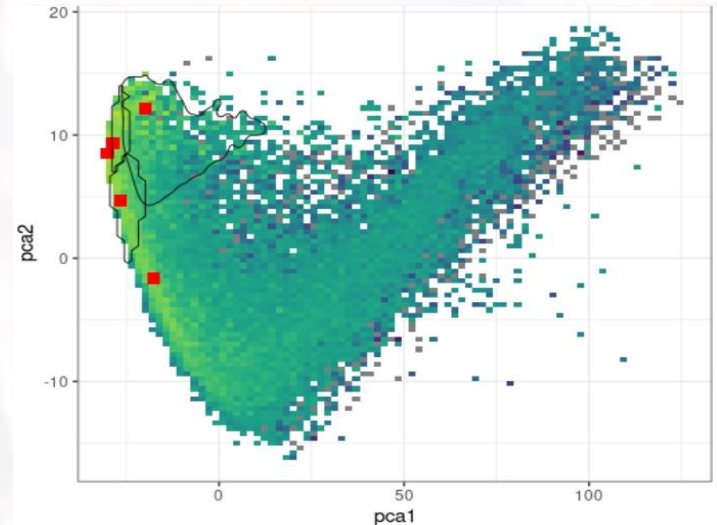
OUTPUTS PROJECTION

Complex numerical results:

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

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

- Supervised (prob. of occurrence)
/ unsupervised
- Non-significant "latent" space
- Desirable mathematical properties



CLUSTERING

Objective:
sparse & synthetic sampling

- Some **prototypes** / centroids
- Weighted / probabilistic classes
- "Real" <-> "Latent" space projection

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**

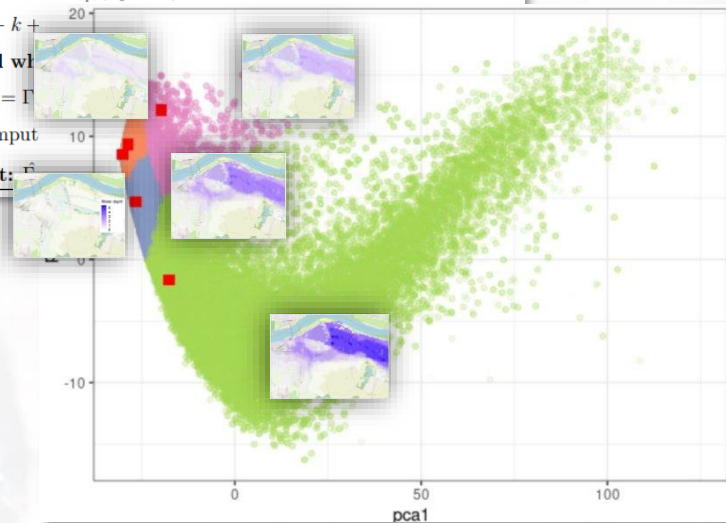
$$\gamma_j^{[k+1]} \leftarrow \mathbb{E} \left[Y(X) \mid Y(X) \in C_j^{\Gamma_\ell^{[k]}} \right], j \in \{1, \dots, \ell\}.$$

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 $\Gamma_\ell^{[0]} \leftarrow \{\gamma_0^{[0]}, \dots, \gamma_\ell^{[0]}\} \in \mathcal{Y}^{\mathcal{X}}$
- 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 $\hat{\Gamma}_\ell^*$

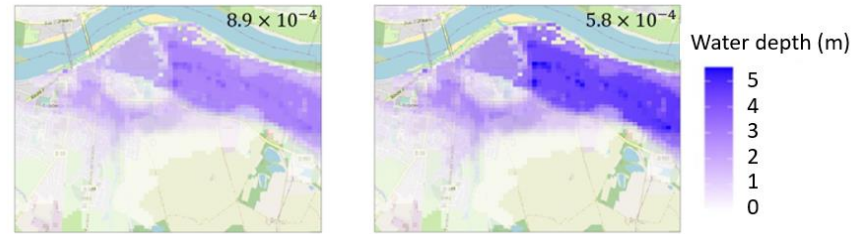
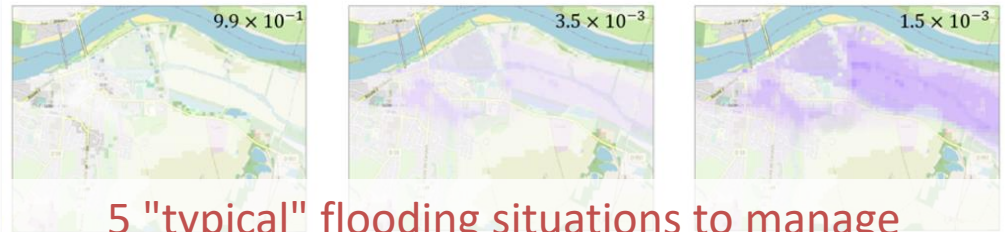
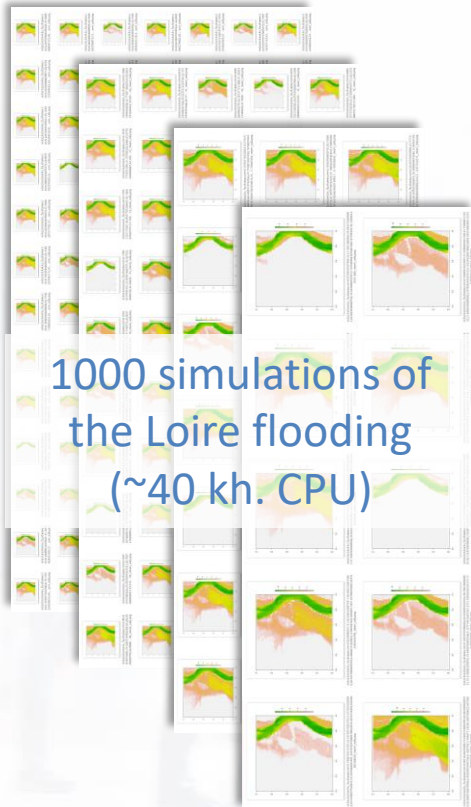
Output: $\hat{\Gamma}_\ell^*$



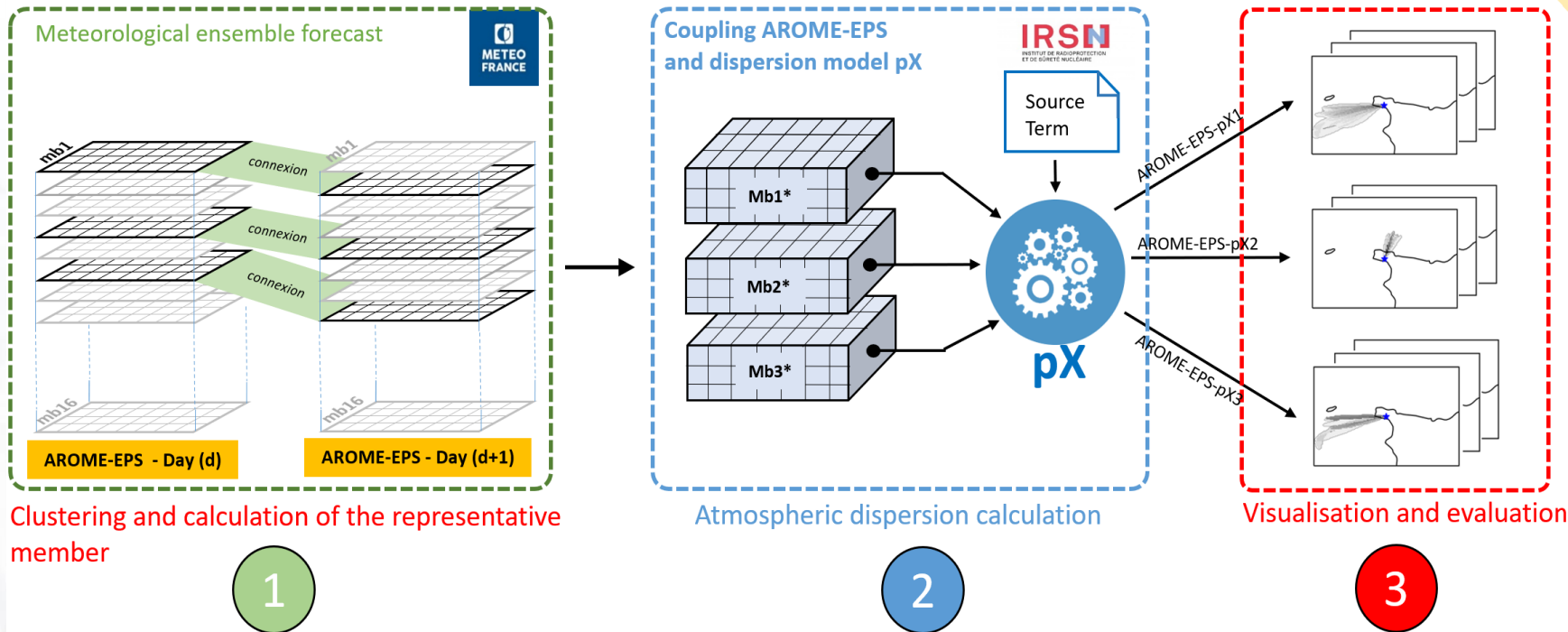
**... INTERPRETING
FOR DECISION-MAKING**

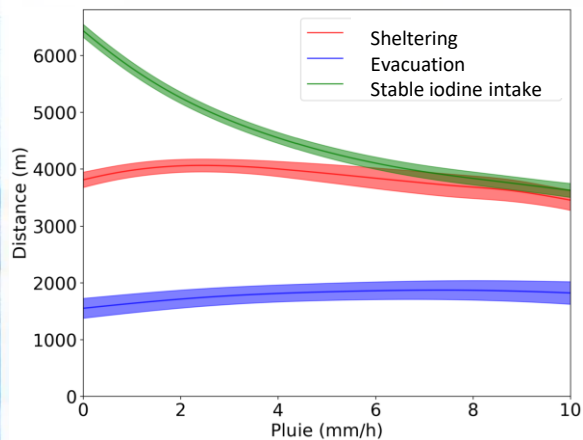


[EX.] FLOODING RISK

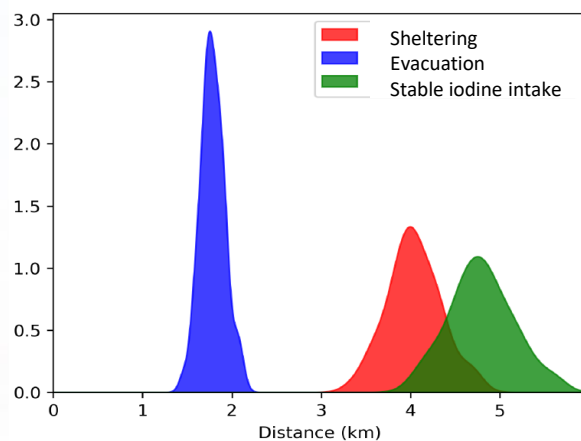


[Ex.] RADIOLOGICAL EMERGENCY RESPONSE



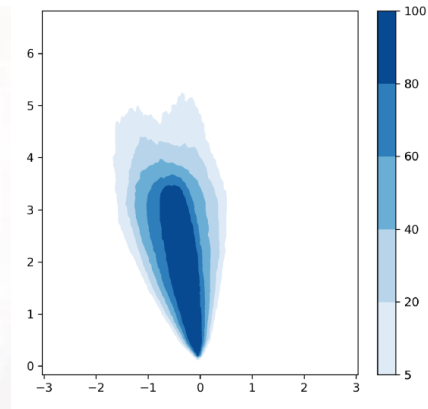


Sensitivity to input parameters (~20s)



Probabilistic assessment of distance and aperture (~20s)

Probability of threshold exceedance (%)
Stable iodine intake



Probability map of threshold exceedance (~2-4 min)



Consortium Industrie Recherche
pour l'Optimisation et la
Quantification d'incertitude
pour les données Onéreuses



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

Younes El-Ouartassy

Membres du jury

V. Thouret

E. Blayo et Y. Roustan

L. Souhac, M. Rochoux et Y. Richet

M. Plu et I. Korsakissok

L. Descamps

Présidente

Rapporteurs

Examinateurs

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



EMULATORS FOR THE RAPID PREDICTION OF CONSEQUENCES IN CASE OF NUCLEAR HAZARD

Raphaël Périllat¹, Sylvain Girard¹, Inès Korsakissok² et Emmanuel Quentrec²
¹ Phimeca Engineering, Paris, France ; ² IRSN, Fontenay-aux-Roses, France



CONTEXT

In case of accidental releases of radionuclides in the atmosphere, the French Institute of Radiation Protection and Nuclear Safety, IRSN, recommends protective actions to be taken on the basis of the practical maximum distance of dose threshold exceedance, as well as the larger spectrum of the territory likely to be concerned. These values are estimated on the basis of the current meteorological situation and a source term assessment. For this purpose, IRSN develops and uses a Gaussian puff model, to perform these calculations. The 3d model is embedded within IRSN's operational platform for emergency response C3E, for the whole calculation chain (including pre- and post-processing, from a few minutes may be required, depending on the configuration).

Pre-calculated sheets for accident scenarios

In the graded approach applied by the IRSN's emergency center, the first response generally relies on pre-calculated scenarios gathered in "Accident Type Sheet" (after IAD in French, for "Type Accident Type"). This database has been constructed in the preparation phase. It relies on calculations made for accident scenarios and simple weather conditions, described by a few parameters assumed to be constant and homogeneous over the simulation domain.

However, all possible meteorological situations cannot be covered in a "single operational sheet". Other weather and dispersion parameters may also vary from the calculation assumptions made in the IAD. Thus, calculations may be required to adjust these parameters to the current situation.

Emergency protection zones after a nuclear accident

Three emergency protection zones are considered in this study:

- a sheltering area, where the simulated dose effective dose over 24 hours is greater than 10 mSv,
- an evacuation area, where the simulated dose effective dose over 24 hours is greater than 10 mSv,
- a clean up/return prohibited area, where the simulated equivalent dose to the thyroid due to inhalation over 24 hours is greater than 10 mSv.

Emulators
Emulators are surrogate models for the original computational model, used to approximate some of its input responses. They are statistical functions built from a large number of simulations, with the aim of predicting new responses with a negligible additional cost. The objective of this study is to use emulators to be able to predict quickly the three emergency protection zones as an operational context.

METHOD

Building the emulators

The construction of emulators relies on the simulation of thousands of cases to cover a wide range of hypotheses. For each input variable, a data is made over a range of variation (see table below), in order to make the emulators flexible over the entire space of possibilities. For this study, 2048 points were drawn in a five-dimensional space. For each point, a simulation was performed. These 2048 simulations were used to training samples to build the emulator.

The maximum distances of threshold exceedance as well as the angular aperture corresponding to the three emergency protection zones were captured from the results generated in these simulations. These quantities were then used in the construction of three emulators (one for each protection zone), with Gaussian priors (or priors), a kernel of "interpolation" that takes into account the distances between the data.

input variable	range of variation	unit
release height	[0, 100]	m
wind module	[0, 10]	m/s
wind direction	[0, 360]	deg
rain intensity	[0, 10]	mm/h
Manufacturing wind factor	[1, 3]	-
Source term magnitude	[100, 1000]	Bq

Classification
The distance of threshold exceedance of protective action guide levels may be all zero for some values of the input variables and that creates a discontinuity that is difficult for emulators to reconstruct. To avoid this problem, we used a classification method to determine, prior to evaluation, whether or not the distance is going to be zero, using a nearest neighbor method. For a given case threshold, the input space of the model can be divided into two subspaces. The purpose of the classification is to remove whether a point in the space is located in one or the other of these subspaces.

Objectives of the emulators

- To obtain new instantaneous results to copy when the meteorological assumptions are not exactly the same as those given by the IAD.
- To obtain a tool for sensitivity studies and training purposes, which would allow to explore some parameters in order to observe their influence on the consequences.
- To be able to estimate uncertainties quickly (in the longer term) using a set of scenarios.

RESULTS

Evaluating the emulators

To evaluate the results of the emulators, an error on uniform test sample of 1000 points to estimate the number of false positives and false negatives by classification and to estimate the error made by the emulator. There are all some false positive or negative cases. Between 1.8 and 0.2%, but these numbers are much lower (even without classification) between 3.7 and 44.0%.

Stability	Area	Mean error	Q95% of error	Ratio false	Ratio error
Normal	Sheltering	53 m	142 m	11	9
	Evacuation	60 m	78 m	49	39
	Return prohibited	42 m	230 m	50	4
Low	Sheltering	103 m	646 m	18	26
	Evacuation	414 m	222 m	20	27
	Return prohibited	87 m	402 m	18	25

It is important to identify which errors are acceptable or not in the context. An error of a few hundred meters over a maximum distance of 1 km may seem less acceptable than an identical error over a maximum distance of 20 km, but in terms of surface, the area of the zone where the emulator gives a bad prediction will be much smaller in the first. In practice, these values are not used in each table, but as an order of magnitude that determines the actual areas over which actions will be carried out (after our administrative estimates). Therefore, the error committed on the estimation of distance must be smaller than the margins that are taken to decision-makers for safety reasons.

OPERATIONAL USE

In an operational context, the emulators can predict the emergency protection zones. A Gaussian Puff model (C3E) can be used to obtain useful charts which allow users to see the effect of each input on the zones. They can also be used for probabilistic approaches where hundred results must be displayed quickly.

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. RICHER¹

¹IRSN

²BRGM

³Mines Saint-Etienne and CNRS, LIMOS



Unlock the future

WHAT'S NEXT ?

Major challenges remain in

- **communicating uncertainties** to decision makers
- **integrating operational constraints** early in expertise

Bridging the gap between social sciences and geoscience / risk assessment and between different kinds of risks



Ambition of [PEPR RISQUES](#), work on Natch [scenarios](#)

Unlock the future