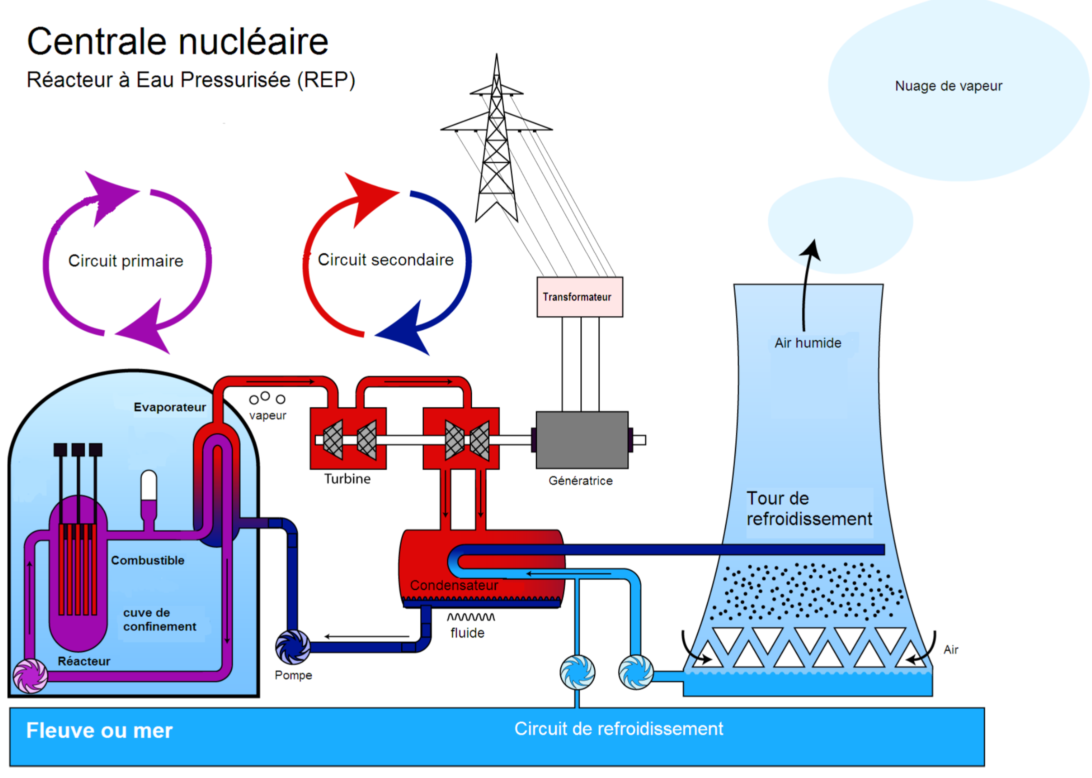
Computer code
Functional medamodeling

Benjamin Auder

CEA Cadarache / UPMC Paris 6

INTRODUCTION

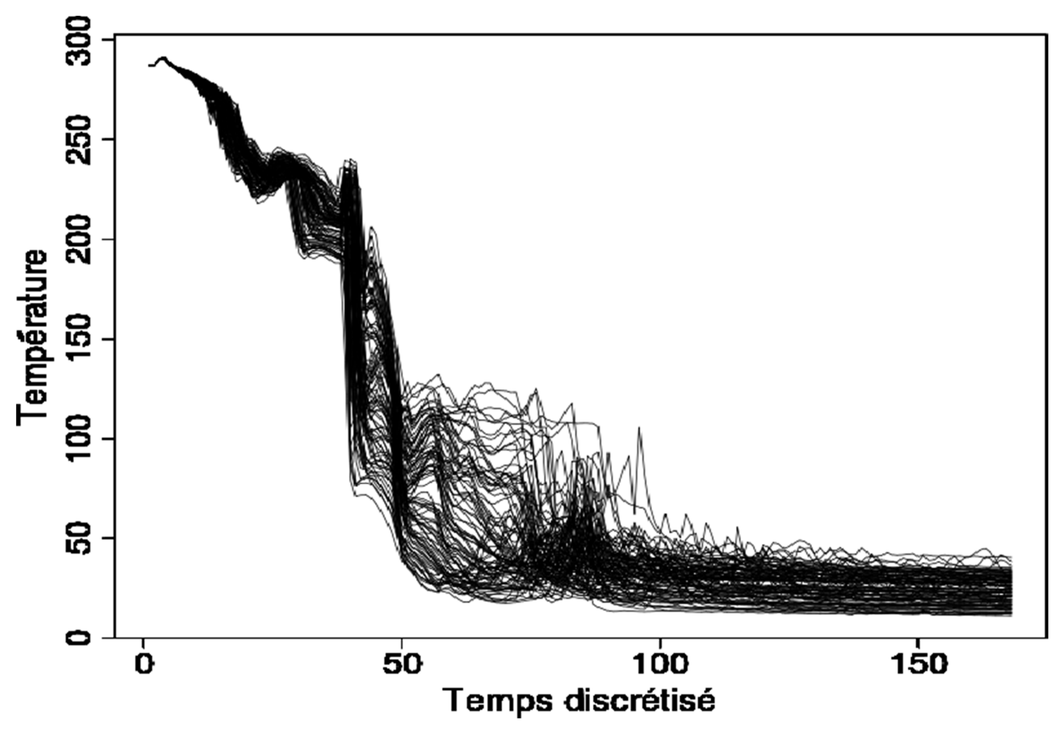


Industrial context = project about life-span of nuclear reactors.

Goal = maximize durability of nuclear reactors at minimal risks.

Sub-project = evaluate the impact of possible accidents on the nuclear vessel
➔ Ensure that some thresholds are not reached.

Chosen approach = study a thermal-hydaulic computer code, and deduce "real world" conclusions from his behaviour.



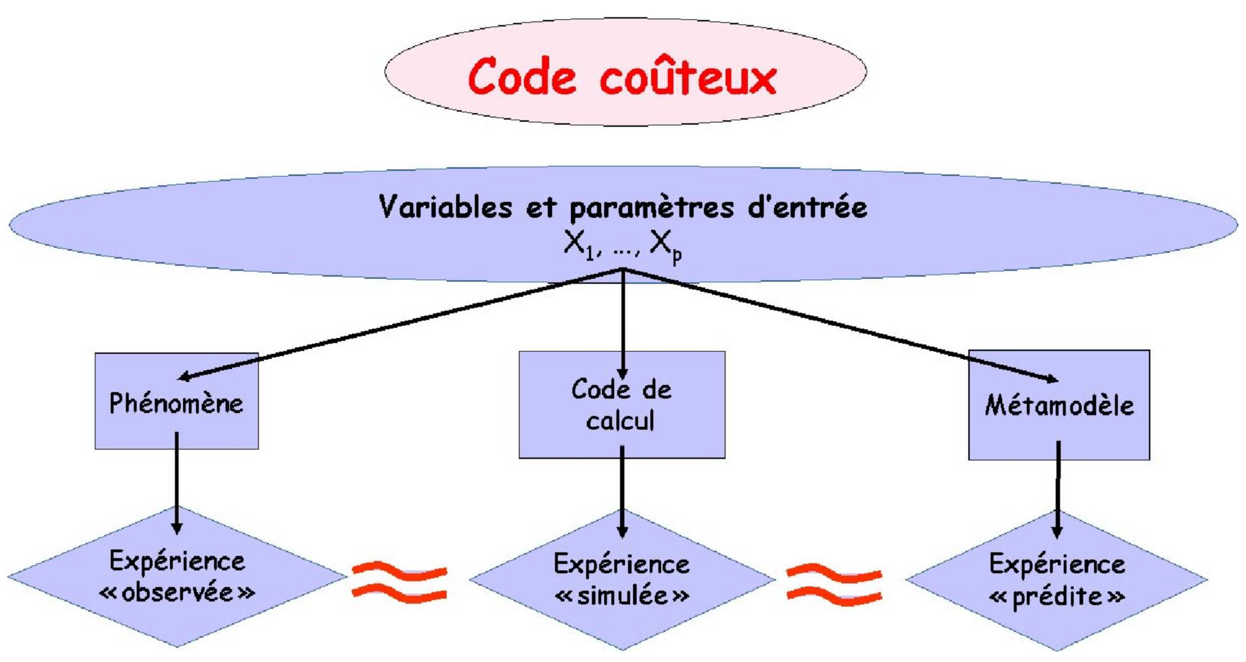
Temperature curves on output of CATHARE code

OBJECTIVES (with words)

Run **statistical analysis** on CATHARE outputs
➔ Give the probabilities that some scenarii lead to accidents.

These probabilities are directly related to the risk thresholds.

Statistical analysis need a lot of code results.
Problem: code is very slow - several hours for one run.
➔ Solution = approximate the code, and use this "metamodel" for analysis.



Summary of the situation

OBJECTIVES (with maths)

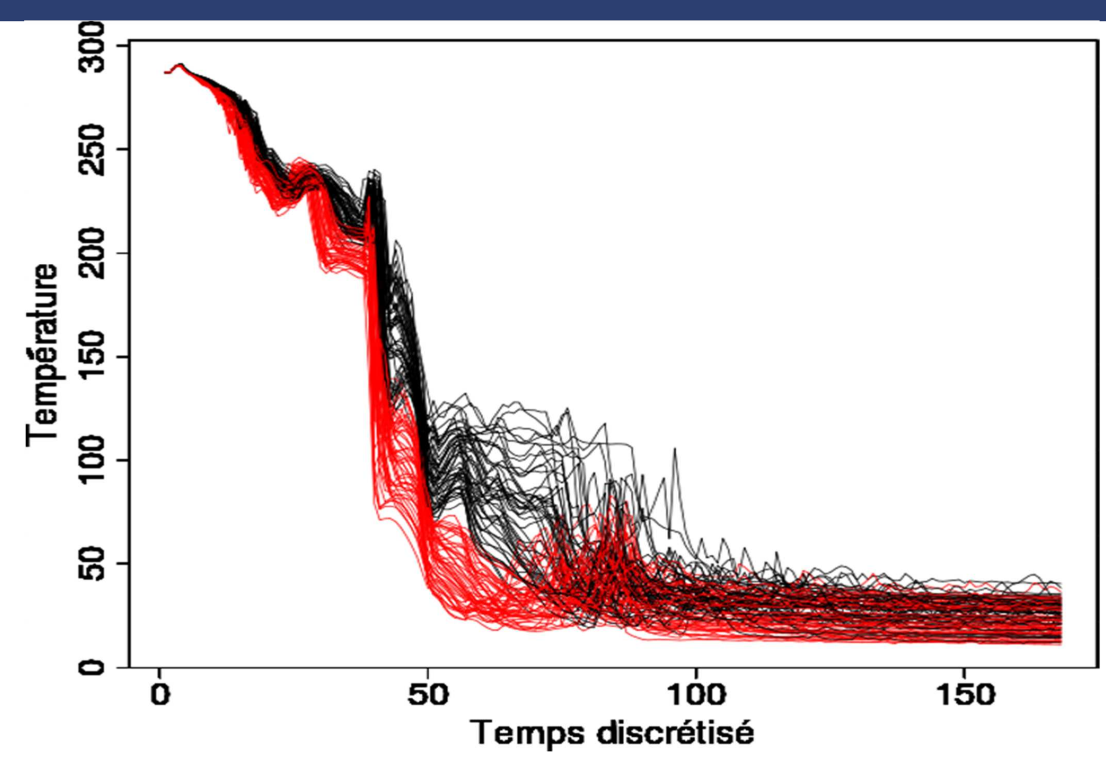
Available = n inputs-outputs (n ≈ a few hundreds) (x_i, y_i) where
- x_i is a **vector** of $\mathbf{R}^p$
- y_i is a **function** from [a,b] to **R**

What we want from (x_i, y_i) = a "metamodel" Φ, such that

$$\Phi(x) \approx \text{CATHARE}(x)$$

for a new input x

METHOD

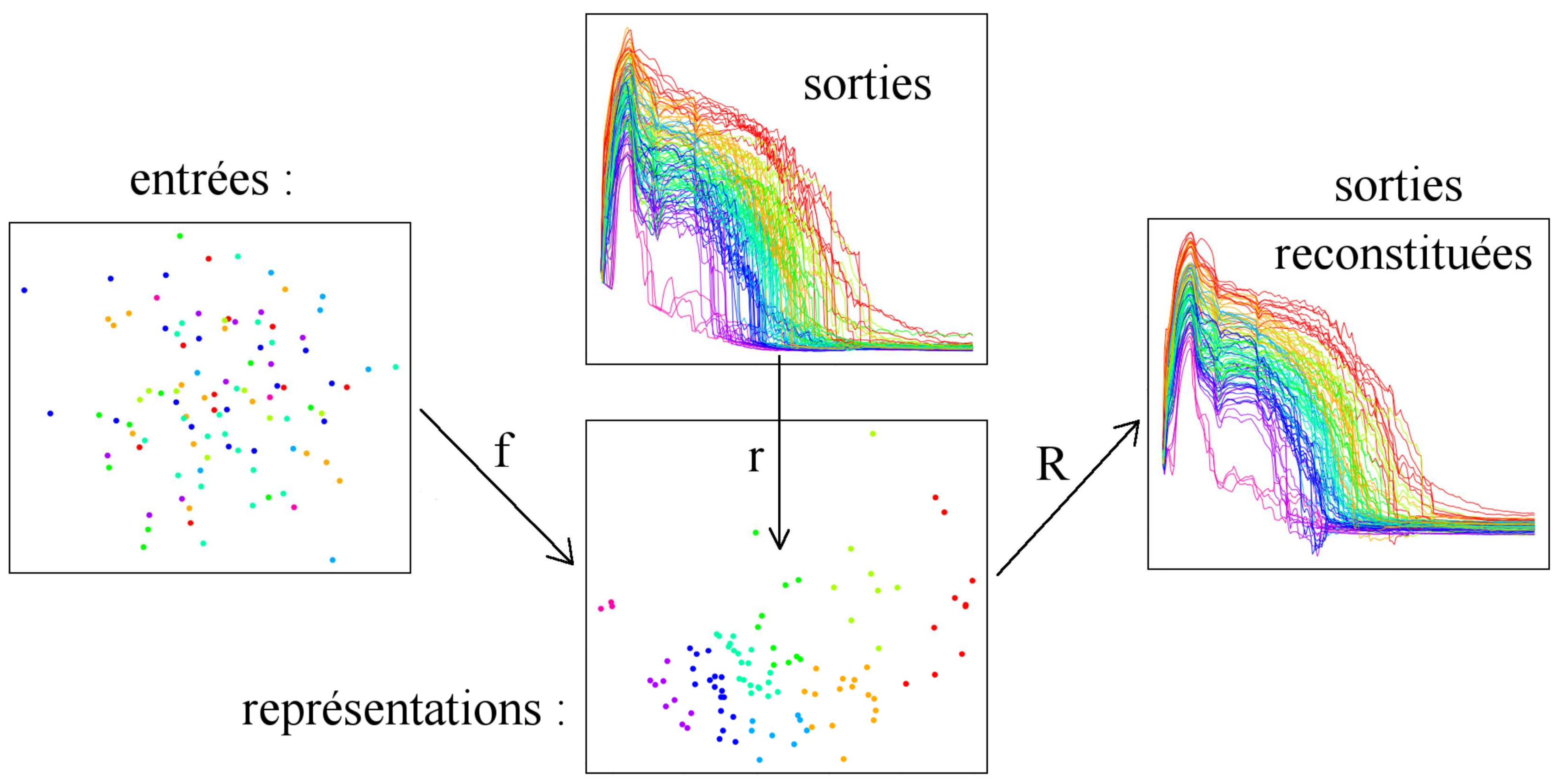


Visually, several (two) groups of curves show different patterns
➔ **First step** = cluster data into meaningful group of curves

---

Method:
1. Represent data through a graph, and compute the random walk euclidian commute-time distance (ECTD).
2. Apply hierarchical, clustering based on the matrix of ECTD distances.

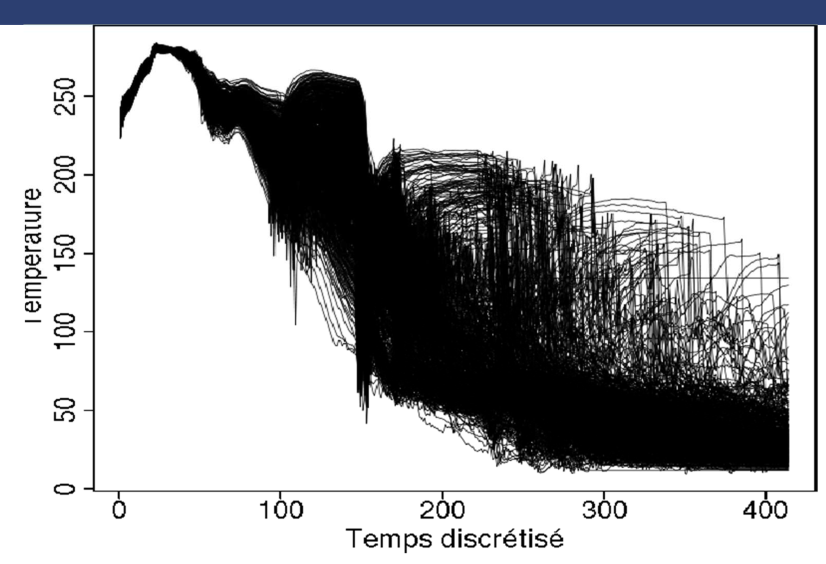
Within each cluster:
1. Reduce the dimension ➔ fall back to usual $\mathbf{R}^p \rightarrow \mathbf{R}^d$ regression. [*]
2. Apply the Projection Pursuit regression to each component
3. Map back the reduced prediction from $\mathbf{R}^d$ to functional space.



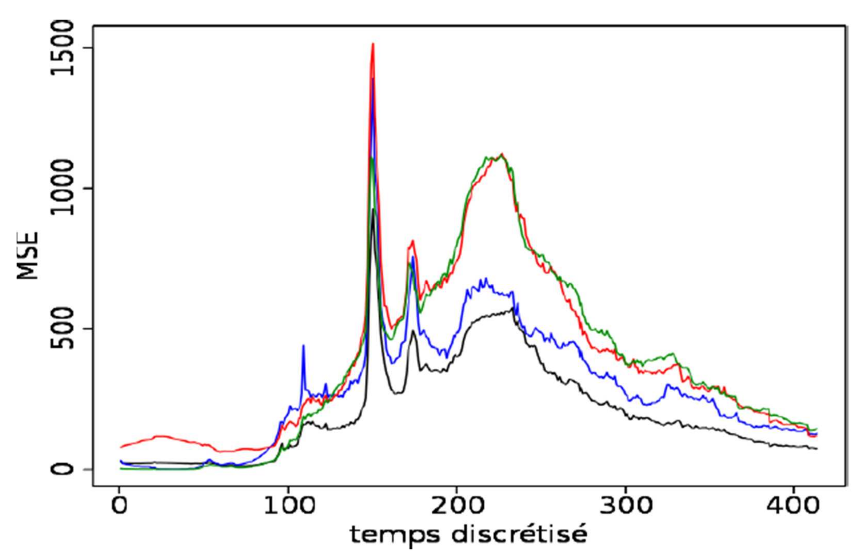
[*] : method used = (functional) Principal Component Analysis, FPCA
and Riemannian Manifold Learning [RML; non-linear modeling]

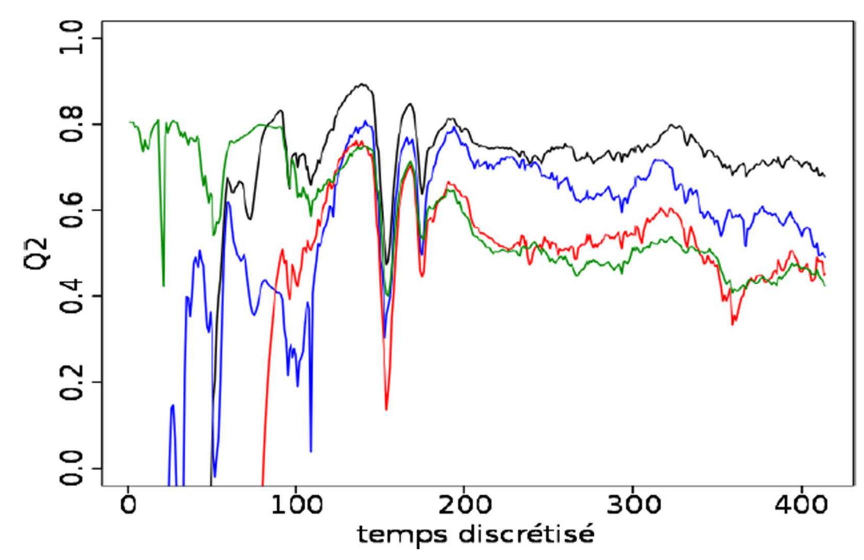
RESULTS

600 model runs / curves
11 input dimensions [vector]
High dimension: 414 sampling points
→ Reduced to d = 8



Leave-30-out: MSE at l., Q2 at r.
Black = FPCA - blue = RML
green= k-NN regression - red = local FPCA





CONCLUSIONS

Linear, non-linear, w/wo clustering approaches ≈ complementary
Main achievement = **R package modelcf** [available on CRAN]

Possible continuations:
- Taking inter-curves correlations into account.
- Develop "dedicated sensitivity analysis" using this metamodel ...