CENTRE EUROPÉEN DE RECHERCHE ET DE FORMATION AVANCÉE EN CALCUL SCIENTIFIQUE

Meta-modeling and GSA in large dimension in fluvial hydrodynamics

CERFACS-OpenTURNS November, 22nd 2021

S. Ricci

Acknowledgment: S. El Garroussi, M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LISN)

Uncertainty Quantification (UQ) @ Cerfacs

Data Driven Modeling Axis @CERFACS

Objective: develop and apply uncertain quantification algorithms for complex problems

Actions:

1 - Ensemble generation and Design of Experiment for functional variables with space reduction solutions

2 - Development and evaluation of objective oriented cost-effective Surrogate Model methodologies for large dimension problems

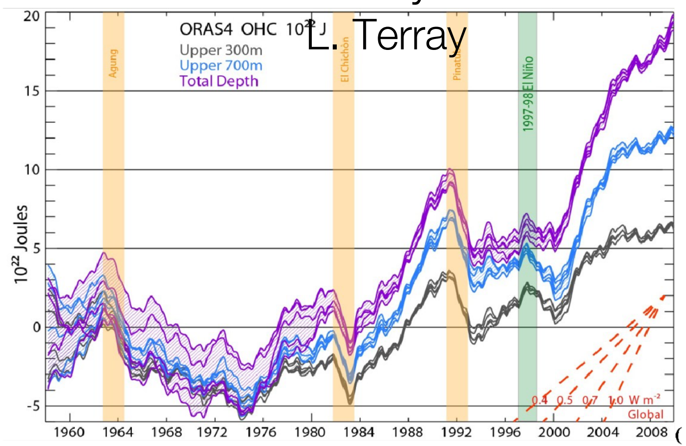
3- Efficient use of surrogate models for sensitivity analysis, optimization, data assimilation

4 - Development of HPC efficient algorithms for stochastic estimation with solvers of increasing complexity (multi-fidelity, Multi-Level Monte Carlo)

Uncertainty Quantification @ CERFACS

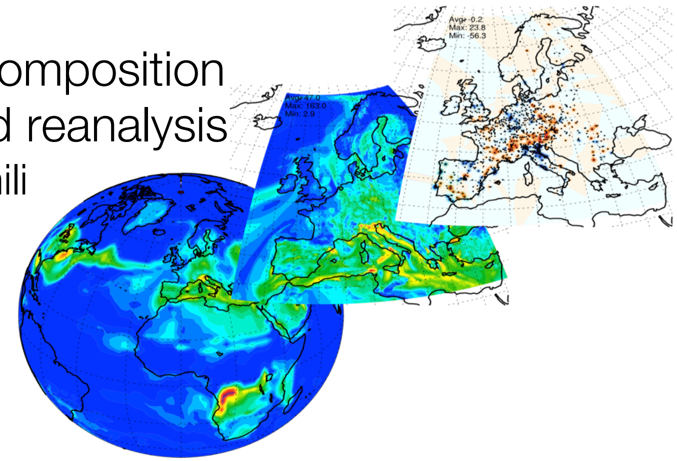
Climate/ocean forecasting and reanalysis

L. Terray



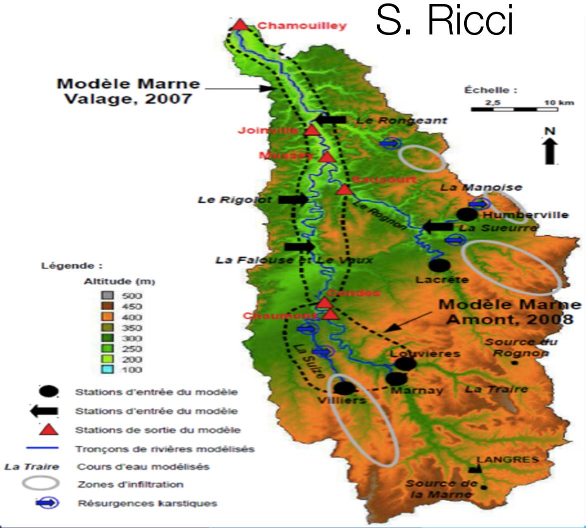
Atmospheric composition forecasting and reanalysis

E. Emili



Flood forecasting

S. Ricci

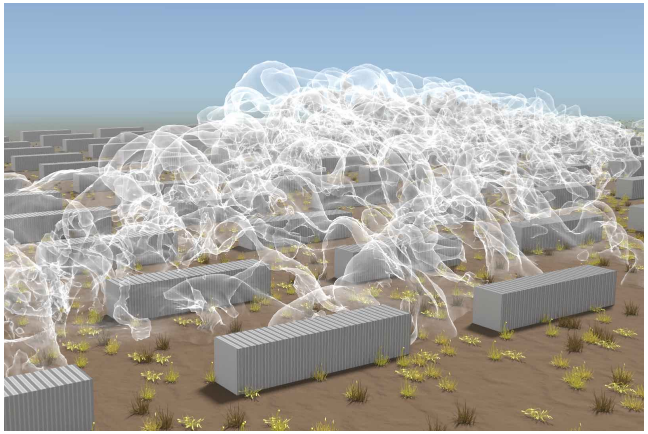


Algorithms and methods

P. Mycek

Pollutant dispersion reanalysis

M. Rochoux, T. Jaravel



Use Cases

- Use Case #1: **Development of surrogate models for sensitivity analysis and ensemble-based data assimilation. Application in hydraulics for large dimension uncertain variables.**
- Use Case #2: Development of UQ algorithms for exascale computing with efficient, scalable and resilient domain decomposition algorithms
- Use Case #3: Surrogate models for large-eddy simulations of atmospheric boundary layer to emulate micro-scale meteorology and assess structural and parametric uncertainty. Development for pollutant dispersion application.
- Use Case #4: Application of Multifidelity-MLMC algorithms for industrial CFD, multi-disciplinary systems and geosciences.
- Use Case #5: Building testbed for global climate projection uncertainty. Use emulation/simple climate models to test constraints on future climate projections and improve existing scenario projections, assess uncertainties in long term warming trajectories and improve consistency between physical system models and socioeconomic models.

Resources

PhDs/Post-docs @CERFACS:

- B. Nony (M. Rochoux), appli dispersion polluants
- E. Lumet (M. Rochoux), appli assim capteurs mobiles
- S. Peatier (B. Sanderson), appli climat
- S. Salle (N. Gourdain-ISAE, S. Ricci,), GSA impact climat sur aviation
- R. Despoeys (S. Ricci, M. Balesdent- ONERA), Multi-fidelité pour optimisation robuste et dépassement de seuil
- R. El Amri (P. Mycek, S. Ricci, M. De Lozzo – IRT), Multi-fidelité, GSA pour MDO-MDA avec UQ
- S. El Garroussi (S. Ricci), appli hydrodynamique fluviale

Surrogate modeling for flood in Hydrodynamics

[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

From deterministic to uncertain flood forecast

S. El Garroussi et al., Simhydro 2021

Vigicrues.gouv.fr: carte 24/11/2016

LeMonde.fr/article/2020/10/05/bilan-meurtrier-precipitations-records-evolution

- Uncertainties in data,
- Uncertainties in human expertise,
- Uncertainties related to the nature of certain events,
- The limits of vigilance at the departmental level, ...

- Uncertainties related to observations,
- Uncertainties related to numerical models for meteorological, hydrological and hydraulic forecasting.

CEPRI, Prévision et anticipation des crues et des inondations, 2018

Shallow water equations (2D formulation) [1871]

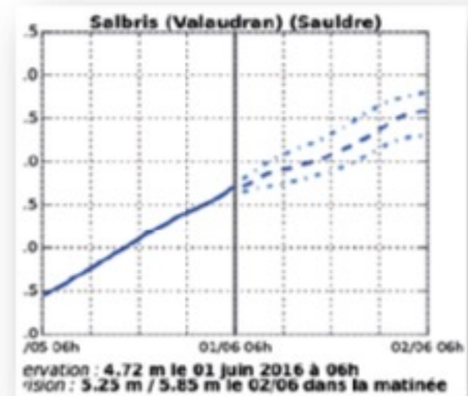
$$\frac{\partial h}{\partial t}+div(h\vec{u})=0$$

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}=-g\frac{\partial Z_s}{\partial x}+\boxed{F_x}+\frac{1}{h}div(h\nu_e\, grad(u))$$

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}=-g\frac{\partial Z_s}{\partial y}+\boxed{F_y}+\frac{1}{h}div(h\nu_e\, grad(v))$$

Advection | Gradient of hydrostatic pressure | Source terms, friction | Diffusion Turbulence Dispersion

- Assumptions of the model
- Boundary conditions
- Initial conditions
- Bathymetry
- Hydraulic forcing



Example of a forecast graph distributed by the SPC "Loire-Cher-Indre" during the floods of June 2016

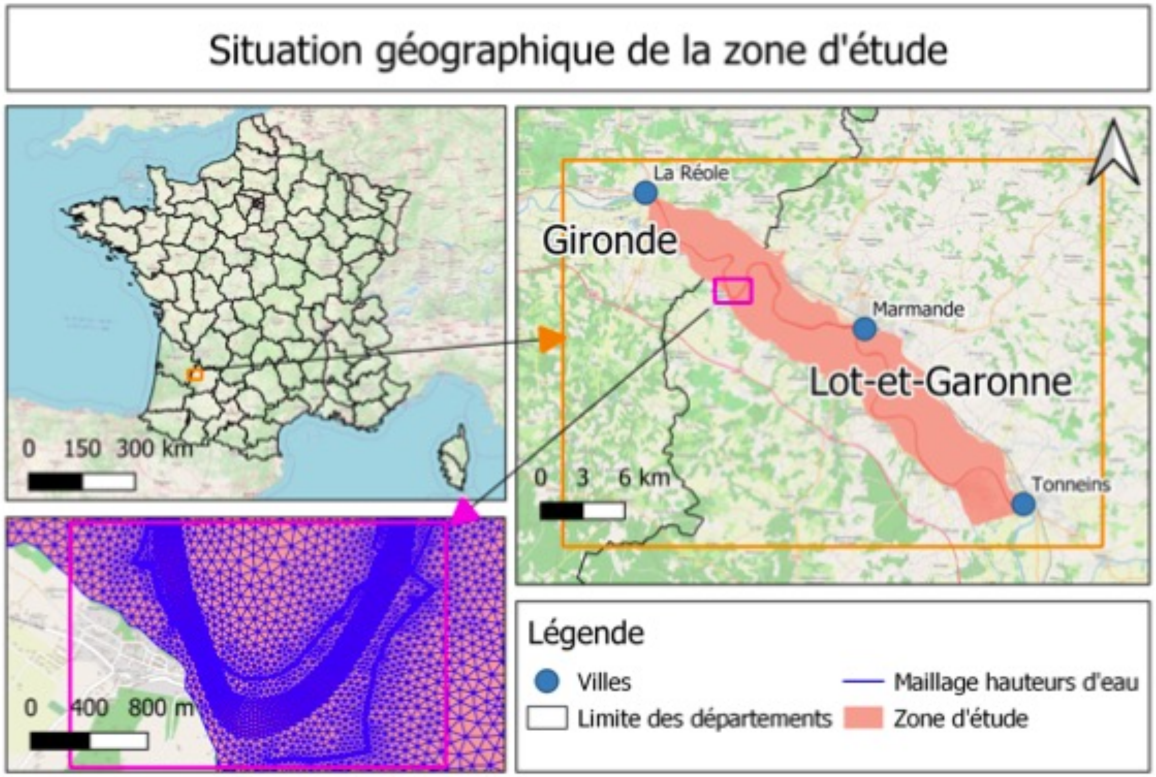
Flood forecasting models are prone to uncertainties; these uncertainties must be reduced to improve their prediction.

[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

Study area and experimental settings

S. El Garroussi et al., UNECECOMP 2021

- The study area extends over a 50 km reach of the Garonne river from Tonneins (upstream) to the confluence with the rivers Lot and La Réole (downstream).
- Uncertain parameters: upstream discharge and 4 friction coefficients; assumed to be independent.
- Variable of interest: water depth map (41,416 mesh nodes).
- The upstream discharge is assumed to be normal and the friction coefficients are assumed to be uniform.
- A T2D simulation of a 3-day flood event lasts 20 minutes.
- The size of the training sample for the surrogate model is 1,000 T2D simulations. The validation sample size is 500 T2D simulations.



Study area of the Garonne river (southwest France) 50-km reach between Tonneins (upstream) and La Réole (downstream), the blue circles represent the in-situ Vigicrue stations and the pink zoom shows the spatial discretization.

Surrogate modeling for flood in Hydrodynamics

[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

PhD S. El Garroussi

Building a meta-model for for large output space in the presence of non linearities between uncertain inputs and outputs

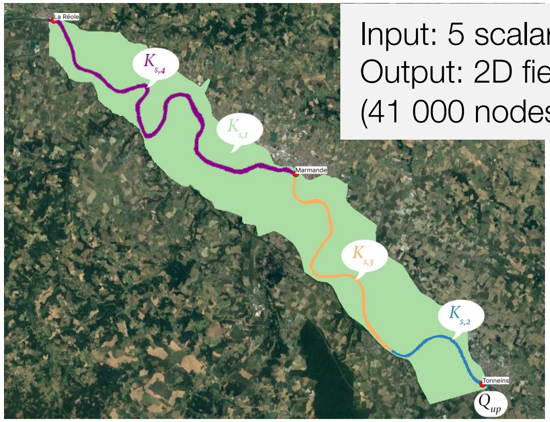
General Objective

Build a surrogate to approximate 2D WSE in flood plains

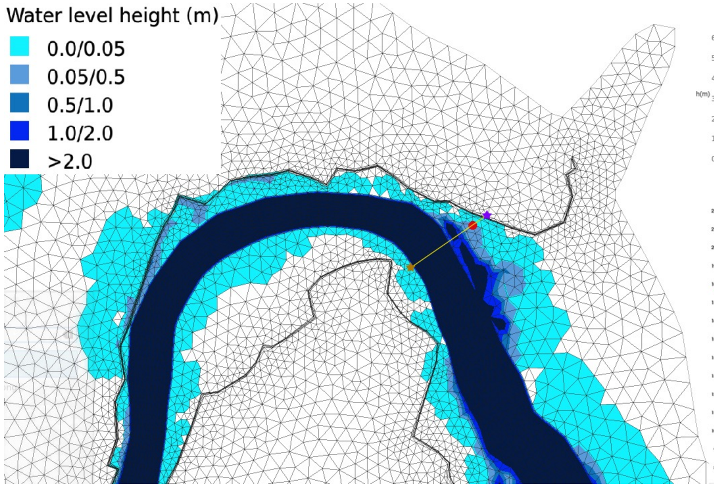
- carry out a GSA with respect to friction and inflow
- identify control vect. for uncertanity reduction with Data Assimilation

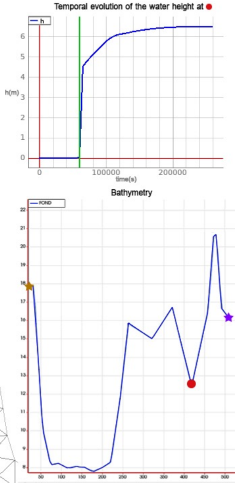
Difficulties

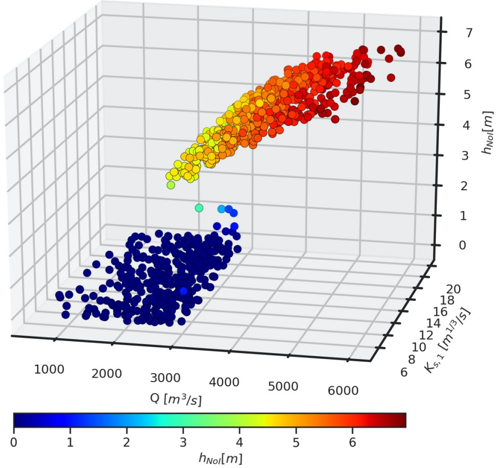
- non-linearities between input and output space, especially in region with strong bathymetry gradients
- large dimension output space



Input: 5 scalars
Output: 2D field
(41 000 nodes)







[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

Uncertainty quantification in a two-dimensional river hydraulic model

El Garroussi et al., UNECECOMP, 2019

Classical approach

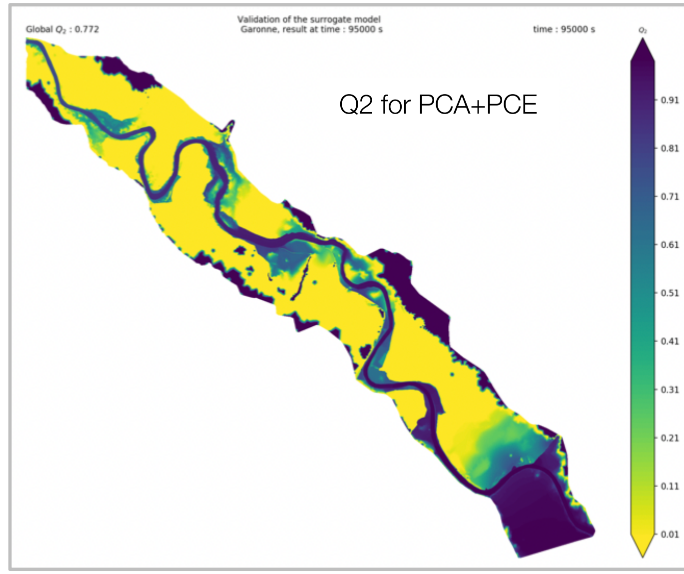
- Principal Component Analysis for reduction of the output space dimension
- A unique Polynomial Chaos Expansion in the reduced input space over the entire output domain

Polynomial chaos expansions (PCE) of (2nd order) stochastic processes (Wiener, 1938):

$$\underline{y} = \sum_{\alpha \in \mathbb{N}^d} \gamma_\alpha \psi_\alpha(\underline{x}),$$

where:

- $\underline{x} = (\underline{x}_1, \ldots, \underline{x}_d)$ a set of independent second order random variables with given joint density $f_{\underline{x}}$,
- $\psi_\alpha(\mathrm{x}) = \prod_{i=1}^{d} \psi_{i,\alpha_i}(x_i)$ is a tensor product of univariate orthonormal polynomials, i.e. $\mathbb{E}\left[\psi_{i,j}(\underline{x}_i)\psi_{i,k}(\underline{x}_i)\right] = \int_{D_{x_i}} \psi_{i,j}(x_i)\psi_{i,k}(x_i) f_{\underline{x}_i}(x_i) dx_i = \delta_{jk}$,
- γ_α is the deterministic coefficient associated with ψ_α. Knowledge of the γ_α, using least square approximation, for example, fully characterizes the process $\underline{y}$.



Merits and Limitations

+ Thanks to the orthogonality of the basis, one can immediately deduce the Sobol decomposition of the PCE

- A unique Polynomial Chaos Expansion over the entire domain does not account for non-linearities (Li and Ghanem 1998) or stochastic discontinuities (Najm, 2009)

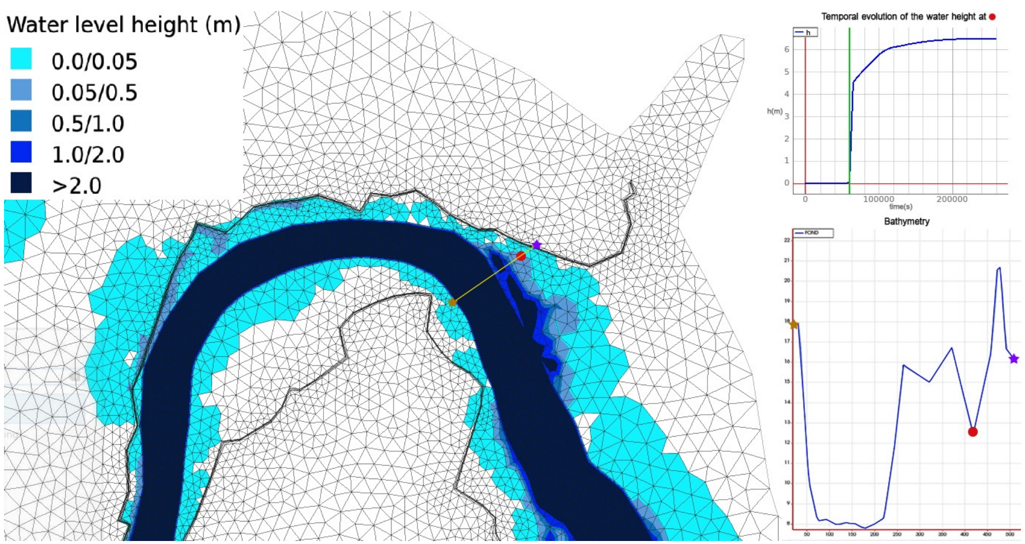
Surrogate modeling for flood in Hydrodynamics

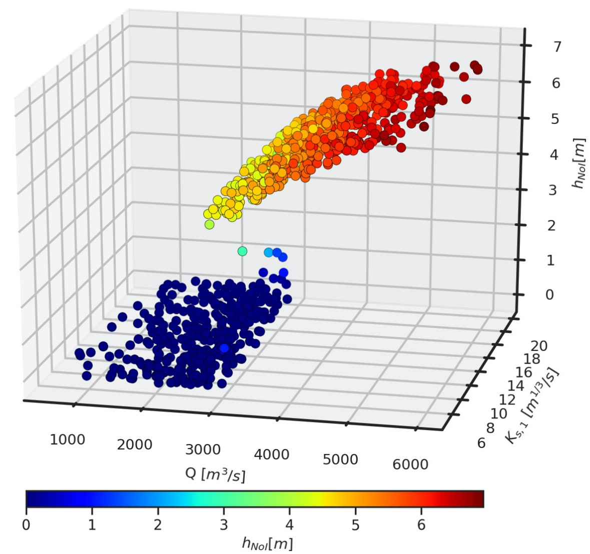
[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

Assessing uncertainties in flood forecasts using a mixture of generalized polynomial chaos expansions

S. El Garroussi et al.
Proceedings TUC2020

Focus:
Dealing with non linearities, **at one critical node with MoE**





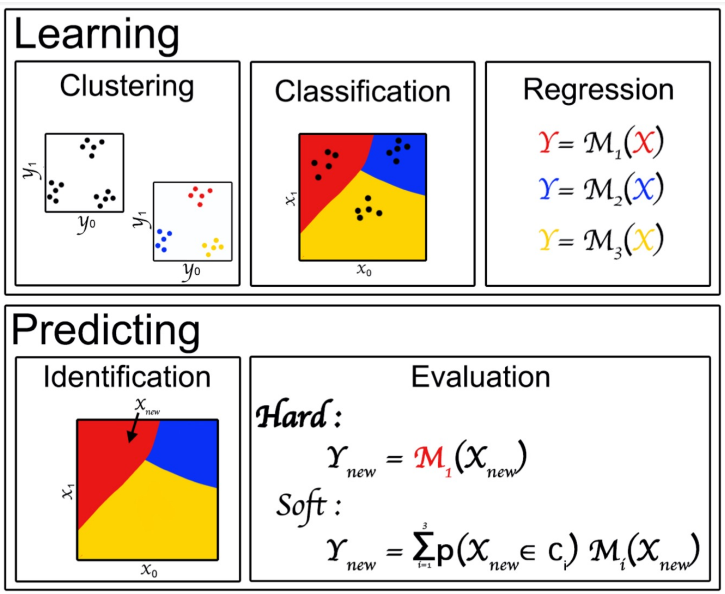
Advanced methodology

- Dimension reduction for non linear functional outputs
- Mixture of experts for each grid point

Surrogate modeling for flood in Hydrodynamics

[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

Mixture of experts for local solution **at one critical node**



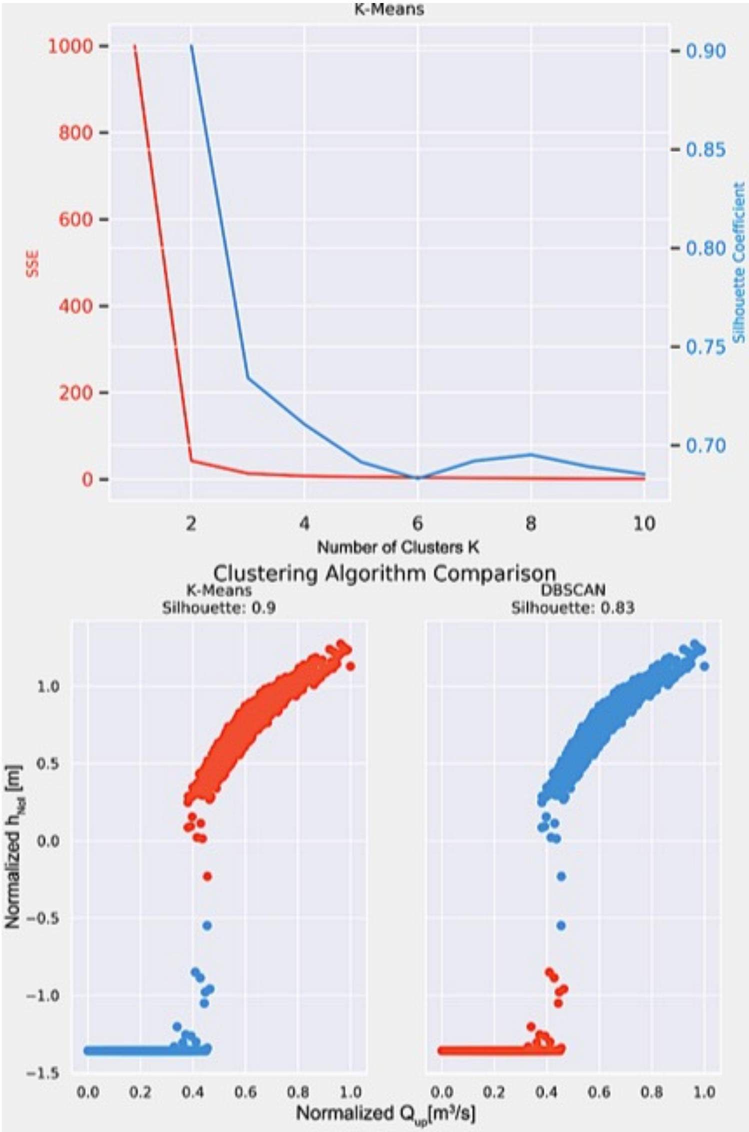


Figure 5: A. The top curve represents the evolution of the two metrics: the SSE in red, and the silhouette coefficient in blue for the K-Means clustering method. B. Bottom left, a comparison of clustering methods: K-Means and DBSCAN.

Clustering - classifying

Machine Learning technics to partition the input space into clusters where non linearities do not occur

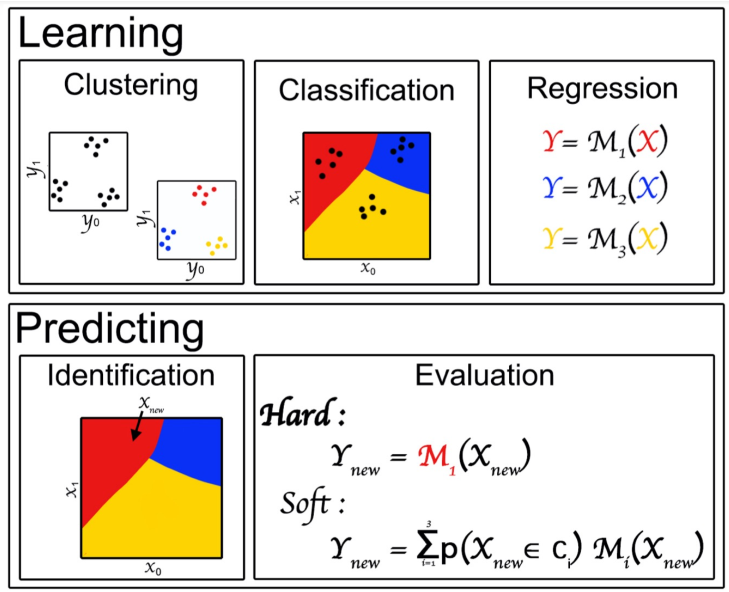
GMM, K-Means from Scikit Learn

S. El Garroussi et al.
Proceedings TUC2020

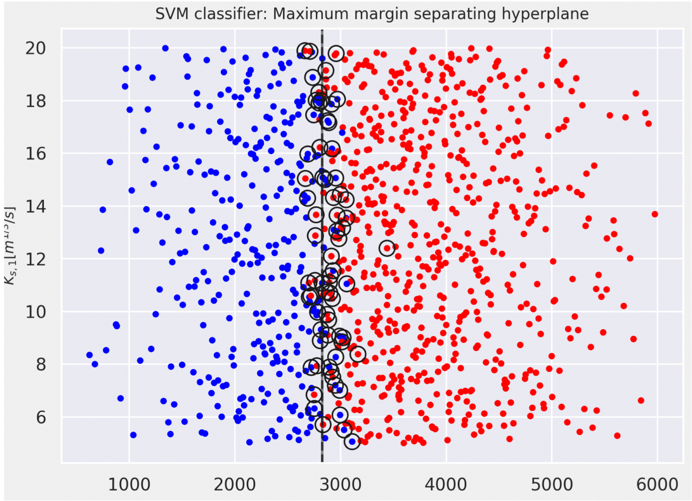
Surrogate modeling for flood in Hydrodynamics

[S. El Garroussi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

Mixture of experts for local solution **at one critical node**



Clustering with K-means in input space



Clustering - classifying

Machine Learning technics to partition the input space into clusters where non linearities do not occur

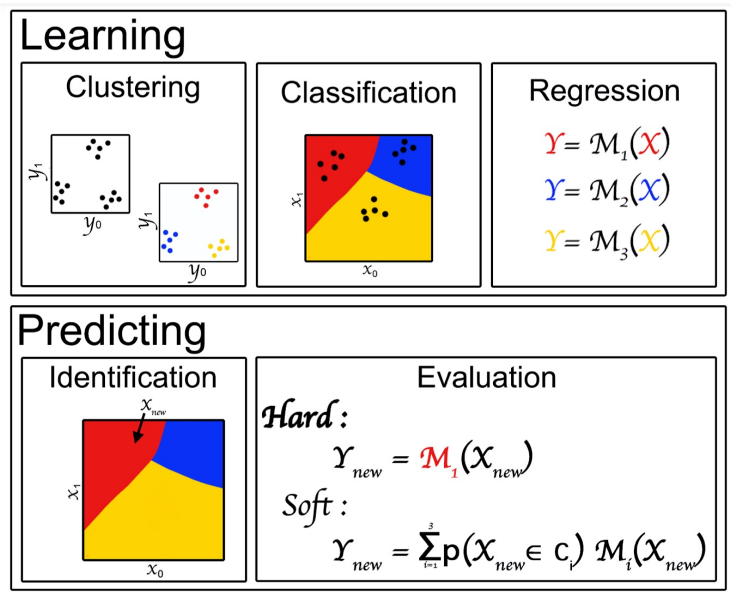
GMM, K-Means from Scikit Learn

S. El Garroussi et al.
Proceedings TUC2020

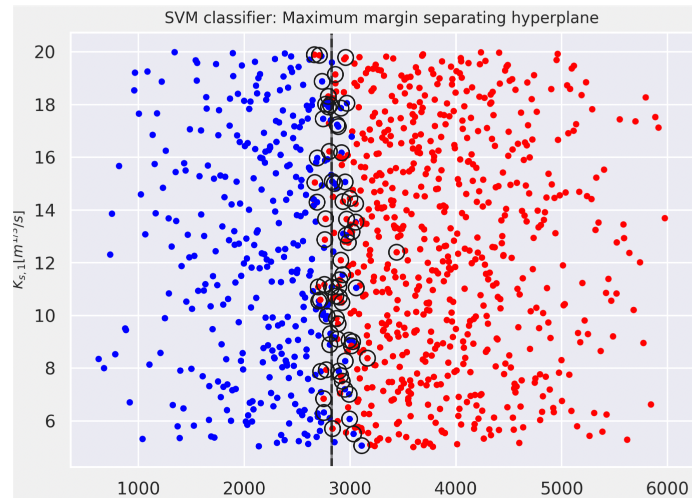
Surrogate modeling for flood in Hydrodynamics

[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

Mixture of experts for local solution **at one critical node : Divide and conquer**



Clustering with K-means in input space



Clustering - classifying

Machine Learning technics to partition the input space into clusters where non linearities do not occur

GMM, K-Means from Scikit Learn

S. El Garroussi et al.
Proceedings TUC2020

Surrogate modeling for flood in Hydrodynamics

[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

S. El Garroussi et al.
Proceedings TUC2020

Mixture of experts for local solution at **all critical nodes** (one gPCE by critical node) and **GSA**

Surrogate models' performance	*gPCE*	*MoE-gPCE*
RMSE [m]	0.92	0.02
Q_2	0.54	0.98

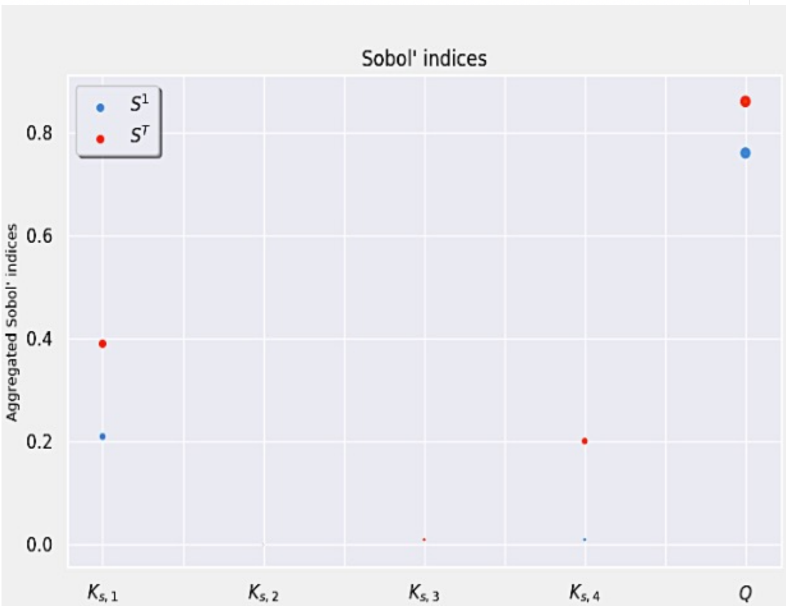


Fig. 7: Sobol' indices for water height at node NoI.

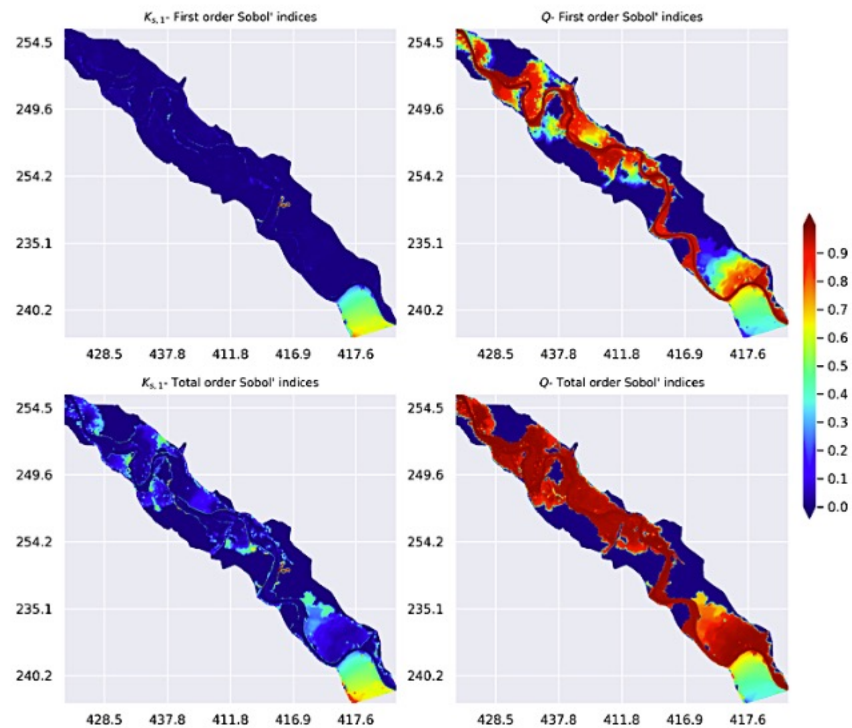


Fig. 8: Spatial aggregated Sobol' indices, first at the top and total in the bottom, for the water height discretized over the mesh following the uncertain inputs: $K_{s,1}$ on the left and Q on the right.

Limitations

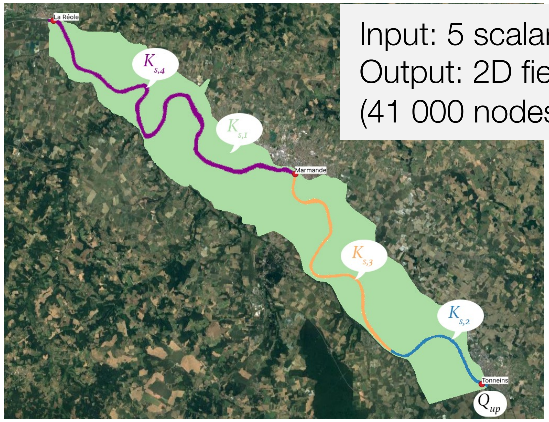
- Computational cost of building 72 000 gPCE (3 000 critical nodes, 8 values PCD degree, 3 classes)
- Need to work on the output dimension in the context of MoE strategy

Surrogate modeling for flood in Hydrodynamics

[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

Tackling random fields non-linearities with unsupervised clustering of polynomial chaos expansion in latent space: application to global sensitivity analysis of river flooding

S. El Garroussi, S. Ricci, M. De Lozzo, N. Goutal, D. Lucor, SERRA, 2021



Input: 5 scalars
Output: 2D field
(41 000 nodes)

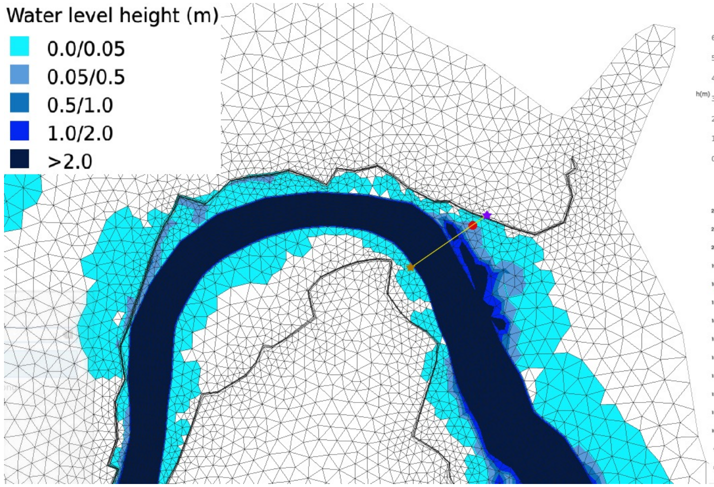
General Objective

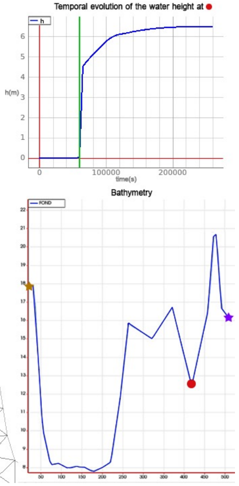
Build a surrogate to approximate 2D WSE in flood plains

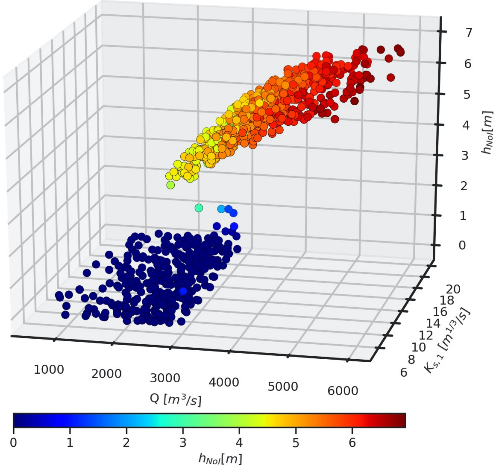
- carry out a GSA with respect to friction and inflow
- identify control vect. for uncertanity reduction with Data Assimilation

Difficulties

- **non-linearities between input and output space, especially in region with strong bathymetry gradients**
- **large dimension output space**







Surrogate modeling for flood in Hydrodynamics

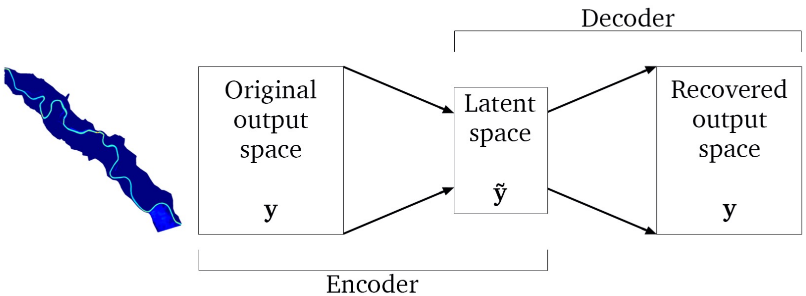
[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

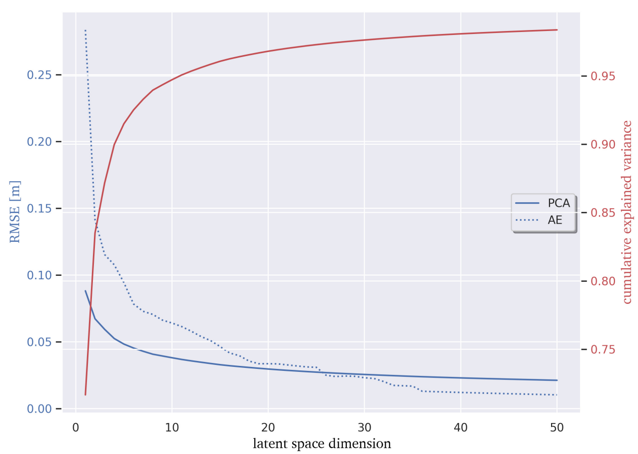
Advanced methodology

- Mixture of experts
- Dimension reduction for non linear functional outputs

Dimension Reduction of the output space

- Need to lower the cost of the surrogate formulation when working over the entire field





RMSE of WSE from learning DB and reconstructed WSE with for PCA and AE

DIMENSION REDUCTION

- Principal Component Analysis
- Auto-encoder (Keras Tensor Flow) based on unsupervised artificial neural network

37 components :
RMSE PCA = 3.82 cm
RMSE AE = 1.27 cm

Comp. Cost AE (2h) >> Comp. Cost PCA (3min)

Max Absolure Error PCA >3m
Max absolute Error AE < 1cm

S. El Garroussi, S. Ricci, M. De Lozzo, N. Goutal, D. Lucor, SERRA, 2021

Surrogate modeling for flood in Hydrodynamics

[I. Raguet (ENM), L. Gibert(ENM), M. Magnan (ENM), A. Mengin (ENM), S. El Garrousi (CERFACS), S. Ricci (CERFACS)]

- Analyse en Composantes Principales (PCA – Scikit Learn)
- Analyse en Composantes Principales à Noyau (KPCA – Scikit Learn) : prise en compte des non linéarités
- Transformation de Karhunen-Loève (KL-SVD - OpenTURNS) : prise en compte des hétérogénéïtés du maillage

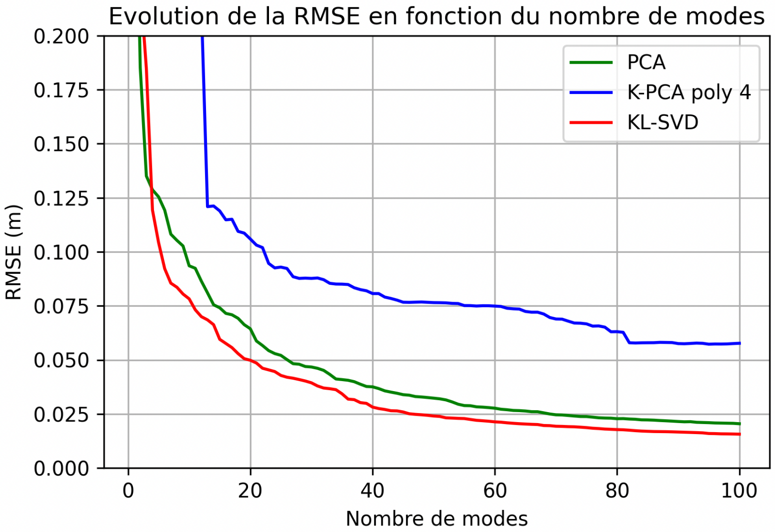
Métriques globales	PCA
MAE [cm]	1
RMSE [cm]	2,66
R^2 [%]	95,93

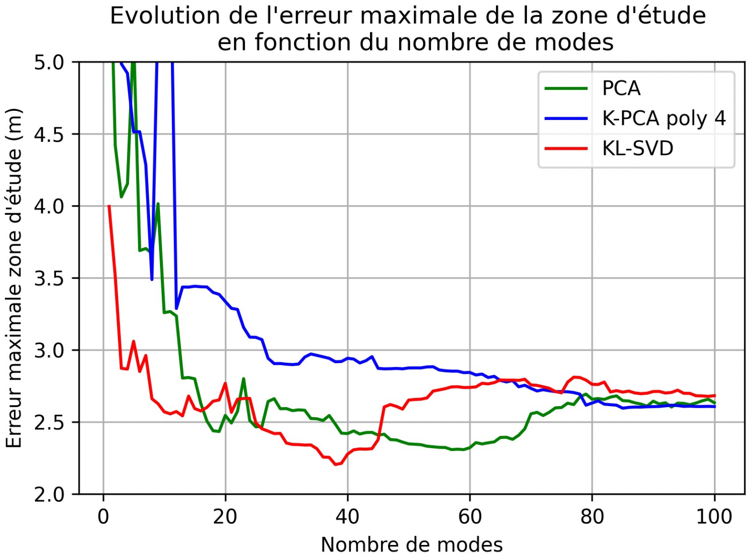
Maximum error PCA = 2.35 m

Noyaux	RMSE [m]	R^2 [%]	MAE [m]	Erreur maximale moyenne globale [m]	Erreur Maximale [m]
Poly-2	0,09	91,71	0,04	0,29	2,66
Poly-3	0,07	94,36	0,03	0,24	2,70
Poly-4	0,07	95,05	0,03	0,24	2,81
Cosine	0,11	86,28	0,04	0,69	2,71
RBF	0,26	64,33	0,99	1,51	3,78

Métriques globales	KL-SVD
MAE [cm]	0,78
RMSE [cm]	2,06
R^2 [%]	94,37

Maximum error KL-SVD = 2.6 m





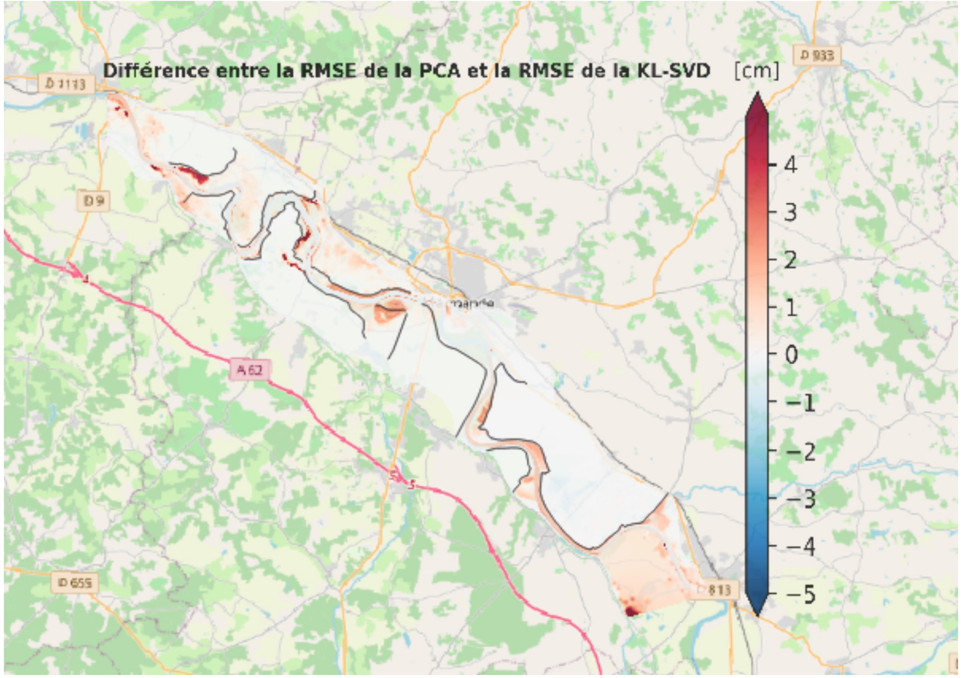
Projet Modélisation ENM-CERFACS (stage court)

- $\mathrm{RMSE}^{KL-SVD} < \mathrm{RMSE}^{PCA} << \mathrm{RMSE}^{KPCA}$
- Erreur maximale $>$ 2,3 m pour les trois méthodes

Surrogate modeling for flood in Hydrodynamics

[I. Raguet (ENM), L. Gibert(ENM), M. Magnan (ENM), A. Mengin (ENM), S. El Garrousi (CERFACS), S. Ricci (CERFACS)]

- Analyse en Composantes Principales (PCA – Scikit Learn)
- Analyse en Composantes Principales à Noyau (KPCA – Scikit Learn) : prise en compte des non linéarités
- Transformation de Karhunen-Loève (KL-SVD - OpenTURNS) : prise en compte des hétérogénéïtés du maillage



PCA
- Temps de calcul courts (10s), simple
- Pas de prise en compte des non linéarités et hétérogénéïtés de la géométrie

KPCA
- Temps de calcul courts (10s)
- Nécessité de tester plusieurs noyaux
- Pas d'amélioration de l'erreur maximale locale

KL-SVD
- Meilleure reconstruction des hauteurs d'eau dans le lit majeur intra-digues où les maillage varie, les gradient de bathymétrie forts et où les vitesses d'écoulement sont fortes
- Pas d'amélioration de l'erreur maximale locale
- Temps de calcul long (6 min)

Projet Modélisation ENM-CERFACS (stage court)

Alternative: Auto-encoder

Surrogate modeling for flood in Hydrodynamics

[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI]

Advanced methodology

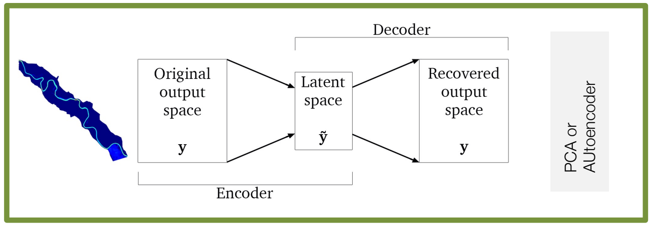
- Mixture of experts
- Dimension reduction for non linear functional outputs

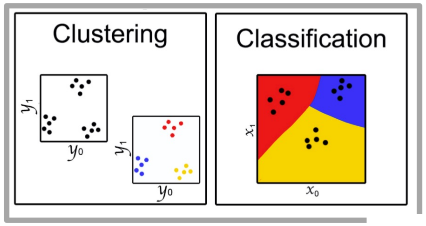
rMPCE strategy

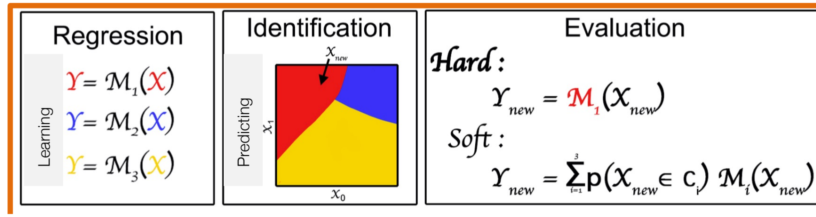
Reduce output dimension taking into account spatial non-linearities with ML

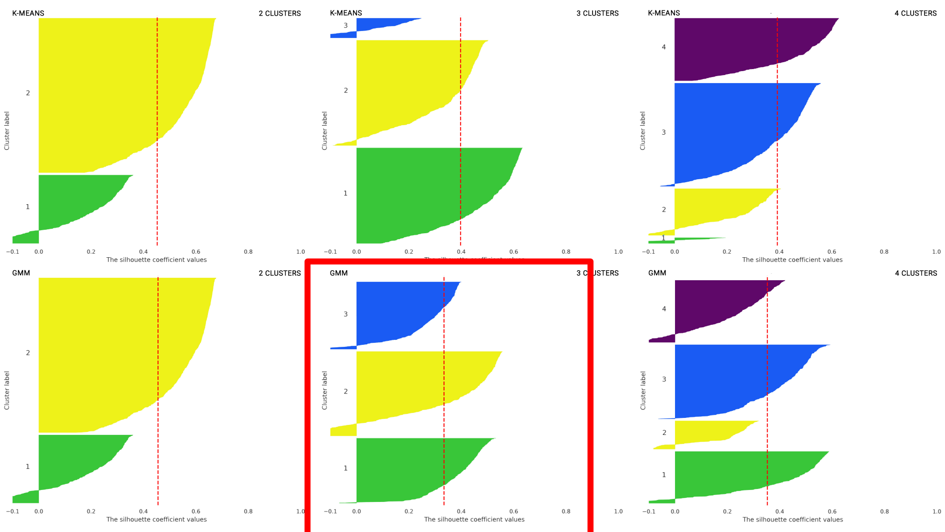
Classify the input/output space with unsupervised and supervised ML

Build a mixture of local polynomial expansion surrogates



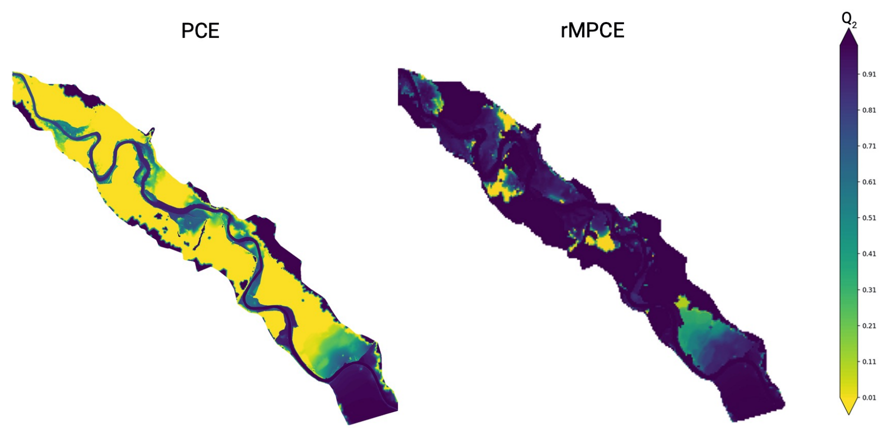






Silhouette plot for K-means (top) and GMM (bottom) with K = 2, 3, 4 (left to right)

Predictive model validation



Improvement of the surrogate model predictive coefficient with rMPCE w.r.t single PCE over 90% of the domain

Surrogate modeling for flood in Hydrodynamics

[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI)

Data: The learning dataset $(\mathbf{X}, \mathbf{Y})$
Result: The rMPCE model

Parameters: The latent space dimension $\tilde{p}$ and the groups' number K

begin
  $\tilde{\mathbf{Y}} \leftarrow$ Reduce the dimension of $\mathbf{Y}$ from p to $\tilde{p}$;
  $\mathcal{L}_1, \ldots, \mathcal{L}_K \leftarrow$ Split $\mathcal{L}$ into K sub-datasets from a clustering on $\tilde{\mathbf{Y}}$;
  $C : \mathbf{x} \mapsto C_1(\mathbf{x}), \ldots, C_K(\mathbf{x}) \leftarrow$ Build a classifier from $\mathbf{X}, \mathcal{L}_1, \ldots, \mathcal{L}_K$;
  for $k \leftarrow 1$ **to** K **do**
    $\widetilde{\mathbf{Y}^{(k)}} \leftarrow$ Reduce the dimension of $\mathbf{Y}^{(k)} = \left(y_j^{(i)}\right)_{\substack{i \in \mathcal{L}_k \\ 1 \leq j \leq p}}$ to $\tilde{p}$;
    $\text{PCE}_k \leftarrow$ Build a PCE from $\mathbf{X}^{(k)} = \left(x_j^{(i)}\right)_{\substack{i \in \mathcal{L}_k \\ 1 \leq j \leq d}}$ and $\widetilde{\mathbf{Y}^{(k)}}$;
  end
  rMPCE $\leftarrow \text{PCE}_1, \ldots, \text{PCE}_K$ and the classifier C;
end

Algorithm 1: Learning stage of the rMPCE

Data: A new input data $\mathbf{x}$ and the rMPCE model.
Result: The predicted output data $\hat{\mathbf{y}}$.

begin
  $C_1(\mathbf{x}), \ldots, C_K(\mathbf{x}) \leftarrow$ Compute the degree of membership to the K sub-groups;
  for $k \leftarrow 1$ **to** K **do**
    $\hat{\tilde{\mathbf{y}}}_k \leftarrow$Compute the k^{th} local prediction with $\text{PCE}_k(\mathbf{x})$;
  end
  $\hat{\tilde{\mathbf{y}}} \leftarrow$ Combine the predictions with $\sum_{k=1}^{K} C_k(\mathbf{x})\ \hat{\tilde{\mathbf{y}}}_k$;
  $\hat{\mathbf{y}} \leftarrow$ Expand the output dimension from $\hat{\tilde{\mathbf{y}}}$ (*decoding*);
end

Algorithm 2: Prediction stage with the rMPCE model

1. Reduction of the output variable dimension,
2. Unsupervised clustering of the learning output data into K groups,
3. Classification of the input space into K subspaces, based on the clustering results,
4. Construction of a 2D-functional output PCE surrogate for each cluster.

then use rMPCE for GSA

Surrogate modeling for flood in Hydrodynamics

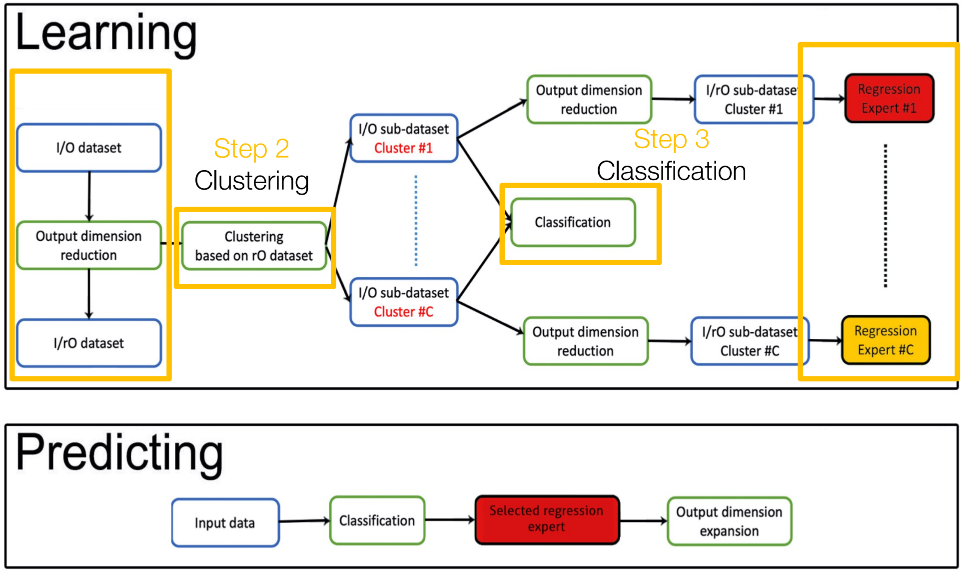
[S. El Garrousi (CERFACS), S. Ricci (CERFACS), M. De Lozzo (IRT), N. Goutal (EDF, LHSV), D. Lucor (LIMSI)

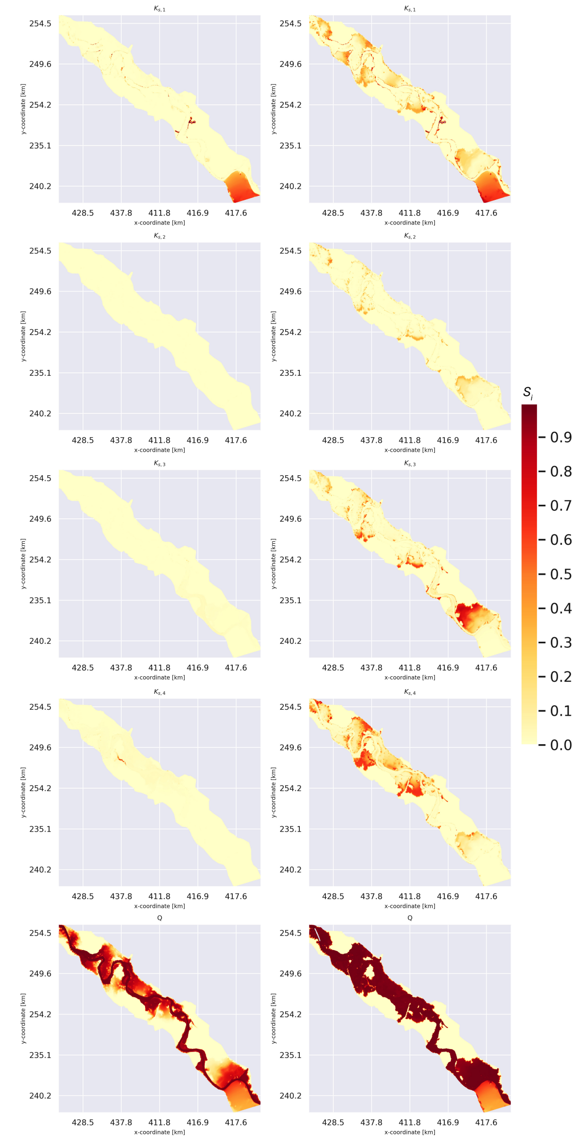
Step 1
Dimension Reduction

Step 2
Clustering

Step 3
Classification

Step 4
Regression





Sobol indices highlight that upstream forcing is the predominant source of uncertainty in water level

Perspectives

- Compute surrogate over time for real flood event by extending input space
- Use surrogate for SA and covariance estimation (low cost EnKF)

El Garroussi, S. et al. (Jan. 2019). "Uncertainty quantification in a two-dimensional river hydraulic model". In: *Uncertainty Quantification in Computational Sciences and Engineering, UNCECOMP*, pp. 243–262. DOI: 10.7712/120219.6339.18380.

El Garroussi, S. et al. (Oct. 2020). "Assessing uncertainties in flood forecasts using a mixture of generalized polynomial chaos expansions". In: *XXVIIth Telemac User Conference*, pp. 84–90.

Galland, J., N. Goutal, and J. Hervouet (1991). "TELEMAC: A new numerical model for solving shallow water equations". In: *Advances in water ressources*.

Li, R. and R. Ghanem (1998). "Adaptive polynomial chaos expansions applied to statistics of extremes in nonlinear random vibration". In: *Probabilistic Engineering Mechanics* 13.2, pp. 125–136. ISSN: 0266-8920. DOI: https://doi.org/10.1016/S0266-8920(97)00020-9.

Marelli, Stefano et al. (2021). "Stochastic spectral embedding". In: *International Journal for Uncertainty Quantification* 11.2, pp. 25–47. ISSN: 2152-5080.

Najm, H.N. (2009). "Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics". In: *Annual Review of Fluid Mechanics* 41.1, pp. 35–52. DOI: 10.1146/annurev.fluid.010908.165248.

Wiener, N. (1938). "The Homogeneous Chaos". In: *American Journal of Mathematics* 60.4, pp. 897–936.

El Garroussi et al (2021), "Tackling random fields non-linearities with unsupervised clustering of polynomial chaos expansion in latent: space: application to global sensitivity analysis of river flooding", Stochastic Environmental Research and Risk Assessment journal

Thank you for your attention