

# Apprentissage d'une variété pour la construction d'un métamodèle fonctionnel

Benjamin Auder

CEA - UPMC

25 novembre 2010

Directeur de thèse : Gérard Biau (UPMC)

Encadrant EDF : Bertrand Iooss

Encadrant CEA : Michel Marquès

- 1 Hypothèses et motivations
- 2 Estimation de la dimension  $d$
- 3 Riemannian Manifold Learning (Lin et al. 2006)
- 4 Applications

# Contexte industriel

Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagées.

*But* : estimer leurs probabilités.

# Contexte industriel

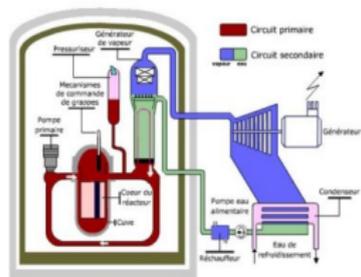
Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagés.

*But* : estimer leurs probabilités.

## Méthodologie

### Modélisation



# Contexte industriel

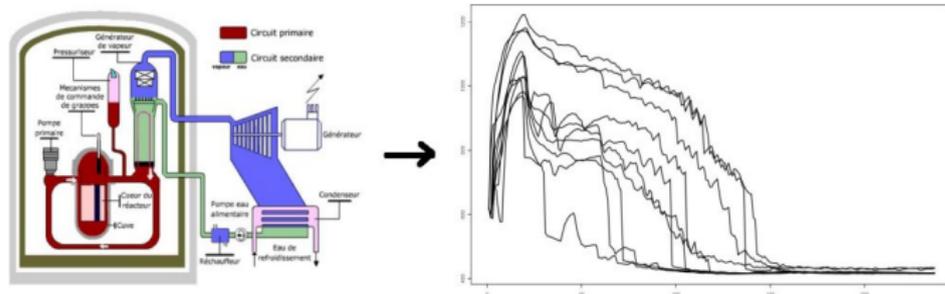
Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagées.

*But* : estimer leurs probabilités.

## Méthodologie

Modélisation → Simulation



# Contexte industriel

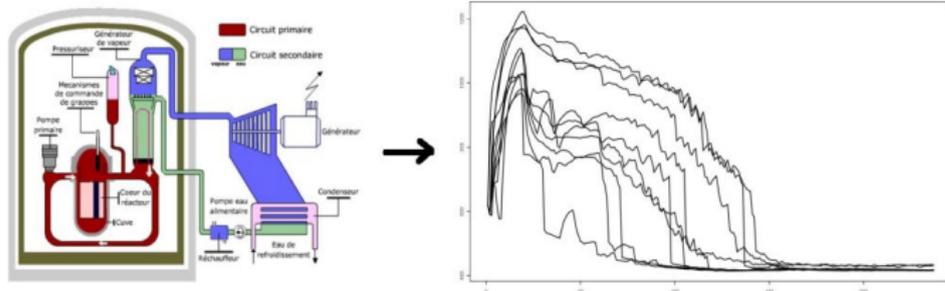
Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagées.

*But* : estimer leurs probabilités.

## Méthodologie

Modélisation → Simulation → Calculs.



→ Analyse de sensibilité, propagation d'incertitudes ..etc.

# Contexte industriel

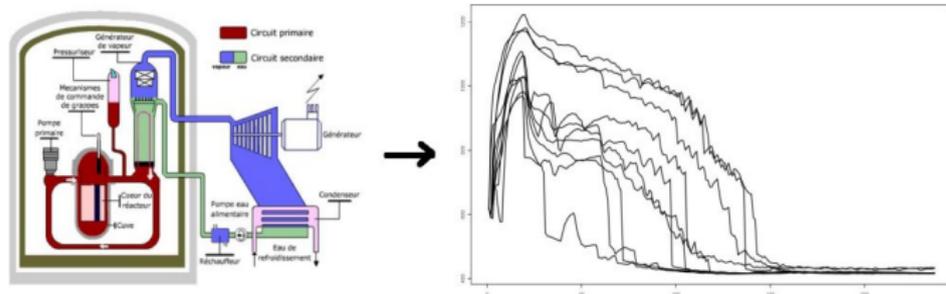
Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagées.

*But* : estimer leurs probabilités.

## Méthodologie

Modélisation → **Simulation** → Calculs.



→ Analyse de sensibilité, propagation d'incertitudes ..etc.

*Améliorer la phase simulation, pour effectuer des calculs plus fiables*

# Au CEA - DER/SESI/LSMR ...

Code thermo-hydraulique "boîte noire" CATHARE coûteux en temps.

Entrées  $x_i$  → CATHARE → CASTEM → probabilité de défaillance

  
code thermo-mécanique

Code thermo-hydraulique "boîte noire" CATHARE coûteux en temps.

Entrées  $x_i \rightarrow$  CATHARE  $\rightarrow$  CASTEM  $\rightarrow$  probabilité de défaillance

  
code thermo-mécanique

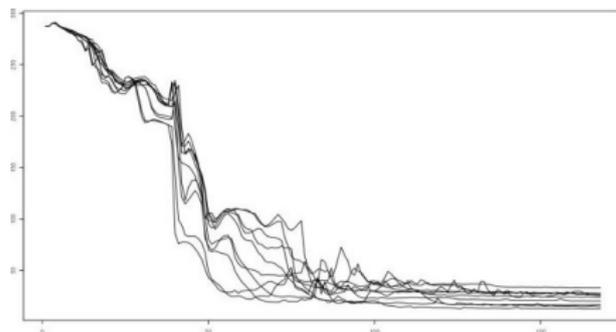


FIGURE: Transitoires de température.

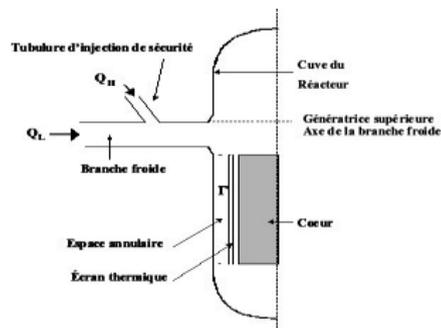


FIGURE: Zone modélisée

Code thermo-hydraulique "boîte noire" CATHARE coûteux en temps.

Entrées  $x_i \rightarrow$  CATHARE  $\rightarrow$  CASTEM  $\rightarrow$  probabilité de défaillance

  
code thermo-mécanique

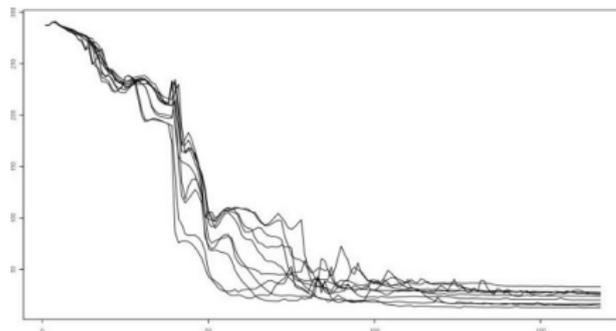


FIGURE: Transitoires de température.

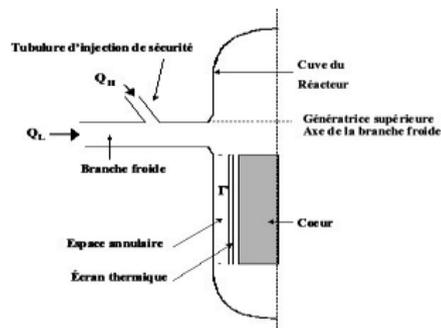
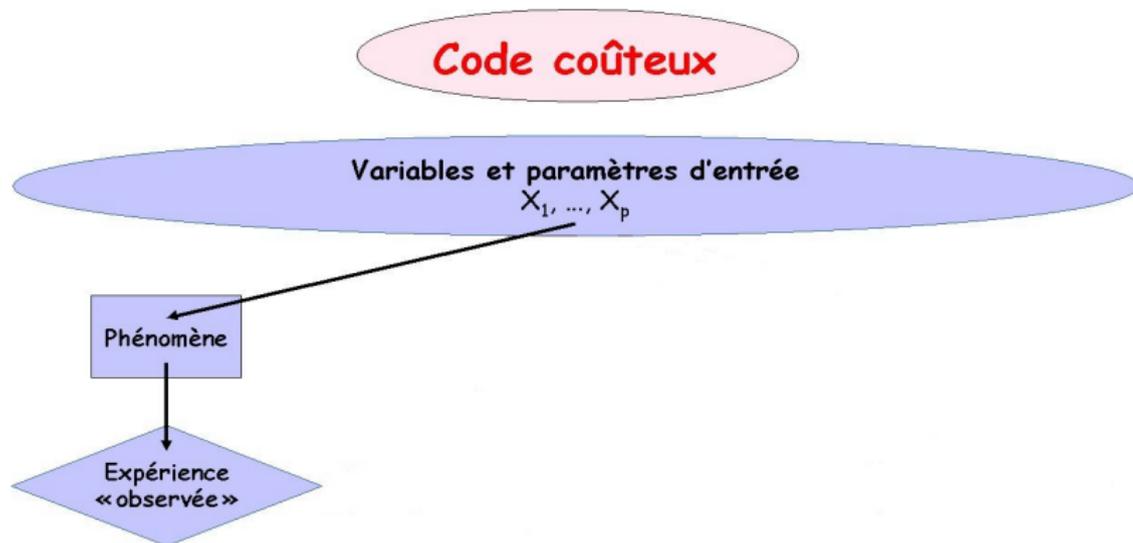


FIGURE: Zone modélisée

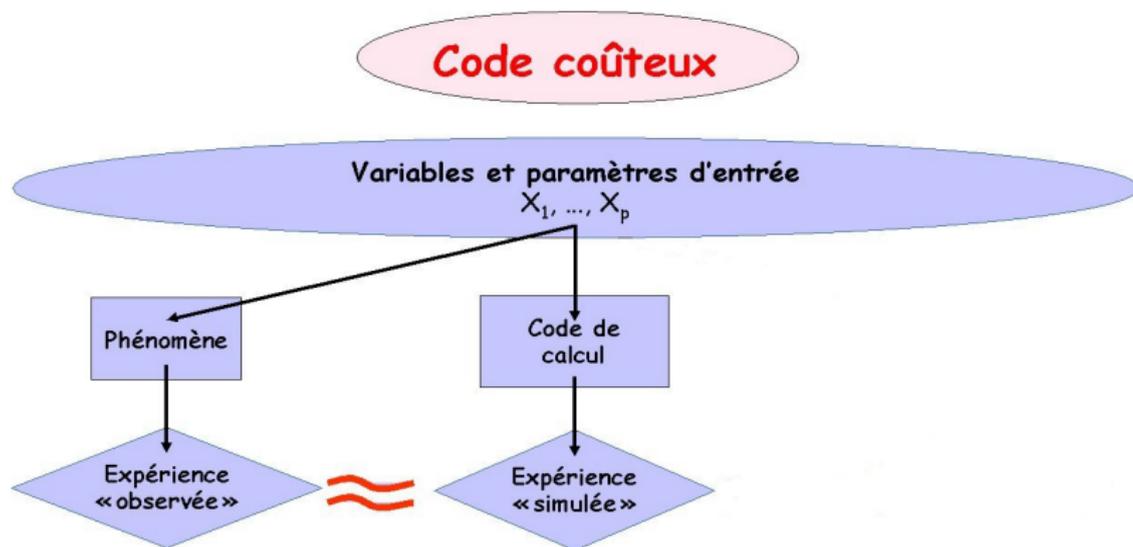
"Accélérer" l'exécution du code CATHARE

# Solution ?



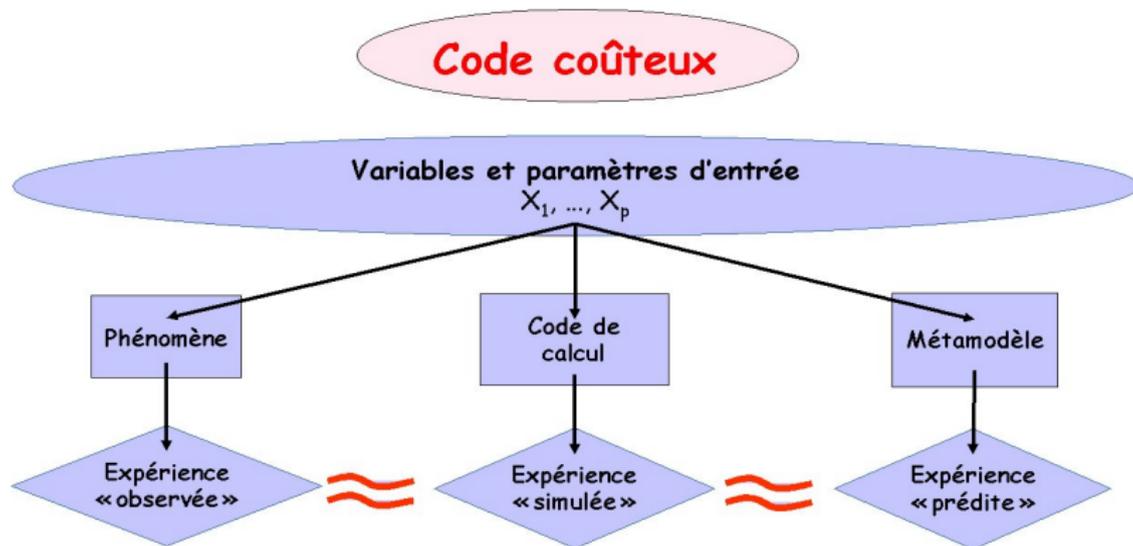
analyses statistiques sur le phénomène

# Solution ?



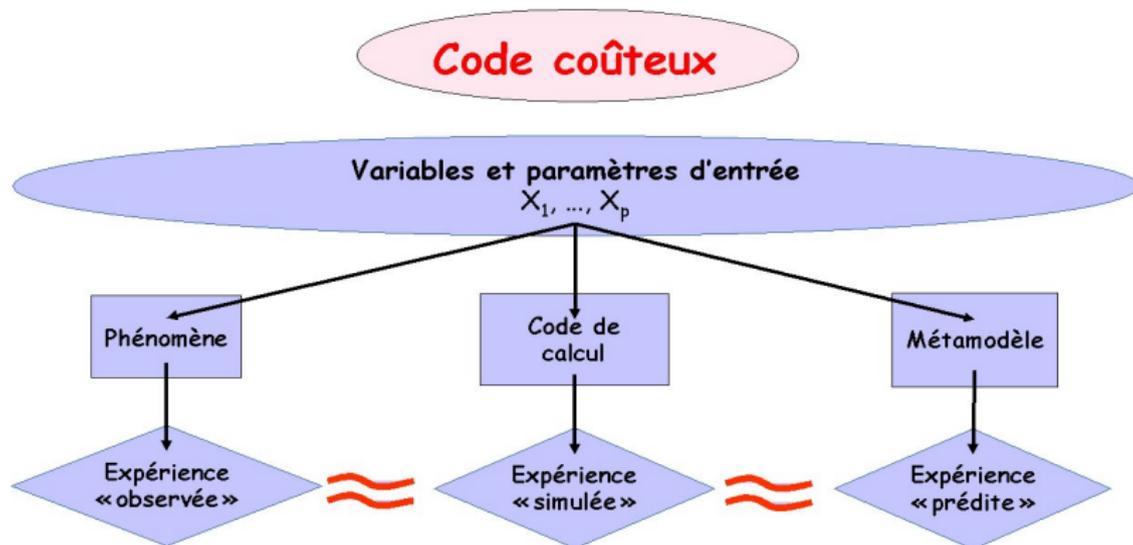
*nécessitent* analyses statistiques sur le phénomène  
beaucoup de résultats de code

# Solution : métamodèle



*nécessitent* analyses statistiques sur le phénomène  
*obtenus avec* beaucoup de résultats de code  
un métamodèle (= modèle du code).

# Solution : métamodèle



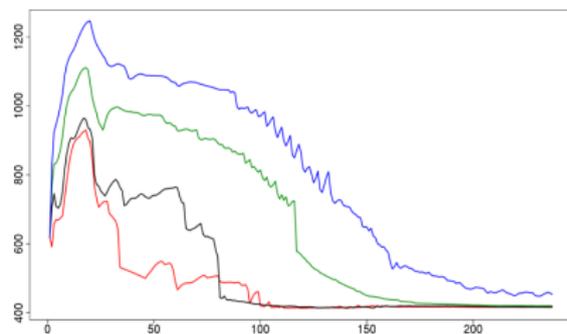
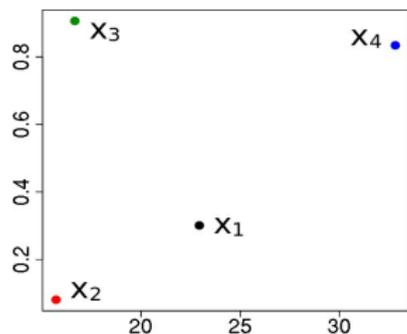
*nécessitent* analyses statistiques sur le phénomène  
*obtenus avec* beaucoup de résultats de code  
un métamodèle (= modèle du code).

*Construire un modèle du code CATHARE*

# Reformulation finale

$n$  couples  $(x_i, y_i)$  connus :

- entrées  $x_i \in \mathbb{R}^p =$  état initial du système physique ;
- sorties  $y_i \in \mathcal{C}([a, b], \mathbb{R}) =$  évolution des paramètres.



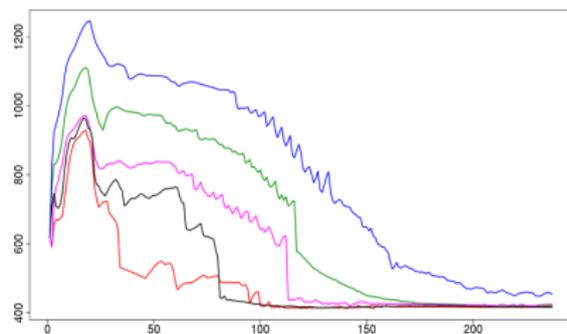
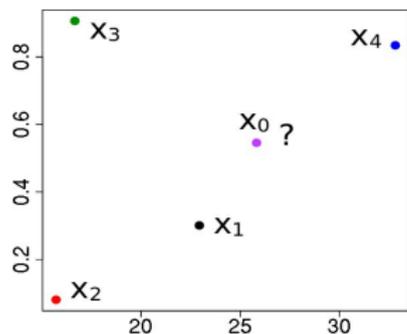
# Reformulation finale

$n$  couples  $(x_i, y_i)$  connus :

- entrées  $x_i \in \mathbb{R}^p =$  état initial du système physique ;
- sorties  $y_i \in \mathcal{C}([a, b], \mathbb{R}) =$  évolution des paramètres.

Objectif = **prédiction** de données fonctionnelles :

$$y^{\text{new}} \simeq \varphi(x^{\text{new}}).$$



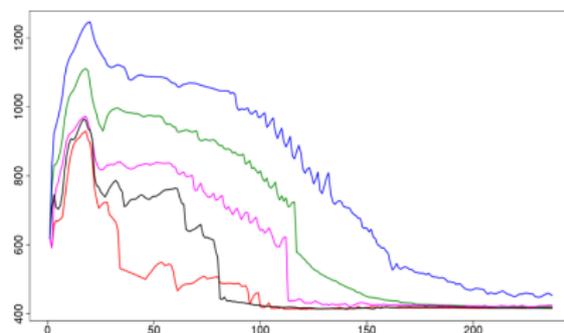
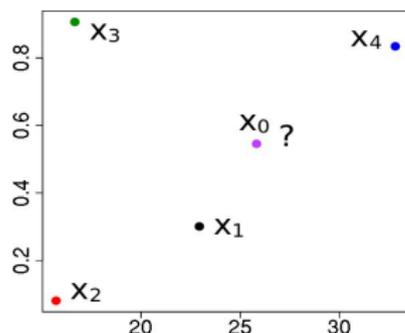
# Reformulation finale

$n$  couples  $(x_i, y_i)$  connus :

- entrées  $x_i \in \mathbb{R}^p =$  état initial du système physique ;
- sorties  $y_i \in \mathcal{C}([a, b], \mathbb{R}) =$  évolution des paramètres.

Objectif = **prédiction** de données fonctionnelles :

$$y^{\text{new}} \simeq \varphi(x^{\text{new}}).$$



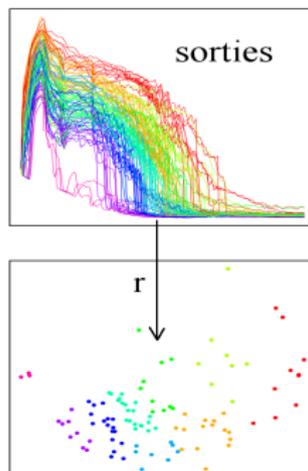
*Apprentissage statistique "régression"  $\mathbb{R}^p \rightarrow \mathcal{C}([a, b], \mathbb{R})$*

Comment se ramener au cas "simple"  $y_i \in \mathbb{R}^d$  ?

# Comment se ramener au cas "simple" $y_i \in \mathbb{R}^d$ ?

① réduction de la dimension :

$$r : \mathcal{C}([a, b], \mathbb{R}) \rightarrow \mathbb{R}^d \text{ (représentation) ;}$$



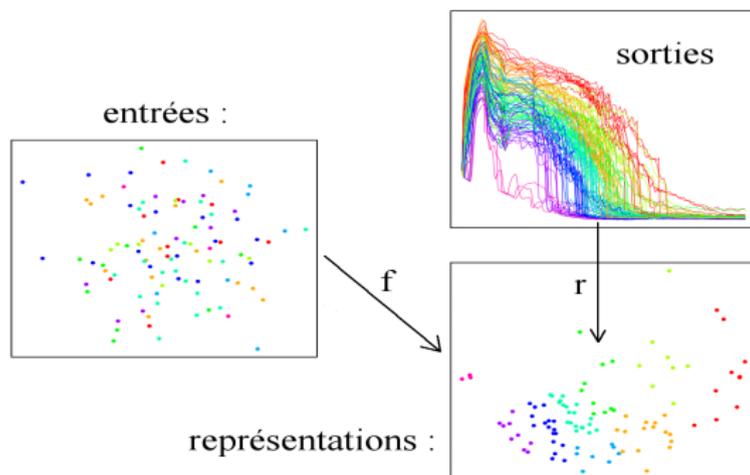
# Comment se ramener au cas "simple" $y_i \in \mathbb{R}^d$ ?

- 1 réduction de la dimension :

$$r : \mathcal{C}([a, b], \mathbb{R}) \rightarrow \mathbb{R}^d \text{ (représentation) ;}$$

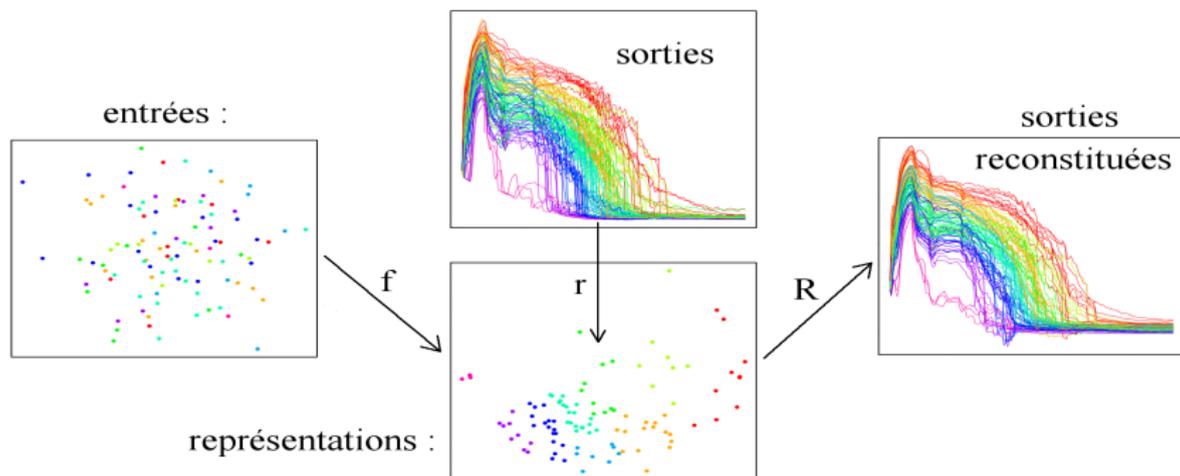
- 2 apprentissage statistique classique :

$$f : \mathbb{R}^p \text{ (entrées) } \rightarrow \mathbb{R}^d \text{ (sorties réduites) ;}$$



# Comment se ramener au cas "simple" $y_i \in \mathbb{R}^d$ ?

- 1 réduction de la dimension :  
 $r : \mathcal{C}([a, b], \mathbb{R}) \rightarrow \mathbb{R}^d$  (représentation) ;
- 2 apprentissage statistique classique :  
 $f : \mathbb{R}^p$  (entrées)  $\rightarrow \mathbb{R}^d$  (sorties réduites) ;
- 3 paramétrage de l'espace des sorties :  
 $R : \mathbb{R}^d \rightarrow \mathcal{C}([a, b], \mathbb{R})$  (reconstruction).



## Méthodes "classiques"

- Régression linéaire fonctionnelle :  
Faraway, 1997 ; Ramsay & Silverman, 2005, ...

## Méthodes "classiques"

- Régression linéaire fonctionnelle :  
Faraway, 1997 ; Ramsay & Silverman, 2005, ...
- Décomposition sur une base orthonormée puis apprentissage des coefficients  $d$ -dimensionnels :  
Chiou et al., 2004 ; Govaerts & Noël, 2005 ;  
Bayarri et al., 2007 ; Marrel, 2008 ; Monestiez & Nerini, 2009

## Méthodes "classiques"

- Régression linéaire fonctionnelle :  
Faraway, 1997 ; Ramsay & Silverman, 2005, ...
- Décomposition sur une base orthonormée puis apprentissage des coefficients  $d$ -dimensionnels :  
Chiou et al., 2004 ; Govaerts & Noël, 2005 ;  
Bayarri et al., 2007 ; Marrel, 2008 ; Monestiez & Nerini, 2009

But : minimiser la dimension de représentation  $d$ , pour

- simplifier le modèle ;
- éviter le surapprentissage,

en conservant de bonnes performances.

- 1 Hypothèses et motivations
- 2 Estimation de la dimension  $d$
- 3 Riemannian Manifold Learning (Lin et al. 2006)
- 4 Applications

- 1 Hypothèses et motivations
- 2 Estimation de la dimension  $d$
- 3 Riemannian Manifold Learning (Lin et al. 2006)
- 4 Applications

## Espace topologique

Ensemble  $E$  + famille  $\mathcal{O}$  d'ouverts  $O \subset E$ ,  
stable par intersection finie et union quelconque.

Exemple :  $\mathbb{R}^n$  + topologie engendrée par les boules ouvertes de  $\|\cdot\|_2$ .

## Espace topologique

Ensemble  $E$  + famille  $\mathcal{O}$  d'ouverts  $O \subset E$ ,  
stable par intersection finie et union quelconque.

Exemple :  $\mathbb{R}^n$  + topologie engendrée par les boules ouvertes de  $\|\cdot\|_2$ .

## Homéomorphisme = "déformation continue"

Application  $f$  bijective continue ( $f^{-1}(O_F) \in \mathcal{O}_E$ ) entre deux espaces topologiques  $E$  et  $F$ , la réciproque étant continue également.

Exemple : projection de la demi-sphère  $3D$  sur le disque  $2D$ .

# Variété connexe

## Variété (connexe)

Espace topologique connexe séparé, à base dénombrable d'ouverts  
**localement homéomorphe à  $\mathbb{R}^d$ .**

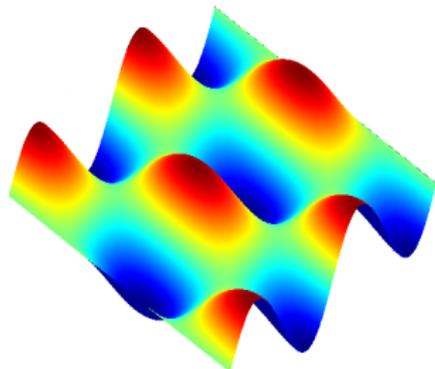


FIGURE: Visualisation d'une variété

# Variété connexe

## Variété (connexe)

Espace topologique connexe séparé, à base dénombrable d'ouverts  
**localement homéomorphe à  $\mathbb{R}^d$ .**

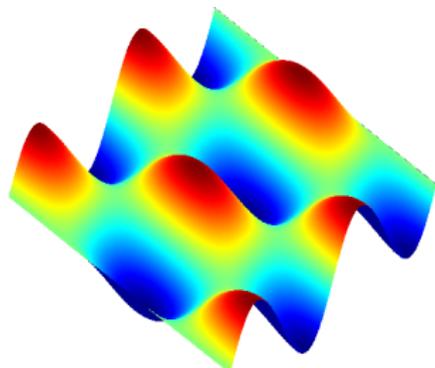


FIGURE: Visualisation d'une variété

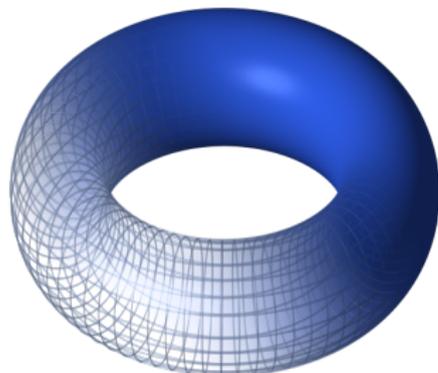


FIGURE: Tore 2D dans  $\mathbb{R}^3$

Tore centré de rayons 1 :  $(\rho - 2)^2 + z^2 = 1$

Paramétrisation :  $(t, \varphi) \mapsto (2 + \cos t, \varphi, \sin t)$

# Estimation de $d$

On se place dans  $E = \mathcal{C}([a, b])$ ,  $\mathbb{R}$ -ev. de dimension infinie.

Données (discrétisées) :  $y_1, \dots, y_n \in V \subset E$ ; structure de  $V$  ?

# Estimation de $d$

On se place dans  $E = \mathcal{C}([a, b])$ ,  $\mathbb{R}$ -ev. de dimension infinie.

Données (discrétisées) :  $y_1, \dots, y_n \in V \subset E$ ; structure de  $V$  ?

$H_1$  :  $V =$  espace vectoriel

- 1 recherche d'une famille libre maximale  $\mathcal{L}$  dans  $V$  ;
- 2 orthonormalisation de  $\mathcal{L}$  (via SVD p.ex.) en  $\mathcal{L}'$ .

$\mathcal{L}'$  bon. de  $V$ , de dimension  $d = |\mathcal{L}'|$ .

# Estimation de $d$

On se place dans  $E = \mathcal{C}([a, b])$ ,  $\mathbb{R}$ -ev. de dimension infinie.

Données (discrétisées) :  $y_1, \dots, y_n \in V \subset E$ ; structure de  $V$  ?

$H_1 : V = \text{espace vectoriel}$

- 1 recherche d'une famille libre maximale  $\mathcal{L}$  dans  $V$  ;
- 2 orthonormalisation de  $\mathcal{L}$  (via SVD p.ex.) en  $\mathcal{L}'$ .

$\mathcal{L}'$  bon. de  $V$ , de dimension  $d = |\mathcal{L}'|$ .

$H_2 : V = \text{variété}$

Le plan tangent en  $f_0 \in V$  est isomorphe à  $\mathbb{R}^d$ .

Approximation :  $\{f_0 + k \text{ plus proches voisins}\} \simeq \text{sev.}$

$\Rightarrow$  méthode précédente localement, puis moyenne.

# Hypothèses comparées

## Espace vectoriel

- + représentation / reconstruction immédiate via une bon.
- + dimension  $d$  facile à estimer (orthonormalisation).

# Hypothèses comparées

## Espace vectoriel

- + représentation / reconstruction immédiate via une bon.
- + dimension  $d$  facile à estimer (orthonormalisation).
- $d$  peut être (beaucoup) trop grand si hypothèse fausse.

# Hypothèses comparées

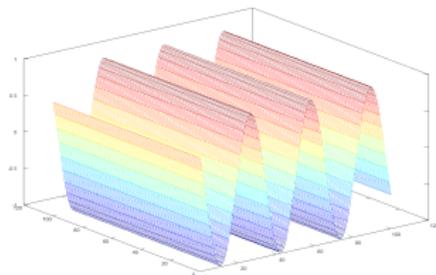
## Espace vectoriel

- + représentation / reconstruction immédiate via une bon.
- + dimension  $d$  facile à estimer (orthonormalisation).
- $d$  peut être (beaucoup) trop grand si hypothèse fausse.

## Variété = *choix retenu*

- + permet d'optimiser le nombre de paramètres à apprendre.
- + paramétrisation "naturelle"  $\Rightarrow$  régression facilitée. [\*]

[\*] Surface  $2D$  à forte courbure :



# Hypothèses comparées

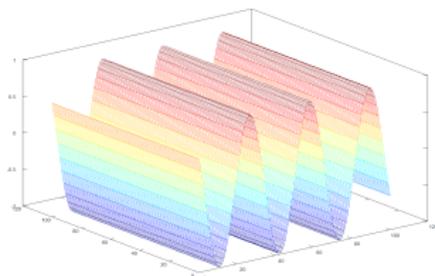
## Espace vectoriel

- + représentation / reconstruction immédiate via une bon.
- + dimension  $d$  facile à estimer (orthonormalisation).
- $d$  peut être (beaucoup) trop grand si hypothèse fausse.

## Variété = *choix retenu*

- + permet d'optimiser le nombre de paramètres à apprendre.
- + paramétrisation "naturelle"  $\Rightarrow$  régression facilitée. [\*]
- (nettement plus) difficile de trouver  $d$  et paramétrisation.

[\*] Surface  $2D$  à forte courbure :



- 1 Hypothèses et motivations
- 2 Estimation de la dimension  $d$
- 3 Riemannian Manifold Learning (Lin et al. 2006)
- 4 Applications

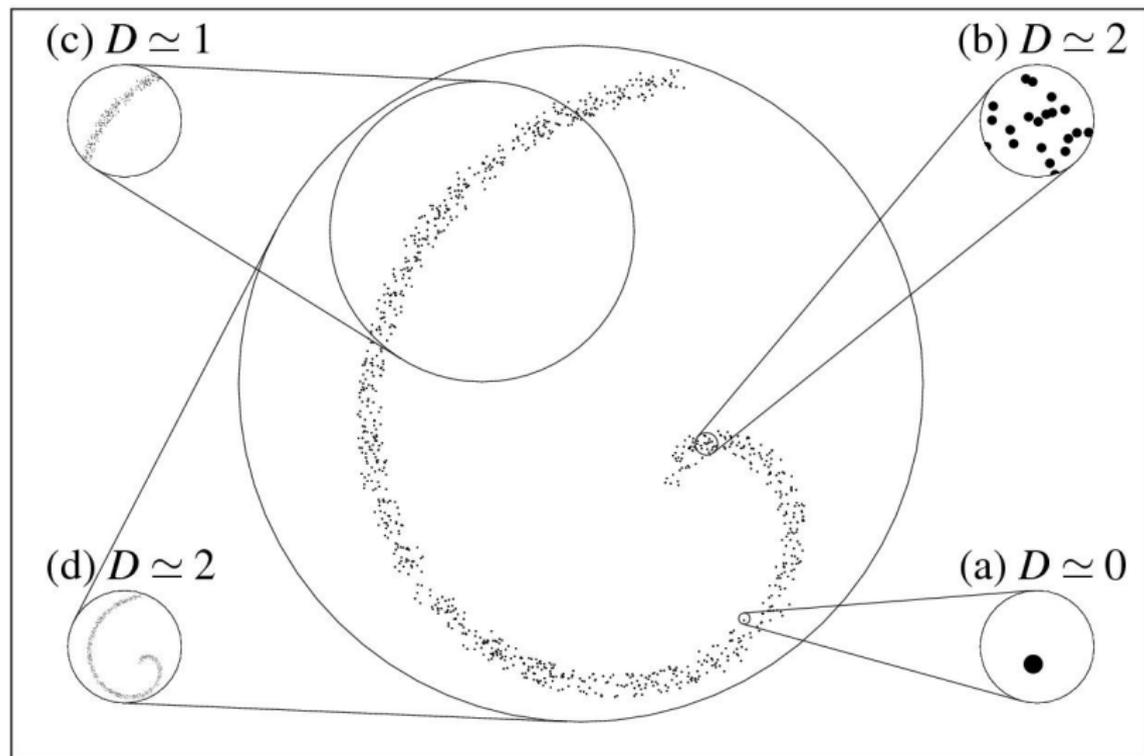


FIGURE: Spirale bruitée à différentes échelles

*Manifold-adaptive dimension estimation*, A. M. Farahmand et al. [07]

## Idée

Pour une distribution uniforme en dimension  $d$ , la probabilité de trouver un élément dans une boule de rayon  $r$  est  $\propto r^d$ .

Supposant la variété régulière et uniformément échantillonnée :

$\mathbb{P}(Y \in B(y_i, r)) = \eta(y_i, r)r^d$ , avec  $\eta(y_i, \cdot) = \eta_0^{(i)}$  au voisinage de 0.

*Manifold-adaptive dimension estimation*, A. M. Farahmand et al. [07]

## Idée

Pour une distribution uniforme en dimension  $d$ , la probabilité de trouver un élément dans une boule de rayon  $r$  est  $\propto r^d$ .

Supposant la variété régulière et uniformément échantillonnée :

$\mathbb{P}(Y \in B(y_i, r)) = \eta(y_i, r)r^d$ , avec  $\eta(y_i, \cdot) = \eta_0^{(i)}$  au voisinage de 0.

Si  $y_i$  a  $k$  voisins "semblables", pour  $N$  "grand" :

$\mathbb{P}(Y \in B(y_i, r_k^{(i)})) \simeq \frac{k'}{N}$ ,

avec  $r_k^{(i)}$  = distance au  $k^{\text{eme}}$  voisin,  $1 \leq k' \leq k$ .

# Formule pour $d$

$$\ln \frac{k}{N} \simeq \ln \eta_0^{(i)} + d \ln r_k^{(i)}, \text{ et}$$

$$\ln \frac{k}{2N} \simeq \ln \eta_0^{(i)} + d \ln r_{[k/2]}^{(i)}.$$

$$\hat{d}_i = \frac{\ln 2}{\ln r_k^{(i)} / r_{[k/2]}^{(i)}}$$

= Estimation locale de la dimension *non entière*.

Deux options :

- $\hat{d} = \frac{1}{N} \sum_{i=1}^N \hat{d}_i$  ;
- $\hat{d} = \arg \max_{d' \in \mathbb{N}^*} \sum_{i=1}^N \mathbb{1}_{\hat{d}_i = d'}$ .

# Convergence théorique

Avec probabilité au moins  $1 - \delta$  :

$$|\hat{d} - d| \leq Cd \left( B \left( \frac{k}{n} \right)^{\frac{1}{d}} + \sqrt{\frac{\log(4/\delta)}{k}} \right),$$

$$C \simeq \left( \min_{i=1..n} \eta_0^{(i)} \right)^{-\frac{1}{d}}.$$

# Convergence théorique

Avec probabilité au moins  $1 - \delta$  :

$$|\hat{d} - d| \leq Cd \left( B \left( \frac{k}{n} \right)^{\frac{1}{d}} + \sqrt{\frac{\log(4/\delta)}{k}} \right),$$

$$C \simeq \left( \min_{i=1..n} \eta_0^{(i)} \right)^{-\frac{1}{d}}.$$

Si  $\frac{k}{n} < (2BCd)^{-d}$ ,

$$\mathbb{P} \left( [\hat{d}] \neq d \right) \leq 4 \exp \left( -k \left( \frac{1}{2Cd} - B \left( \frac{k}{n} \right)^{\frac{1}{d}} \right)^2 \right).$$

$\Rightarrow$  convergence exponentielle (pour  $n$  "grand" ...).

# Convergence théorique

Avec probabilité au moins  $1 - \delta$  :

$$|\hat{d} - d| \leq Cd \left( B \left( \frac{k}{n} \right)^{\frac{1}{d}} + \sqrt{\frac{\log(4/\delta)}{k}} \right),$$

$$C \simeq \left( \min_{i=1..n} \eta_0^{(i)} \right)^{-\frac{1}{d}}.$$

Si  $\frac{k}{n} < (2BCd)^{-d}$ ,

$$\mathbb{P} \left( [\hat{d}] \neq d \right) \leq 4 \exp \left( -k \left( \frac{1}{2Cd} - B \left( \frac{k}{n} \right)^{\frac{1}{d}} \right)^2 \right).$$

$\Rightarrow$  convergence exponentielle (pour  $n$  "grand" ...).

Choix pratique :  $k \simeq \lceil \sqrt{n} \rceil$ .

# Exemple I - hypersphère

$$S^n = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} / \sum_{i=1}^{n+1} x_i^2 = R^2\} \subset \mathbb{R}^{n+1}, \text{ dimension } n.$$

# Exemple I - hypersphère

$$S^n = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} / \sum_{i=1}^{n+1} x_i^2 = R^2\} \subset \mathbb{R}^{n+1}, \text{ dimension } n.$$

Noir : 100 pts. ; rouge : 200 pts. ;  
bleu : 500 pts. ; vert : 1000 pts.

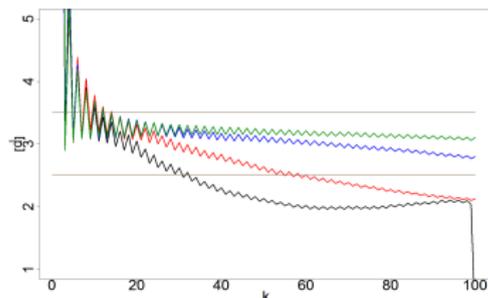


FIGURE:  $S_3 \subset \mathbb{R}^4$ .

# Exemple I - hypersphère

$$S^n = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} / \sum_{i=1}^{n+1} x_i^2 = R^2\} \subset \mathbb{R}^{n+1}, \text{ dimension } n.$$

Noir : 100 pts. ; rouge : 200 pts. ;  
bleu : 500 pts. ; vert : 1000 pts.

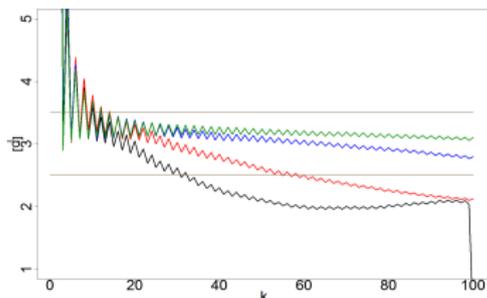


FIGURE:  $S_3 \subset \mathbb{R}^4$ .

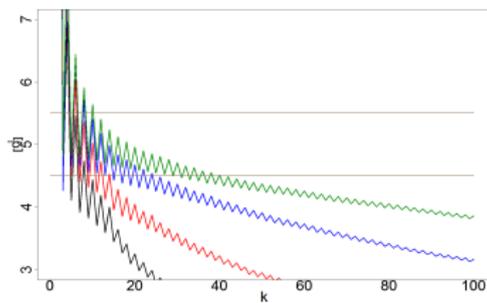


FIGURE:  $S_5 \subset \mathbb{R}^6$ .

# Exemple I - hypersphère

$$S^n = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} / \sum_{i=1}^{n+1} x_i^2 = R^2\} \subset \mathbb{R}^{n+1}, \text{ dimension } n.$$

Noir : 100 pts. ; rouge : 200 pts. ;  
bleu : 500 pts. ; vert : 1000 pts.

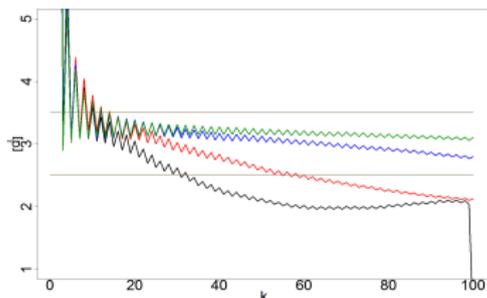


FIGURE:  $S_3 \subset \mathbb{R}^4$ .

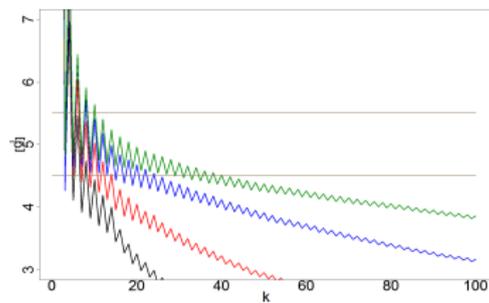


FIGURE:  $S_5 \subset \mathbb{R}^6$ .

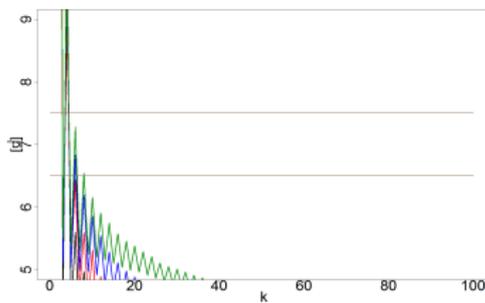
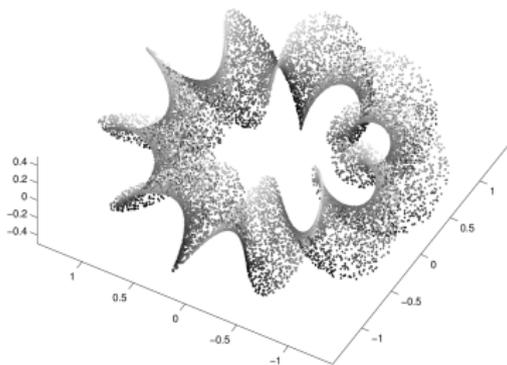


FIGURE:  $S_7 \subset \mathbb{R}^8$ .

## Exemple II - $k$ -twist Möbius

$k$  torsions appliquées à un ruban, puis recollage des deux bouts.

$$x_1(u, v) = \left(1 + \frac{u}{2} \cos\left(\frac{k}{2}v\right)\right) \cos v, \quad x_2(u, v) = \left(1 + \frac{u}{2} \cos\left(\frac{k}{2}v\right)\right) \sin v, \\ x_3(u, v) = \frac{u}{2} \sin\left(\frac{k}{2}v\right), \quad \text{avec } u \in [-1, 1], \quad v \in [0, 2\pi[.$$



**FIGURE:** 16000 points sur le "10-Ruban" de Möbius

## Exemple II - $k$ -twist Möbius

$k$  torsions appliquées à un ruban, puis recollage des deux bouts.

$$x_1(u, v) = \left(1 + \frac{u}{2} \cos\left(\frac{k}{2}v\right)\right) \cos v, \quad x_2(u, v) = \left(1 + \frac{u}{2} \cos\left(\frac{k}{2}v\right)\right) \sin v, \\ x_3(u, v) = \frac{u}{2} \sin\left(\frac{k}{2}v\right), \quad \text{avec } u \in [-1, 1], \quad v \in [0, 2\pi[.$$

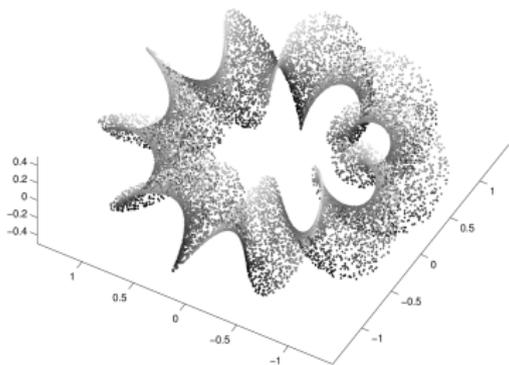


FIGURE: 16000 points sur le "10-Ruban" de Möbius

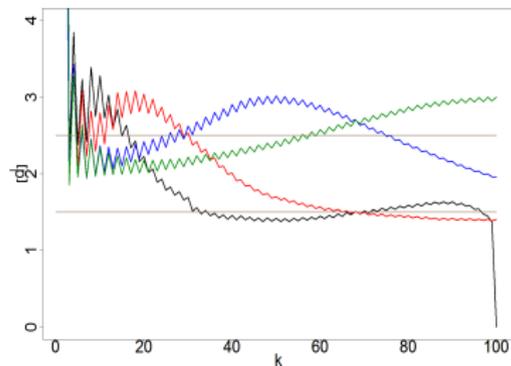
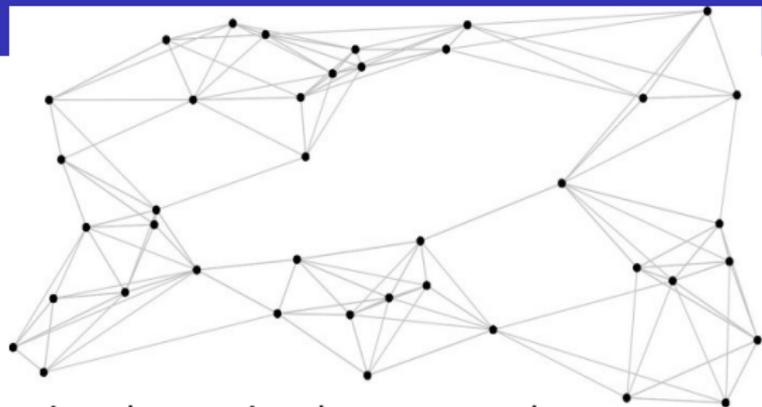


FIGURE: 100 points ; 200 points ;  
500 points ; 1000 points

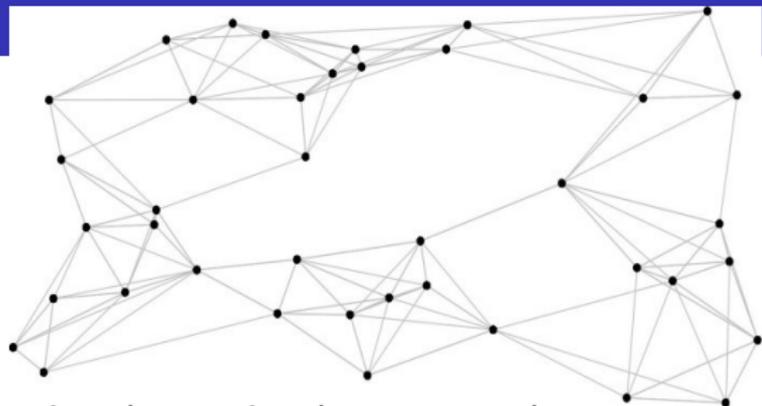
- 1 Hypothèses et motivations
- 2 Estimation de la dimension  $d$
- 3 Riemannian Manifold Learning (Lin et al. 2006)
- 4 Applications



**FIGURE:** Représentation des sorties dans un graphe ;  
sommets = données, arêtes = distances.

$y_i$  se connecte à  $y_j$  si :

- $\|y_i - y_j\| \leq \epsilon \rightarrow \epsilon$ -graphe (symétrique) ;
- $y_j$  est parmi les  $k$  plus proches voisins de  $y_i$  (norme  $L_2$  p.ex.).



**FIGURE:** Représentation des sorties dans un graphe ;  
sommets = données, arêtes = distances.

$y_i$  se connecte à  $y_j$  si :

- $\|y_i - y_j\| \leq \epsilon \rightarrow \epsilon$ -graphe (symétrique) ;
- $y_j$  est parmi les  $k$  plus proches voisins de  $y_i$  (norme  $L_2$  p.ex.).

# Étape préliminaire

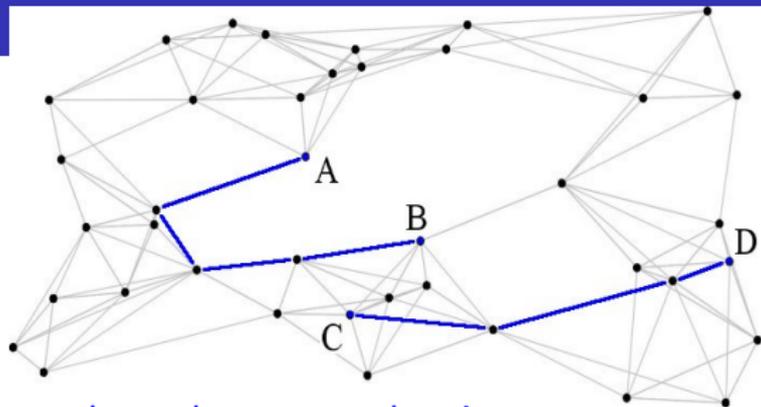


FIGURE: En bleu : deux plus courts chemins ;

$y_i$  se connecte à  $y_j$  si :

- $\|y_i - y_j\| \leq \epsilon \rightarrow \epsilon$ -graphe (symétrique) ;
- $y_j$  est parmi les  $k$  plus proches voisins de  $y_i$  (norme  $L_2$  p.ex.).

## Recherche de $y_0$ (origine)

Pour chaque sommet (= courbe  $y_i$ ), calculer la somme  $S$  des plus courts chemins à tous les autres sommets. Choisir  $y_0 = \arg \min S$

# Étape locale I

Base locale  $Q_0 = (e_1, \dots, e_d)$  de l'espace tangent en  $y_0$   
(ACP sur un voisinage de  $y_0$ )

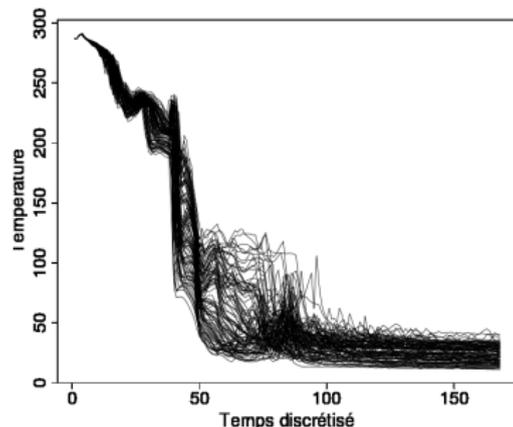


FIGURE: 100 courbes CATHARE I

# Étape locale I

Base locale  $Q_0 = (e_1, \dots, e_d)$  de l'espace tangent en  $y_0$   
(ACP sur un **voisinage de  $y_0$** )

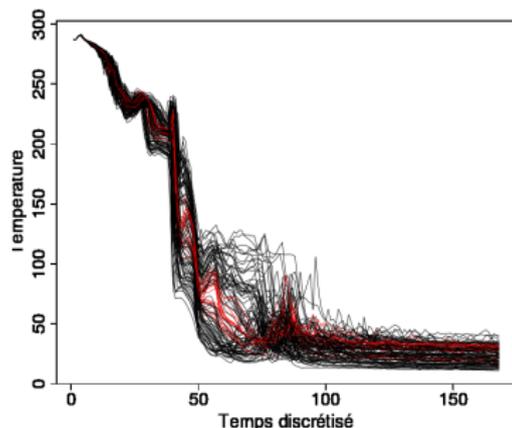


FIGURE: 15 voisins de  $y_0$

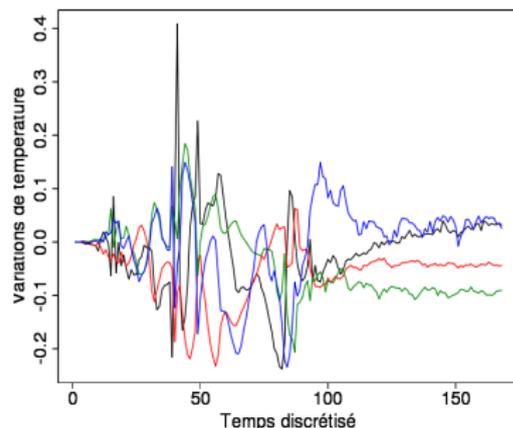


FIGURE: Base locale en  $y_0$  ( $d = 4$ )

# Étape locale II

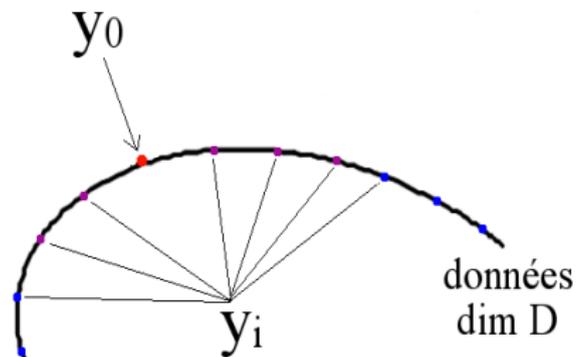


FIGURE: Courbe origine  $y_0$

# Étape locale II

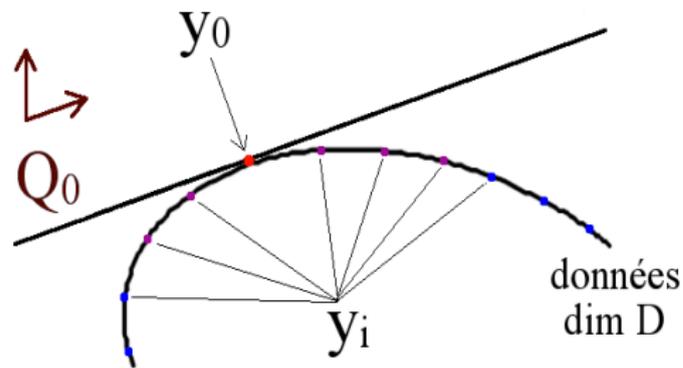


FIGURE: Espace tangent en  $y_0$  + base locale  $Q_0$

# Étape locale II

Coordonnées réduites  $z_i$  calculées par projection sur  $Q_0$ ,

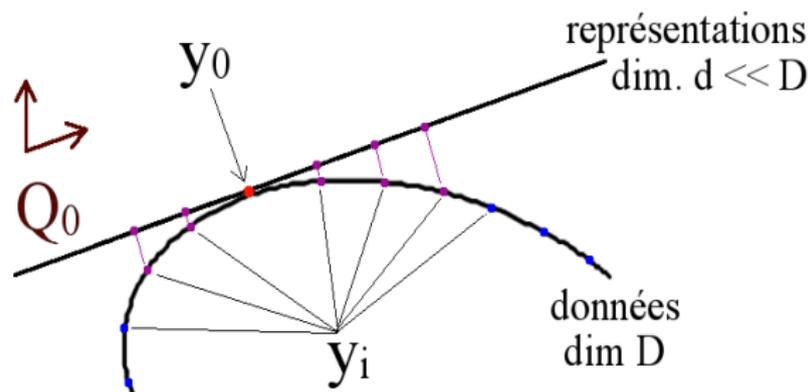


FIGURE: Coordonnées locales  $z_i$  sur l'espace tangent

# Étape locale II

Coordonnées réduites  $z_i$  calculées par projection sur  $Q_0$ ,  
+ normalisation pour satisfaire la contrainte  $\|y_i - y_0\| = \|z_i - z_0\|$ .

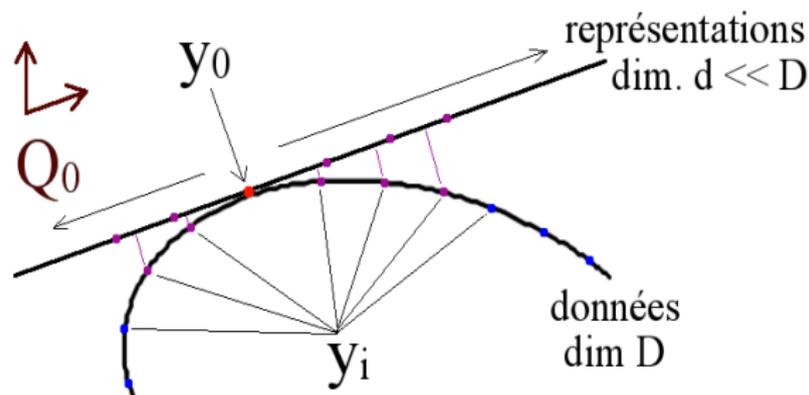


FIGURE: Normalisation des coordonnées  $z_i$

## $y_i$ "loin" de $y_0$

- $y_p$  = prédécesseur de  $y_i$  sur un plus court chemin depuis  $y_0$
- $y_{i_1}, \dots, y_{i_d}$  = voisins de  $y_p$  dont les coordonnées  $z_{i_k}$  sont connues

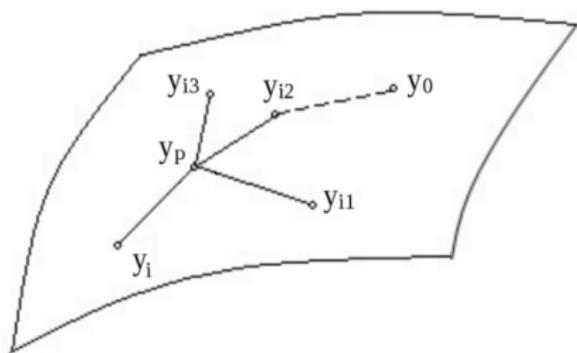


FIGURE: Données  $y_i$  en dim.  $D$

# $y_i$ "loin" de $y_0$

- $y_p$  = prédécesseur de  $y_i$  sur un plus court chemin depuis  $y_0$
- $y_{i_1}, \dots, y_{i_d}$  = voisins de  $y_p$  dont les coordonnées  $z_{i_k}$  sont connues

$z_i = r(y_i)$  calculé en..

- conservant au mieux les angles :  $\cos \widehat{z_i z_p z_{i_j}} \simeq \cos \widehat{y_i y_p y_{i_j}}$   
avec  $\cos \widehat{u_1 u_2 u_3} = \frac{\langle u_1 - u_2, u_3 - u_2 \rangle}{\|u_1 - u_2\| \|u_3 - u_2\|}$  ;
- sous la contrainte de normalisation  $\|y_i - y_p\| = \|z_i - z_p\|$ .

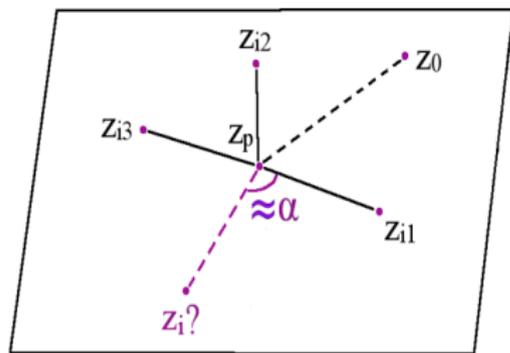
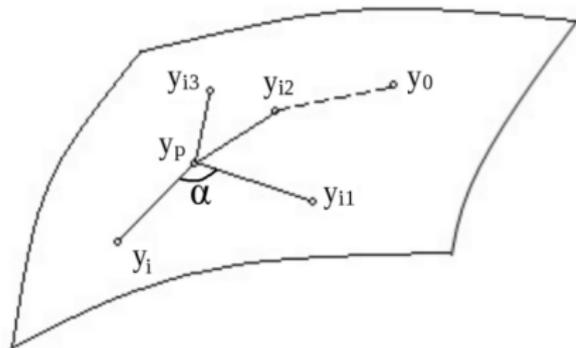


FIGURE: Données  $y_i$  en dim.  $D$

FIGURE:  $z_i = r(y_i)$  en dim.  $d \ll D$

# Exemple I

Swissroll fonctionnel :

$$f_{\alpha,\beta,\gamma}(x \in [0, 4\pi]) =$$

$$\alpha \cos x + \beta \sin x + \frac{\gamma}{2} \sin 2x$$

$$(\alpha = t \cos t, \beta = t \sin t, \gamma = u),$$

$$t \in [0, 2\pi]$$

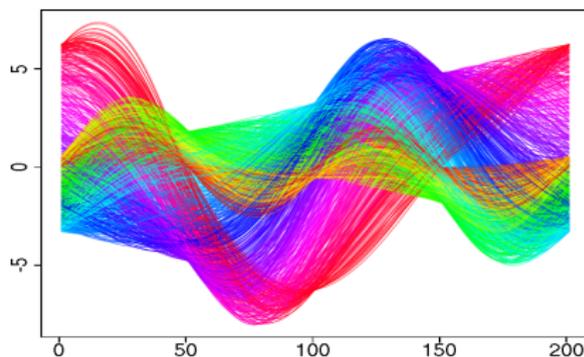


FIGURE: Swissroll fonctionnel

# Exemple 1

Swissroll fonctionnel :

$$f_{\alpha,\beta,\gamma}(x \in [0, 4\pi]) = \alpha \cos x + \beta \sin x + \frac{\gamma}{2} \sin 2x$$

$(\alpha = t \cos t, \beta = t \sin t, \gamma = u),$   
 $t \in [0, 2\pi]$

	ACP	RML
$d = 1$	3.44	0.9
$d = 2$	0.84	0.003
$d = 3$	0	0.001

TABLE: MSE moyenne en fonction de  $d$

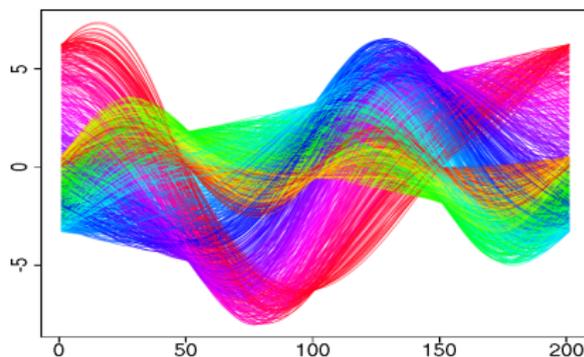


FIGURE: Swissroll fonctionnel

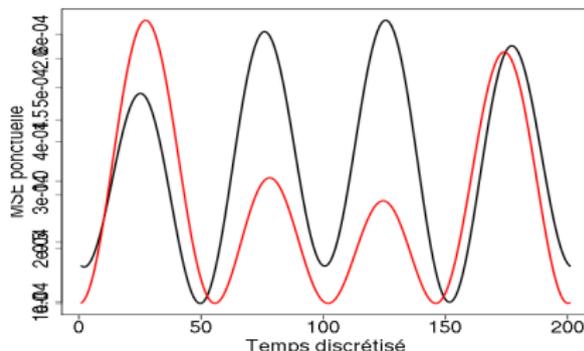


FIGURE: MSE RML (noir) et ACP (rouge) pour  $d = 2$

## Exemple II

$\mathcal{T} = n$  nombres aléatoires dans  $[0, 2\pi]$ .

= paramètres des  $n$  courbes correspondantes

$$f_i(t) = \sum_{j=1}^{d/2} \cos((2j-1)\mathcal{T}[i]) \cos((2j-1)t) + \sum_{j=1}^{d/2} \sin(2j\mathcal{T}[i]) \sin(2jt)$$

## Exemple II

$\mathcal{T} = n$  nombres aléatoires dans  $[0, 2\pi]$ .

= paramètres des  $n$  courbes correspondantes

$$f_i(t) = \sum_{j=1}^{d/2} \cos((2j-1)\mathcal{T}[i]) \cos((2j-1)t) + \sum_{j=1}^{d/2} \sin(2j\mathcal{T}[i]) \sin(2jt)$$

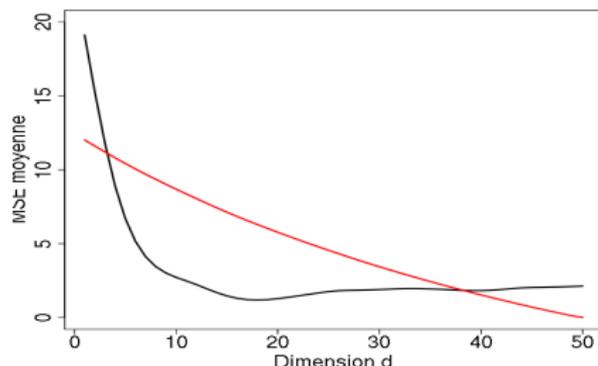
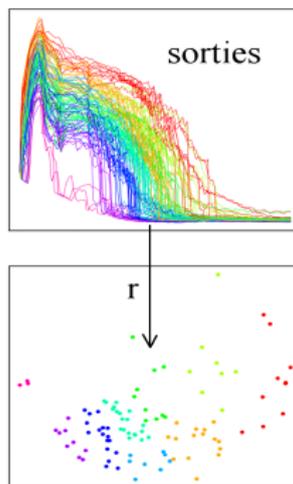


FIGURE: MSE RML (noir) et ACP (rouge) en fonction de  $d$

Croisements entre  $d = 2$  et  $d = 3$  – entre  $d = 38$  et  $d = 39$ .

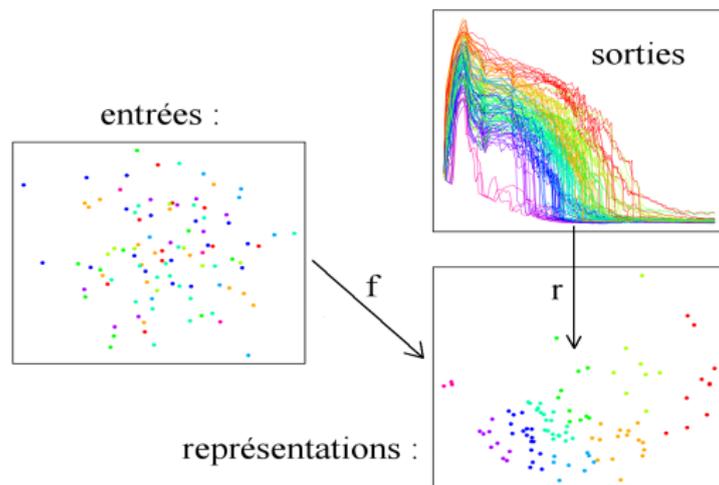
- 1 Hypothèses et motivations
- 2 Estimation de la dimension  $d$
- 3 Riemannian Manifold Learning (Lin et al. 2006)
- 4 Applications

# Résumé des étapes



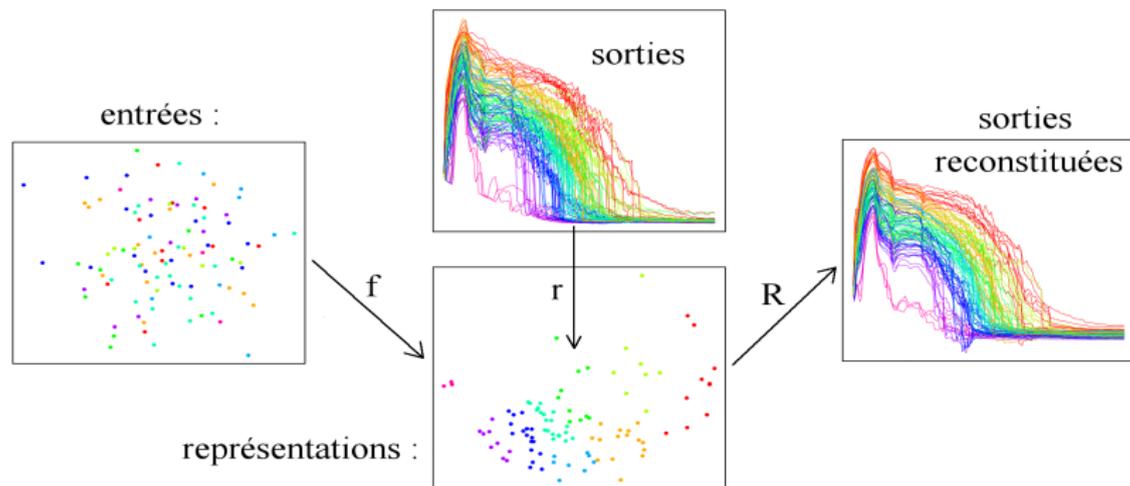
① réduction de dimension  $r : V \rightarrow \mathbb{R}^d \Rightarrow$  représentations  $r_i$  ;

# Résumé des étapes



- 1 réduction de dimension  $r : V \rightarrow \mathbb{R}^d \Rightarrow$  représentations  $r_i$  ;
- 2 apprentissage statistique  $f : \mathbb{R}^p \rightarrow \mathbb{R}^d$   
"entrées  $\mapsto$  représentations",  $f(x_i) \simeq r_i$  ;

# Résumé des étapes



- 1 réduction de dimension  $r : V \rightarrow \mathbb{R}^d \Rightarrow$  représentations  $r_i$  ;
- 2 apprentissage statistique  $f : \mathbb{R}^p \rightarrow \mathbb{R}^d$   
"entrées  $\mapsto$  représentations",  $f(x_i) \simeq r_i$  ;
- 3 prédiction d'une nouvelle courbe :  $\hat{y} = R(f(x)) = R(\hat{r})$ .

# Validation du modèle

Données :

- entraînement =  $\{(x_i, y_i), i = 1, \dots, n\}$  ;
- test =  $\{(x'_i, y'_i), i = 1, \dots, m\}$  ;

Prédictions du modèle :  $\hat{y}'_i = M(x'_i), i = 1, \dots, m$ .

# Validation du modèle

Données :

- entraînement =  $\{(x_i, y_i), i = 1, \dots, n\}$  ;
- test =  $\{(x'_i, y'_i), i = 1, \dots, m\}$  ;

Prédictions du modèle :  $\hat{y}'_i = M(x'_i), i = 1, \dots, m$ .

Mesure "absolue" puis relative de l'erreur ponctuelle

$$MSE[j] = \frac{1}{m} \sum_{i=1}^m (\hat{y}'_i(j) - y'_i(j))^2, \quad j = 1, \dots, D \text{ (discrétisation)}.$$

# Validation du modèle

Données :

- entraînement =  $\{(x_i, y_i), i = 1, \dots, n\}$  ;
- test =  $\{(x'_i, y'_i), i = 1, \dots, m\}$  ;

Prédictions du modèle :  $\hat{y}'_i = M(x'_i), i = 1, \dots, m$ .

Mesure "absolue" puis relative de l'erreur ponctuelle

$$MSE[j] = \frac{1}{m} \sum_{i=1}^m (\hat{y}'_i(j) - y'_i(j))^2, \quad j = 1, \dots, D \text{ (discrétisation).}$$

$$Q_2[j] = 1 - \frac{m \cdot MSE[j]}{\sum_{i=1}^m (\bar{y}(j) - y'_i(j))^2} \text{ (comparaison à la moyenne).}$$

$-\infty < Q_2 \leq 1$  :  $\leq 0 \Rightarrow$  (très) mauvais modèle ;  
 $\simeq 1 \Rightarrow$  modèle parfait.

# Test I - courbes de température CATHARE

100 évaluations du modèle,  
4 dimensions en entrée,  
168 points de discrétisation.

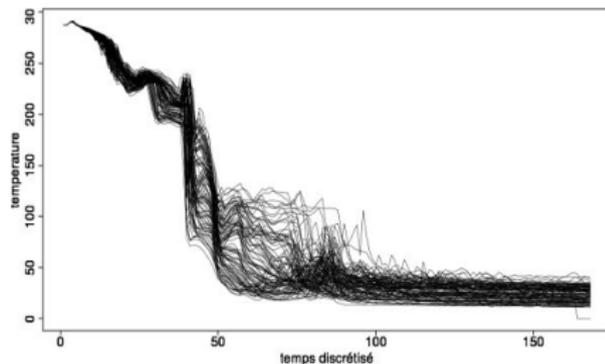


FIGURE: Les 100 courbes en sortie

# Test I - courbes de température CATHARE

100 évaluations du modèle,  
4 dimensions en entrée,  
168 points de discrétisation.

validation croisée  
leave-10-out :

MSE à g.,  $Q_2$  à d. ;  $\hat{d} = 4$

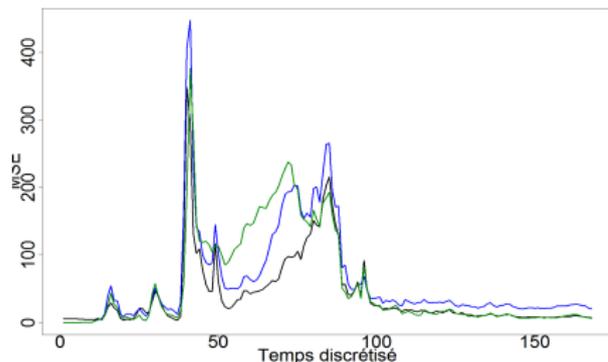


FIGURE: Noir : ACP fonctionnelle ; bleu : RML ; vert : Nadaraya-Watson.

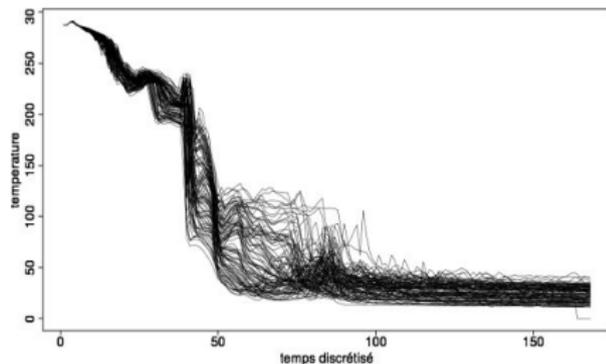
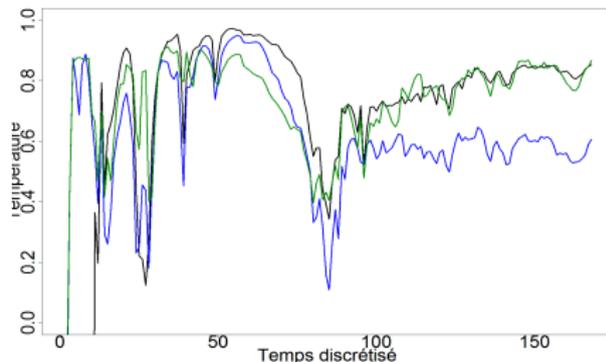


FIGURE: Les 100 courbes en sortie



# 5 courbes prédites

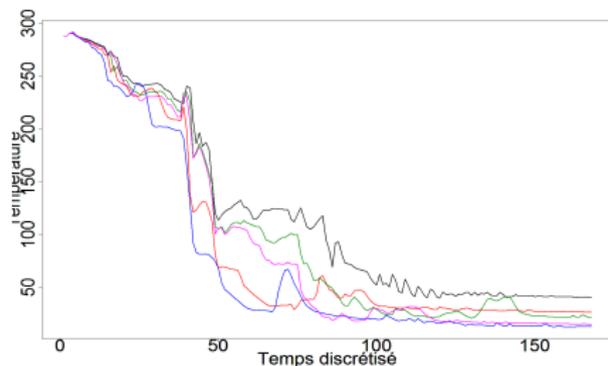


FIGURE: Courbes cibles

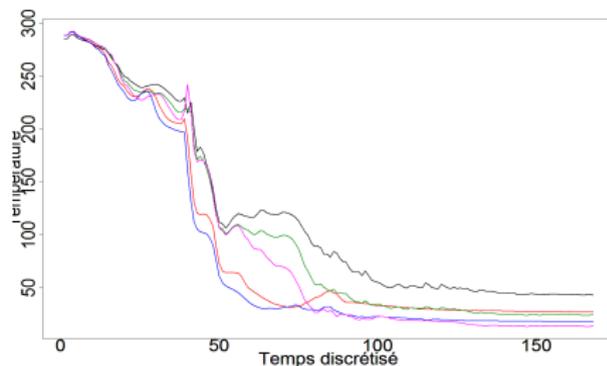


FIGURE: ACP fonc.

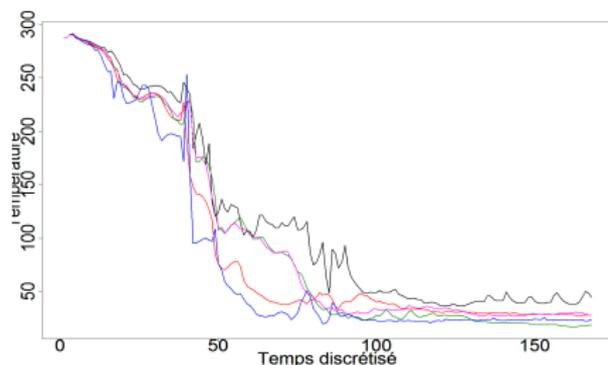


FIGURE: RML

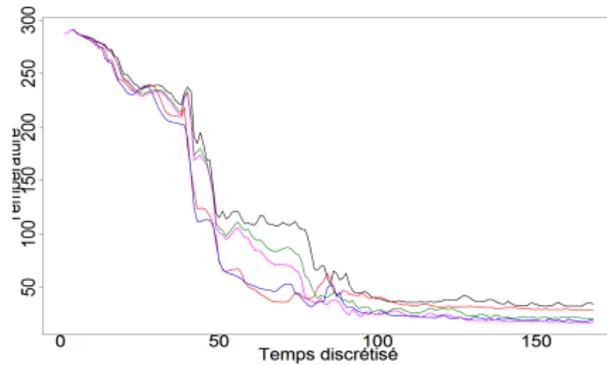


FIGURE: Nadaraya-Watson

# Test II - courbes de température CATHARE

600 évaluations du modèle,  
11 dimensions en entrée,  
414 points de discrétisation.

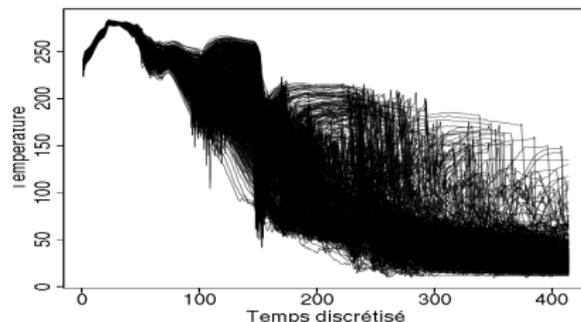


FIGURE: Les 600 courbes en sortie

# Test II - courbes de température CATHARE

600 évaluations du modèle,  
11 dimensions en entrée,  
414 points de discrétisation.

validation croisée  
leave-10-out :

MSE à g.,  $Q_2$  à d. ;  $\hat{d} = 7$

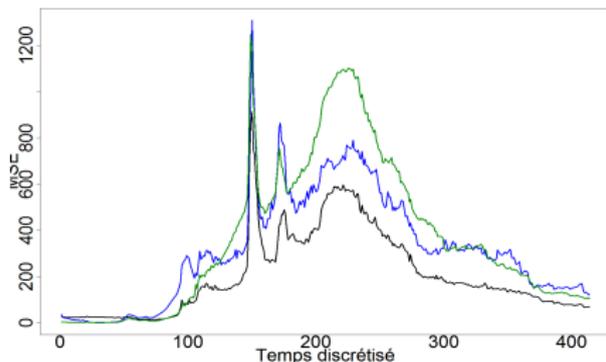


FIGURE: Noir : ACP fonctionnelle ; bleu : RML ; vert : Nadaraya-Watson.

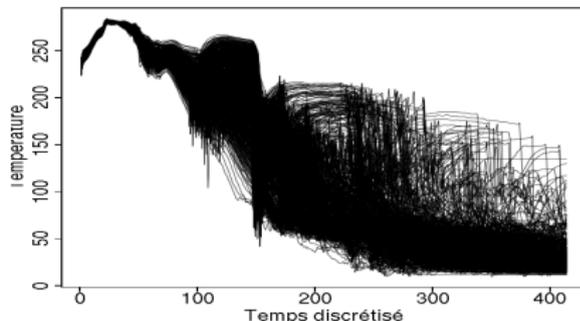
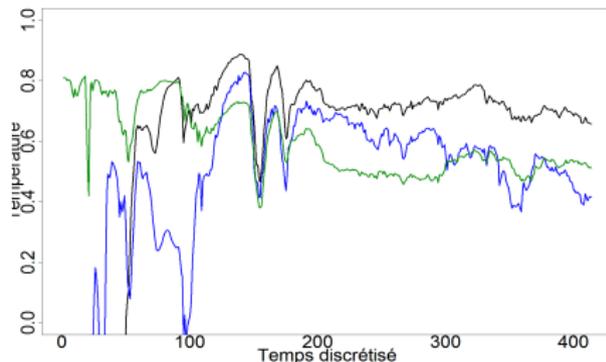


FIGURE: Les 600 courbes en sortie



# 5 courbes prédites

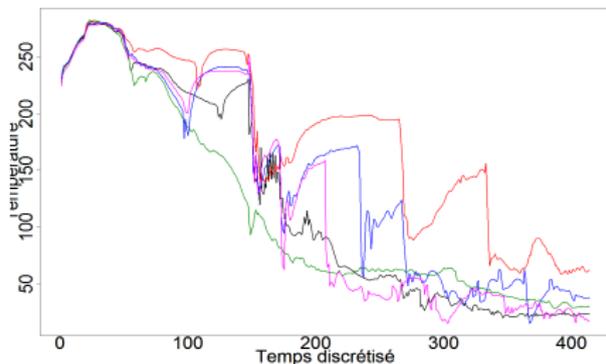


FIGURE: Courbes cibles

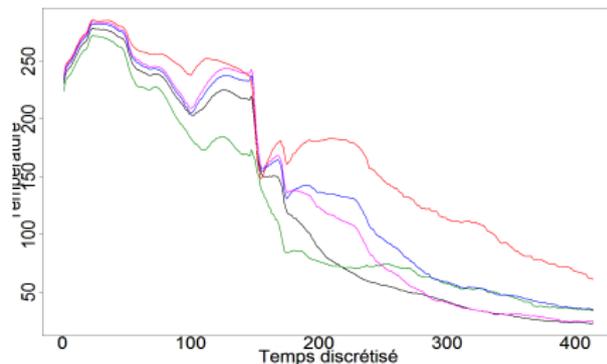


FIGURE: ACP fonc.

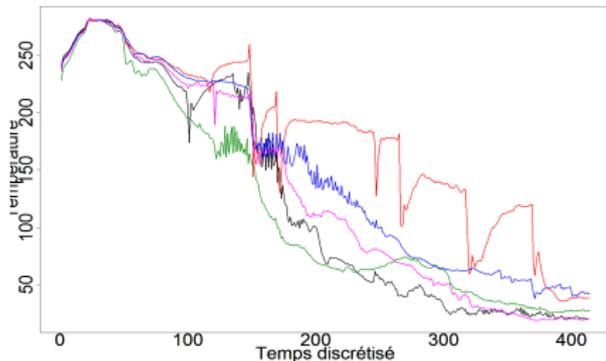


FIGURE: RML

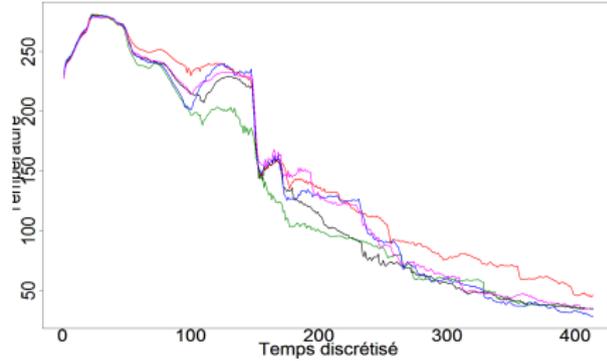


FIGURE: Nadaraya-Watson

## Estimation de $d$ : autres approches (géométriques)

- $\hat{d}$  = cardinal du plus grand simplexe du graphe de voisinage "de visibilité" (Lin et Zha, 2006) ;
- détection des "slivers" (simplexes de volume négligeable) dans le graphe de voisinage (Cheng et Chiu, 2009).

## Estimation de $d$ : autres approches (géométriques)

- $\hat{d}$  = cardinal du plus grand simplexe du graphe de voisinage "de visibilité" (Lin et Zha, 2006) ;
- détection des "slivers" (simplexes de volume négligeable) dans le graphe de voisinage (Cheng et Chiu, 2009).

## Réduction de la dimension

ACPF (linéaire) et RML : complémentaires.

Preuve de convergence (RML + LPcaML) à écrire.

Quelques paramètres ne sont pas optimisés automatiquement ...

# Conclusion

## Estimation de $d$ : autres approches (géométriques)

- $\hat{d}$  = cardinal du plus grand simplexe du graphe de voisinage "de visibilité" (Lin et Zha, 2006) ;
- détection des "slivers" (simplexes de volume négligeable) dans le graphe de voisinage (Cheng et Chiu, 2009).

## Réduction de la dimension

ACPF (linéaire) et RML : complémentaires.

Preuve de convergence (RML + LPcaML) à écrire.

Quelques paramètres ne sont pas optimisés automatiquement ...

Validation : dimension fractale d'une courbe au lieu du MSE ?