

Construction d'un métamodèle fonctionnel

Benjamin Auder

CEA - UPMC

21 octobre 2010

Thèse depuis 02/2008

Directeur de thèse : Gérard Biau (UPMC)

Superviseur CEA : Michel Marquès

Superviseur EDF : Bertrand looss

Contexte industriel

Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagées.

But : estimer leurs probabilités.

Contexte industriel

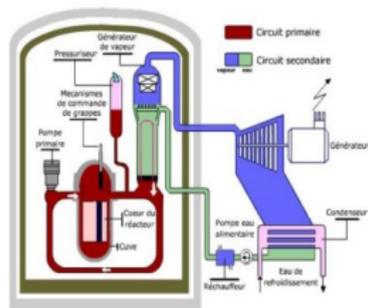
Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagées.

But : estimer leurs probabilités.

Méthodologie

Modélisation



Contexte industriel

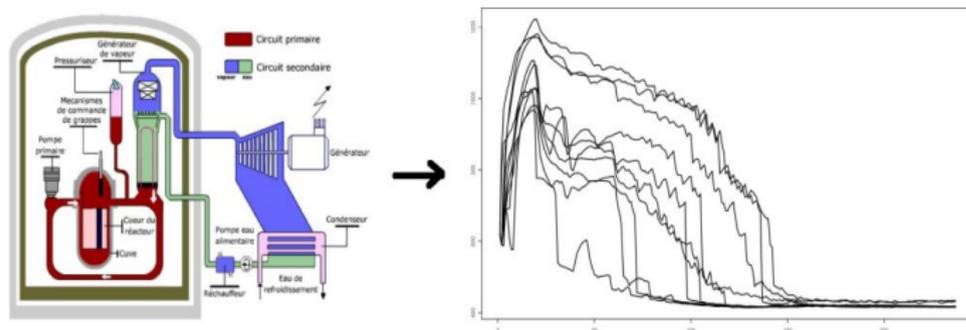
Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagées.

But : estimer leurs probabilités.

Méthodologie

Modélisation → Simulation



Contexte industriel

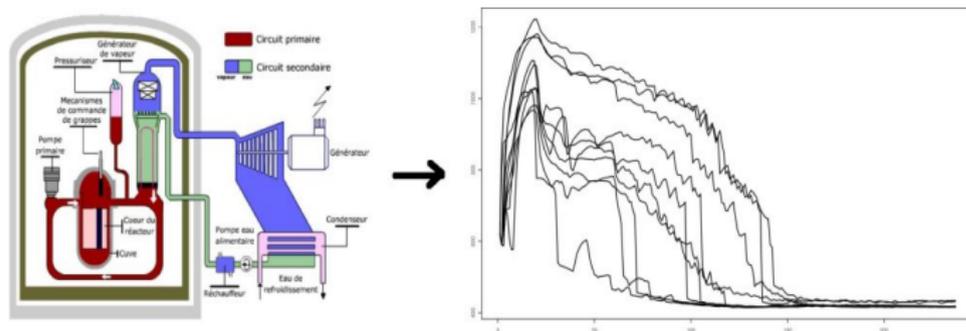
Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagées.

But : estimer leurs probabilités.

Méthodologie

Modélisation → Simulation → Calculs.



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

Contexte industriel

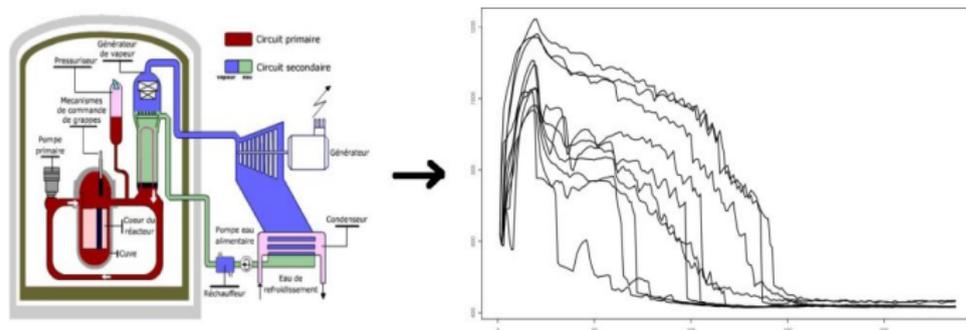
Cadre : durée de vie des cuves.

→ 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 CATHARE coûteux en temps

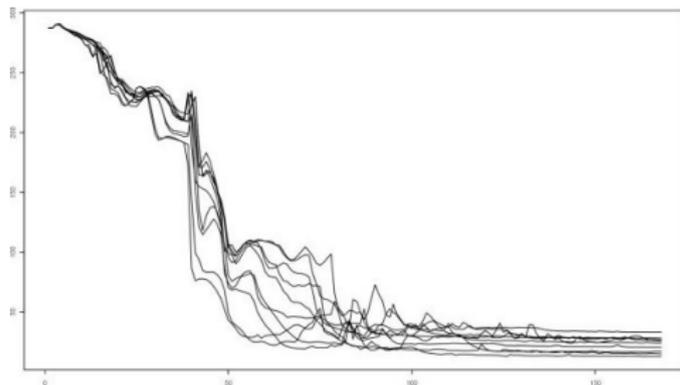


FIG.: Transitoires de température.

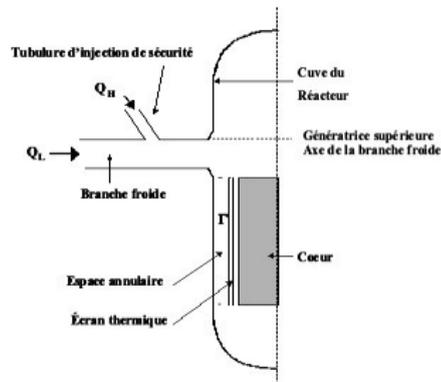


FIG.: Zone modélisée

Au CEA - DER/SESI/LSMR ...

Code thermo-hydraulique CATHARE coûteux en temps

- "boîte noire" ;

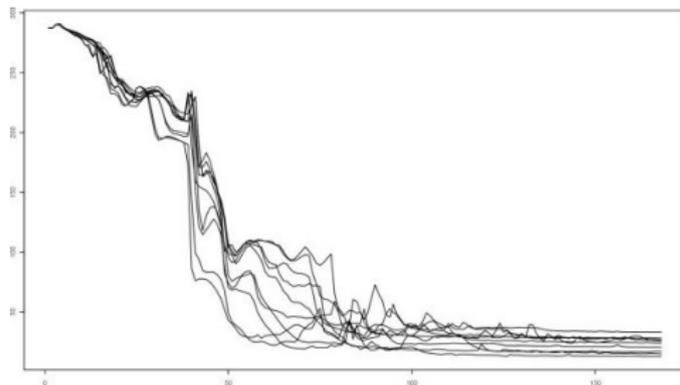


FIG.: Transitoires de température.

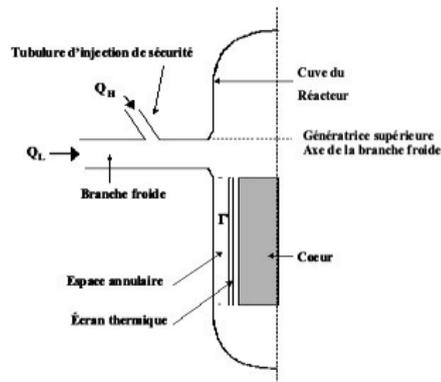


FIG.: Zone modélisée

Au CEA - DER/SESI/LSMR ...

Code thermo-hydraulique CATHARE coûteux en temps

- "boîte noire" ;
- paramètres d'entrée incertains.

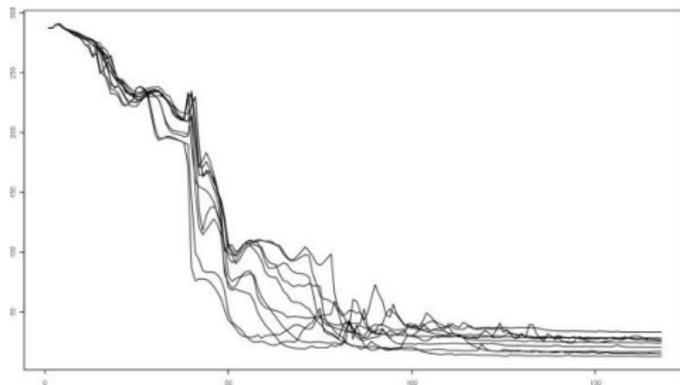


FIG.: Transitoires de température.

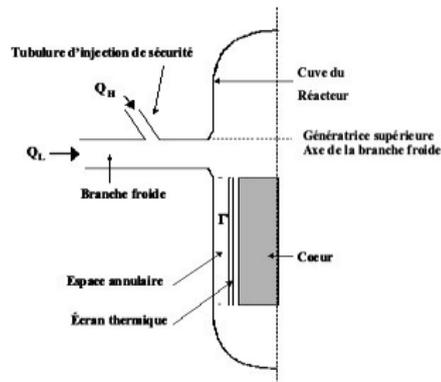


FIG.: Zone modélisée

Au CEA - DER/SESI/LSMR ...

Code thermo-hydraulique CATHARE coûteux en temps

- "boîte noire" ;
- paramètres d'entrée incertains.

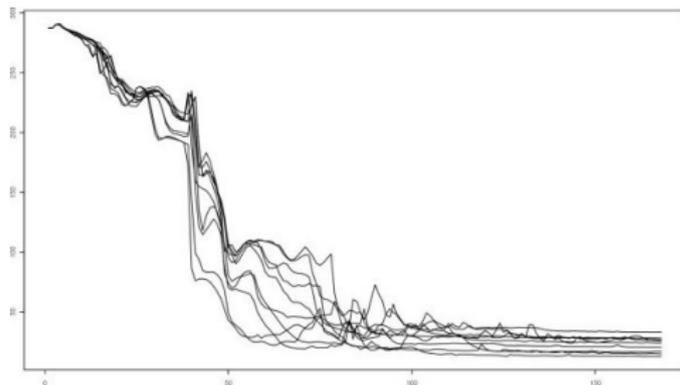


FIG.: Transitoires de température.

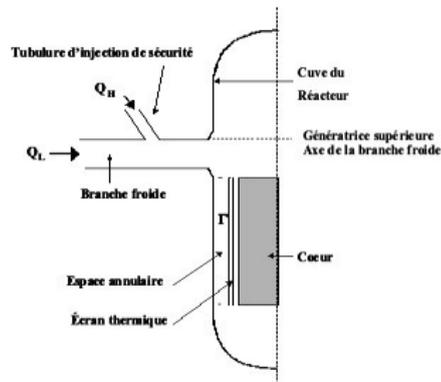


FIG.: Zone modélisée

"Accélérer" l'exécution du code CATHARE

Solution ?

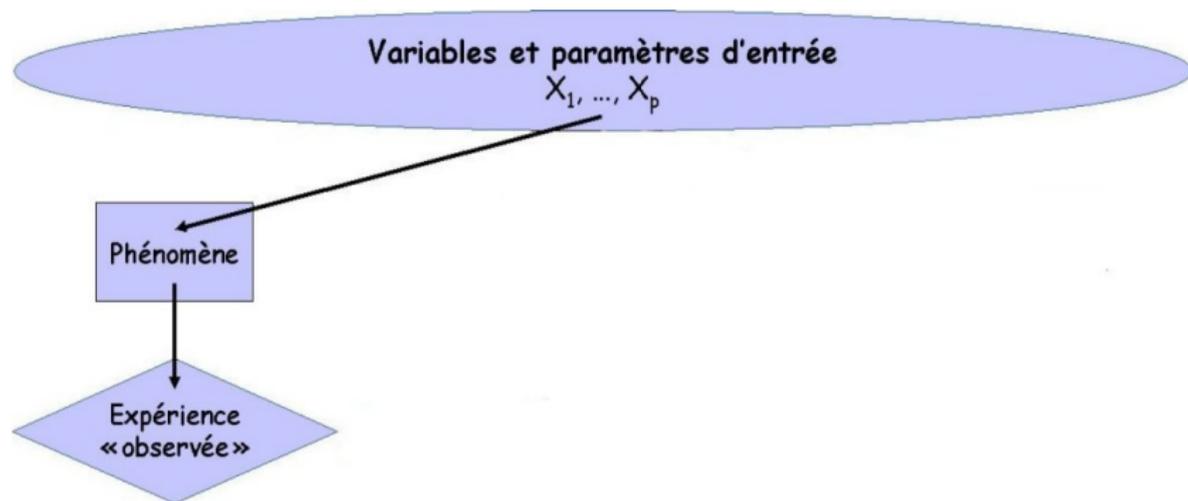


FIG.: Point de départ : phénomène physique.
analyses statistiques sur le phénomène

Solution ?

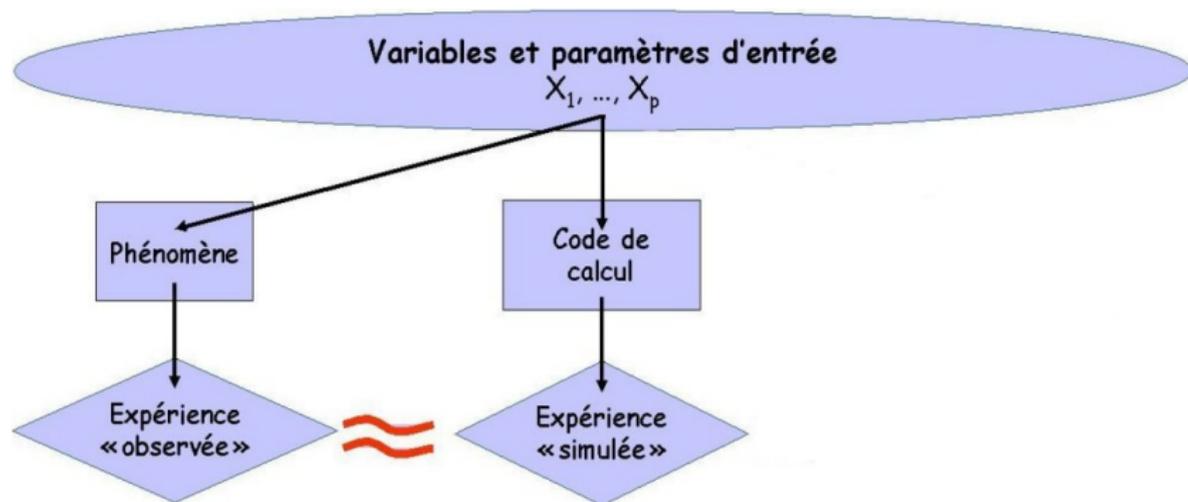


FIG.: Code de calcul = simulation du phénomène.

nécessitent analyses statistiques sur le phénomène
beaucoup de résultats de code

Solution : métamodèle

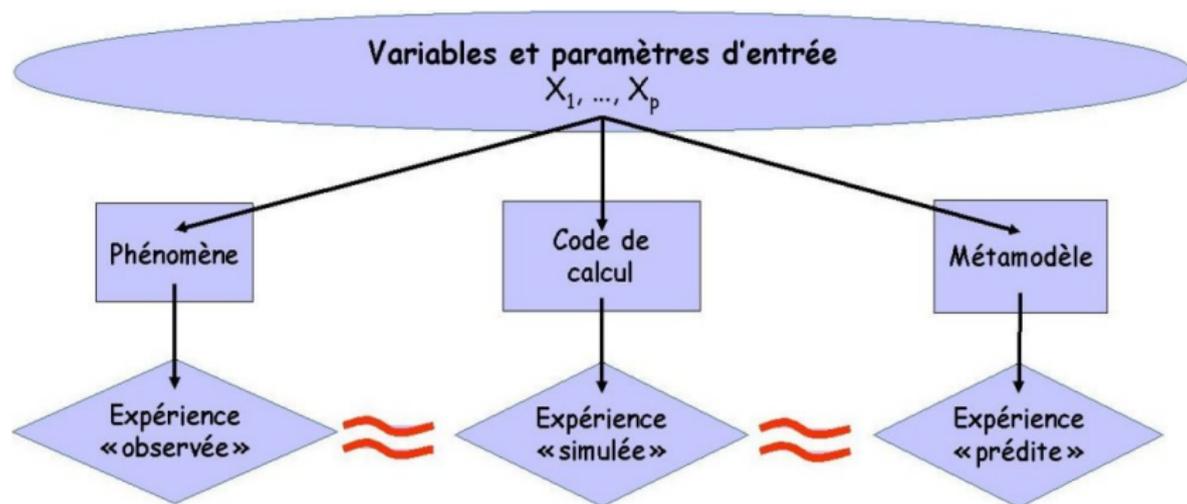


FIG.: Métamodèle = accélération des simulations.

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

Solution : métamodèle

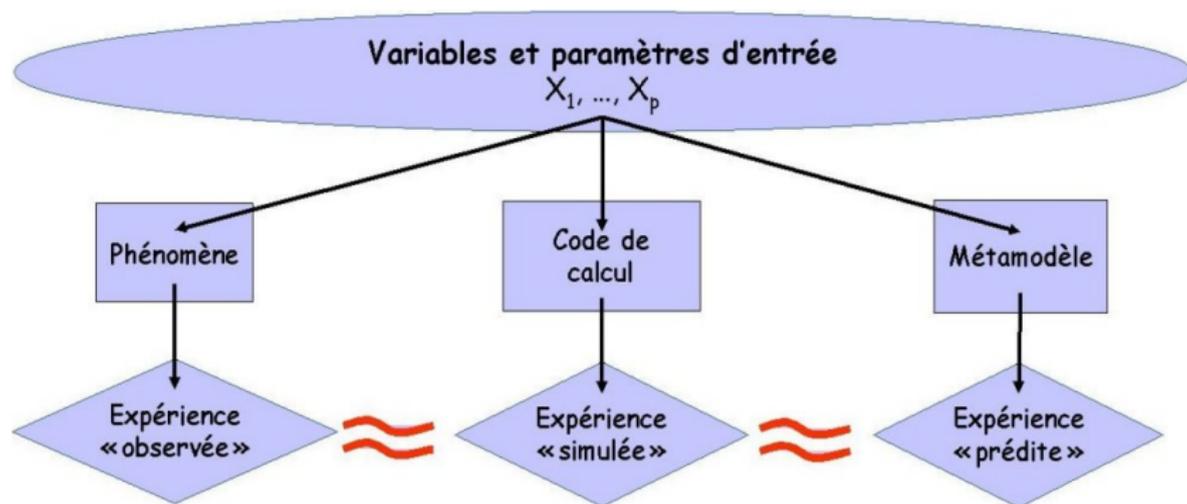


FIG.: Métamodèle = accélération des simulations.

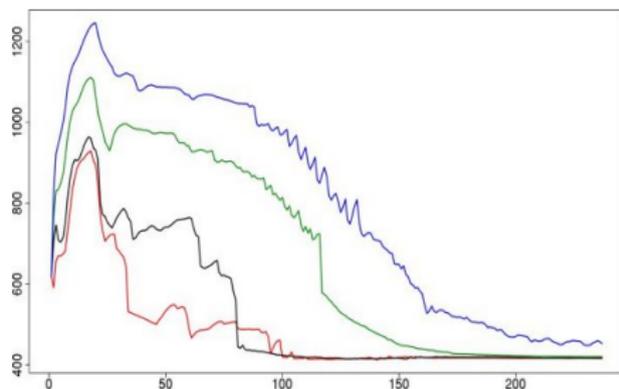
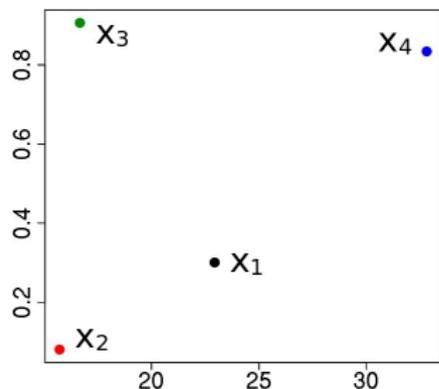
analyses statistiques sur le phénomène
nécessitent beaucoup de résultats de code
obtenus avec un métamodèle (= modèle du code).

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.



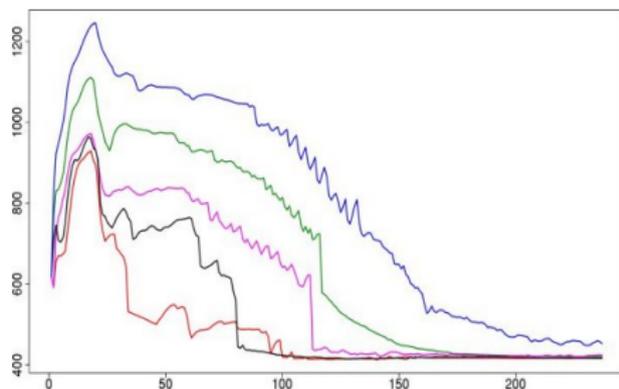
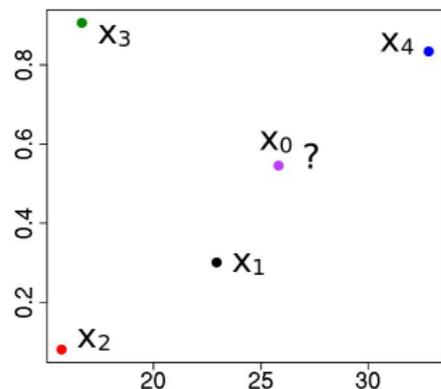
Reformulation finale

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 :

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



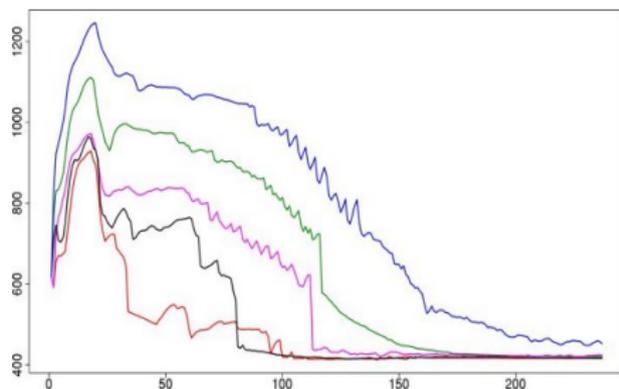
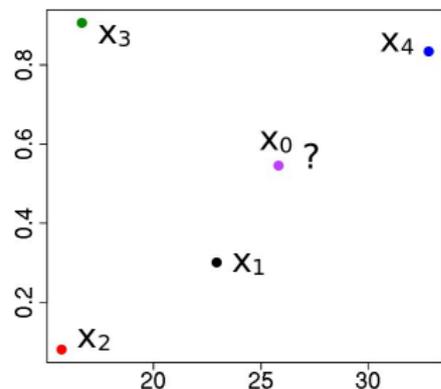
Reformulation finale

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 :

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



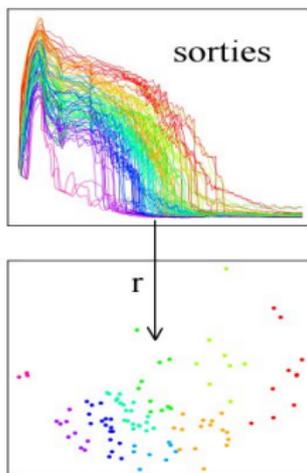
Apprentissage statistique "régression" $\mathbb{R}^p \rightarrow \mathcal{C}([a, b], \mathbb{R})$

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 :

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



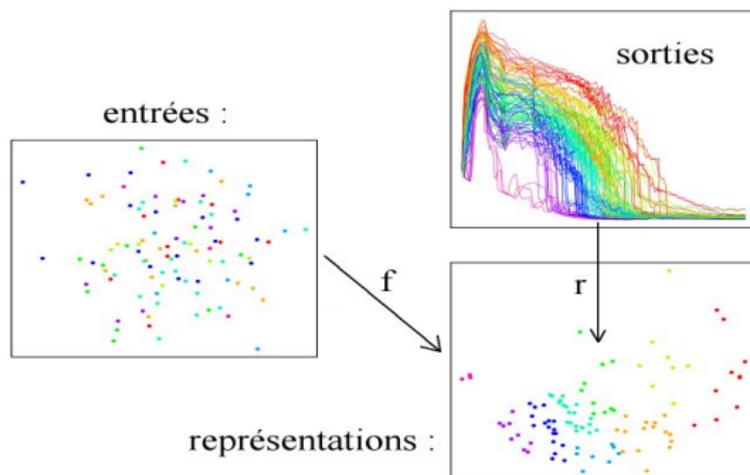
Comment se ramener au cas "simple" $y_i \in \mathbb{R}^d$?

① réduction de la dimension :

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

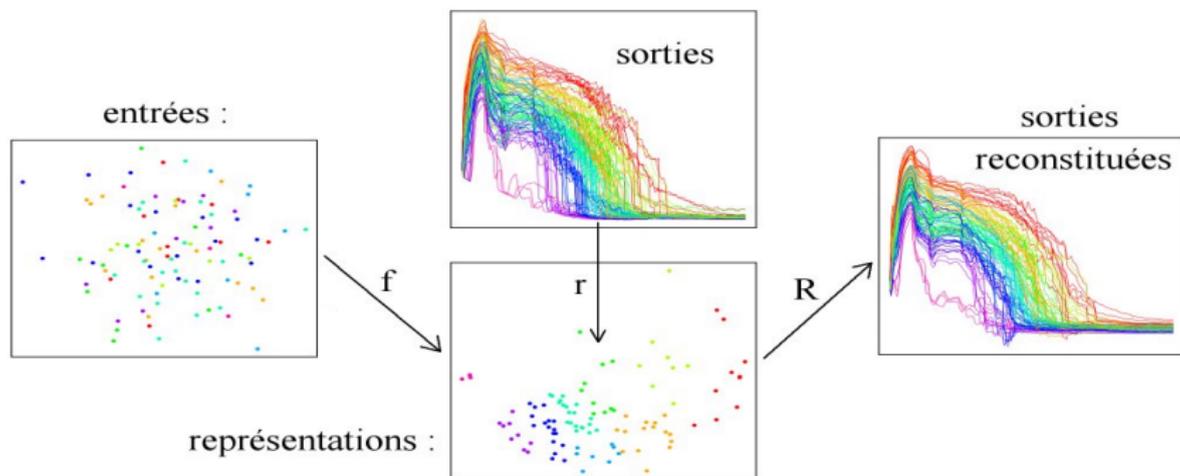
② apprentissage statistique classique :

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



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).



État de l'art

Méthodes "classiques"

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

État de l'art

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

État de l'art

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 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

Estimation de d

Hypothèse principale

Les courbes en sortie sont échantillonnées sur une variété (fonctionnelle) "suffisamment régulière"

Estimation de d

Hypothèse principale

Les courbes en sortie sont échantillonnées sur une variété (fonctionnelle) "suffisamment régulière"

Manifold-adaptive dimension estimation,
A. M. Farahmand et al. 2007

Méthode basée sur..

$$\mathbb{P}(Y \in B(y_i, r)) \simeq \eta_i r^d, \eta_i > 0$$

$$\mathbb{P}(Y \in B(y_i, r_k)) \simeq \frac{k}{n} \quad (r_k = \text{distance au } k^{\text{eme}} \text{ voisin}).$$

Estimation de d

Hypothèse principale

Les courbes en sortie sont échantillonnées sur une variété (fonctionnelle) "suffisamment régulière"

Manifold-adaptive dimension estimation,
A. M. Farahmand et al. 2007

Méthode basée sur..

$$\mathbb{P}(Y \in B(y_i, r)) \simeq \eta_i r^d, \eta_i > 0$$

$$\mathbb{P}(Y \in B(y_i, r_k)) \simeq \frac{k}{n} \quad (r_k = \text{distance au } k^{\text{eme}} \text{ voisin}).$$

\Rightarrow estimation locale d_i , puis moyenne = d .

Types de méthodes

Beaucoup de techniques différentes, dont les méthodes..

- ① linéaires : la représentation d -dimensionnelle s'écrit linéairement en fonction de y (bases orthonormées) ;

Types de méthodes

Beaucoup de techniques différentes, dont les méthodes..

- ① linéaires : la représentation d -dimensionnelle s'écrit linéairement en fonction de y (bases orthonormées) ;
- ② optimisant une formule globale, aboutissant souvent à la recherche de vecteurs propres (Isomap, LLE, ..etc) ;

Types de méthodes

Beaucoup de techniques différentes, dont les méthodes..

- ① linéaires : la représentation d -dimensionnelle s'écrit linéairement en fonction de y (bases orthonormées) ;
- ② optimisant une formule globale, aboutissant souvent à la recherche de vecteurs propres (Isomap, LLE, ..etc) ;
- ③ déterminant les représentations au fur et à mesure, en fonction de celles déjà déterminées (RML, LPcaML, ..etc) ;

Types de méthodes

Beaucoup de techniques différentes, dont les méthodes..

- ① linéaires : la représentation d -dimensionnelle s'écrit linéairement en fonction de y (bases orthonormées) ;
- ② optimisant une formule globale, aboutissant souvent à la recherche de vecteurs propres (Isomap, LLE, ..etc) ;
- ③ déterminant les représentations au fur et à mesure, en fonction de celles déjà déterminées (RML, LPcaML, ..etc) ;
- ④ géométriques, cherchant à déterminer la topologie des données par l'analyse du graphe les représentant. (...)

Types de méthodes

Beaucoup de techniques différentes, dont les méthodes..

- 1 linéaires : la représentation d -dimensionnelle s'écrit linéairement en fonction de y (bases orthonormées) ;
 - 2 optimisant une formule globale, aboutissant souvent à la recherche de vecteurs propres (Isomap, LLE, ..etc) ;
 - 3 déterminant les représentations au fur et à mesure, en fonction de celles déjà déterminées (RML, LPcaML, ..etc) ;
 - 4 géométriques, cherchant à déterminer la topologie des données par l'analyse du graphe les représentant. (...)
- 2 → pas assez de contrôle sur les propriétés (distances, angles..) des représentations des données.
- 4 → peut être intéressant, mais semble surtout utilisé pour des objets géométriques de petite dimension dans \mathbb{R}^D .

Types de méthodes

Beaucoup de techniques différentes, dont les méthodes..

- 1 linéaires : la représentation d -dimensionnelle s'écrit linéairement en fonction de y (bases orthonormées) ;
 - 2 optimisant une formule globale, aboutissant souvent à la recherche de vecteurs propres (Isomap, LLE, ..etc) ;
 - 3 déterminant les représentations au fur et à mesure, en fonction de celles déjà déterminées (RML, LPcaML, ..etc) ;
 - 4 géométriques, cherchant à déterminer la topologie des données par l'analyse du graphe les représentant. (...)
- 2 → pas assez de contrôle sur les propriétés (distances, angles..) des représentations des données.
- 4 → peut être intéressant, mais semble surtout utilisé pour des objets géométriques de petite dimension dans \mathbb{R}^D .

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

Introduction

$y_1, \dots, y_n \in \mathcal{C}([a, b])$.

Objectif = trouver la "meilleure" description du jeu de données par une famille orthonormée e_1, \dots, e_n (tronquée à $d \leq n$).

Introduction

$$y_1, \dots, y_n \in \mathcal{C}([a, b]).$$

Objectif = trouver la "meilleure" description du jeu de données par une famille orthonormée e_1, \dots, e_n (tronquée à $d \leq n$).

Méthode linéaire : $r(y) = (\langle y, e_1 \rangle, \dots, \langle y, e_d \rangle) = z$

$$R(z) = \sum_{j=1}^d z_j e_j \simeq y$$

Introduction

$$y_1, \dots, y_n \in \mathcal{C}([a, b]).$$

Objectif = trouver la "meilleure" description du jeu de données par une famille orthonormée e_1, \dots, e_n (tronquée à $d \leq n$).

Méthode linéaire : $r(y) = (\langle y, e_1 \rangle, \dots, \langle y, e_d \rangle) = z$

$$R(z) = \sum_{j=1}^d z_j e_j \simeq y$$

Reformulation

Maximiser $\text{Var}\{\langle y_i, e_j \rangle\}_{i=1..n}$ avec $e_j \in \mathcal{C}([a, b])$, $j = 1..n$

sous contraintes d'orthonormalité : $\forall j \neq k, \langle e_j, e_k \rangle = 0$.

Calcul

Deux axes de résolution (Ramsay & Silverman 2005) :

- ① discrétisation, ACP sur matrice $n \times D$ puis lissage éventuel ;
- ② décomposition sur une base, ACP sur coefficients ($n \times D$) puis reconstruction.

Calcul

Deux axes de résolution (Ramsay & Silverman 2005) :

- 1 discrétisation, ACP sur matrice $n \times D$ puis lissage éventuel ;
- 2 décomposition sur une base, ACP sur coefficients ($n \times D$) puis reconstruction.

Courbes discrétisées en lignes : $A = \begin{pmatrix} y_{11} & \cdots & y_{1D} \\ \vdots & \vdots & \vdots \\ y_{n1} & \cdots & y_{nD} \end{pmatrix}$

Calcul

Deux axes de résolution (Ramsay & Silverman 2005) :

- 1 discrétisation, ACP sur matrice $n \times D$ puis lissage éventuel ;
- 2 décomposition sur une base, ACP sur coefficients ($n \times D$) puis reconstruction.

Courbes discrétisées en lignes : $A = \begin{pmatrix} y_{11} & \cdots & y_{1D} \\ \vdots & \vdots & \vdots \\ y_{n1} & \cdots & y_{nD} \end{pmatrix}$

Fonctions $e_j =$ vecteurs propres de tAA

Calcul

Deux axes de résolution (Ramsay & Silverman 2005) :

- 1 discrétisation, ACP sur matrice $n \times D$ puis lissage éventuel ;
- 2 décomposition sur une base, ACP sur coefficients ($n \times D$) puis reconstruction.

Courbes discrétisées en lignes :
$$A = \begin{pmatrix} y_{11} & \cdots & y_{1D} \\ \vdots & \vdots & \vdots \\ y_{n1} & \cdots & y_{nD} \end{pmatrix}$$

Fonctions $e_j =$ vecteurs propres de tAA

Décomposition en valeurs singulières : $A = U\Sigma{}^tV$

Calcul

Deux axes de résolution (Ramsay & Silverman 2005) :

- 1 discrétisation, ACP sur matrice $n \times D$ puis lissage éventuel ;
- 2 décomposition sur une base, ACP sur coefficients ($n \times D$) puis reconstruction.

$$\text{Courbes discrétisées en lignes : } A = \begin{pmatrix} y_{11} & \cdots & y_{1D} \\ \vdots & \vdots & \vdots \\ y_{n1} & \cdots & y_{nD} \end{pmatrix}$$

Fonctions $e_j =$ vecteurs propres de tAA

Décomposition en valeurs singulières : $A = U\Sigma^tV$

Propriété

Les $\min(n, D)$ premiers vecteurs propres coïncident avec les premières colonnes de V

Exemple

Swissroll 3D :

$$(t \cos t, t \sin t, u), t \in [0, 2\pi]$$

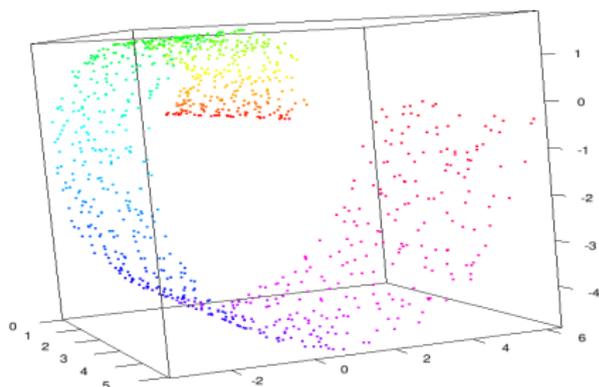


FIG.: Swissroll 3D

Exemple

Swissroll 3D :

$$(t \cos t, t \sin t, u), t \in [0, 2\pi]$$

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

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

$$(\alpha, \beta, \gamma) \in \text{swissroll}$$

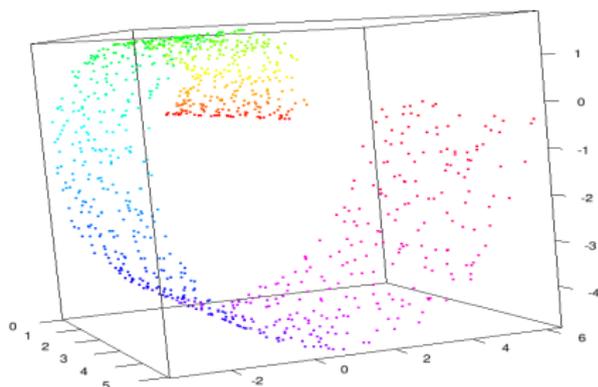


FIG.: Swissroll 3D

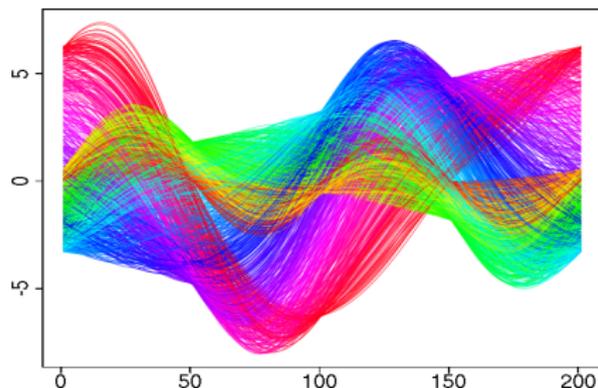


FIG.: Swissroll fonctionnel

Exemple

Swissroll 3D :

$$(t \cos t, t \sin t, u), t \in [0, 2\pi]$$

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

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

$$(\alpha, \beta, \gamma) \in \text{swissroll}$$

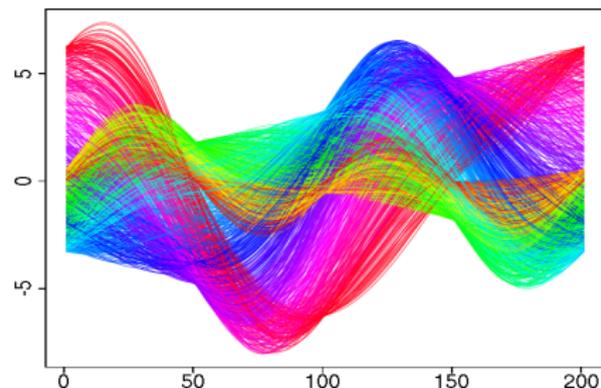


FIG.: Swissroll fonctionnel

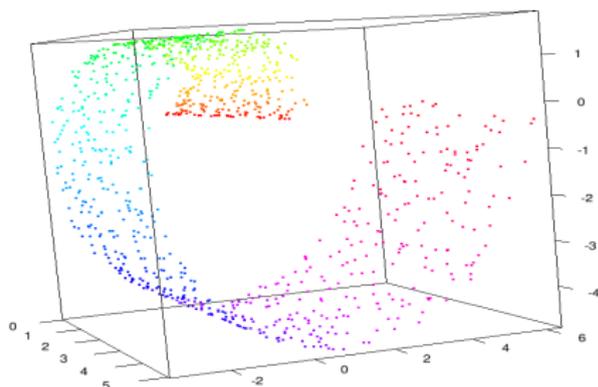


FIG.: Swissroll 3D

Paramétrisation \Rightarrow dimension 2

Variance expliquée avec..

- 1 composante = 45%
- 2 composantes = 80%
- ≥ 3 composantes = 100%

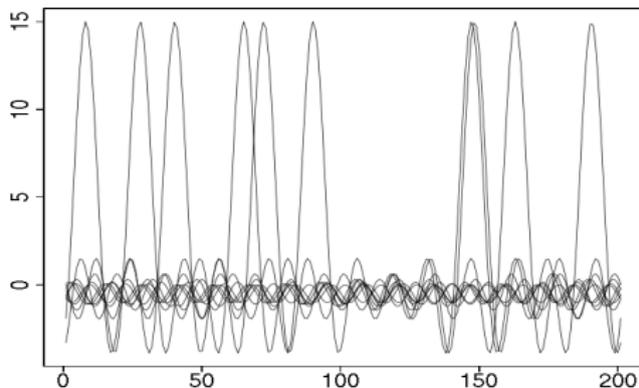
Limitations

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

= paramètres des n courbes correspondantes

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

Tracé de 10 courbes aléatoires f_i
pour $d = 30$



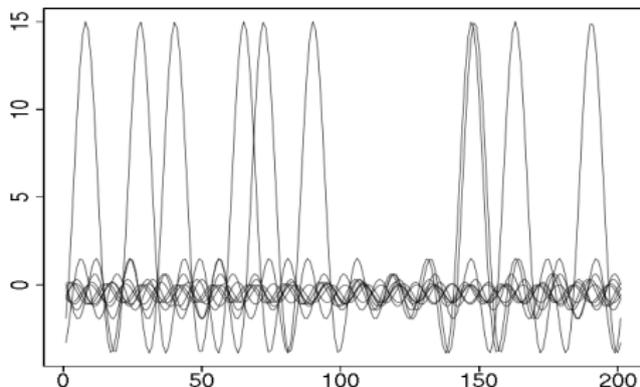
Limitations

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

= paramètres des n courbes correspondantes

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

Tracé de 10 courbes aléatoires f_i
pour $d = 30$



ACP sur matrice des f_i discrétisées $\rightarrow d$ composantes significatives.

Paramètres dépendant de $\cos \mathcal{T}[i]$ et $\sin \mathcal{T}[i]$ seulement

\Rightarrow ("infinitement") sous-optimal.

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

Étape préliminaire

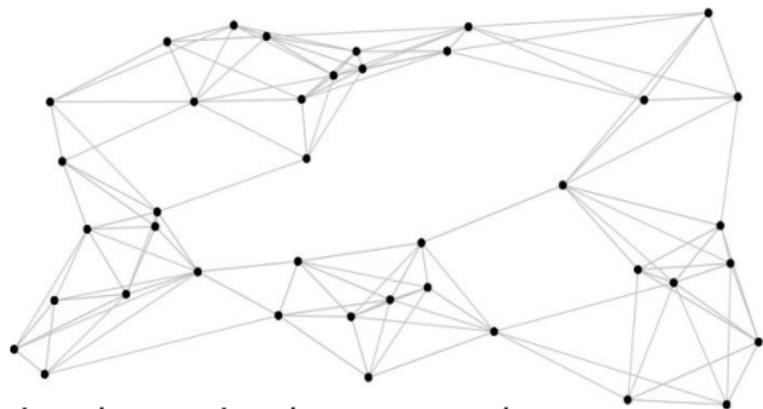


FIG.: Représentation des sorties dans un graphe ;
sommets = données, arêtes = distances.

y_j se connecte à y_i 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

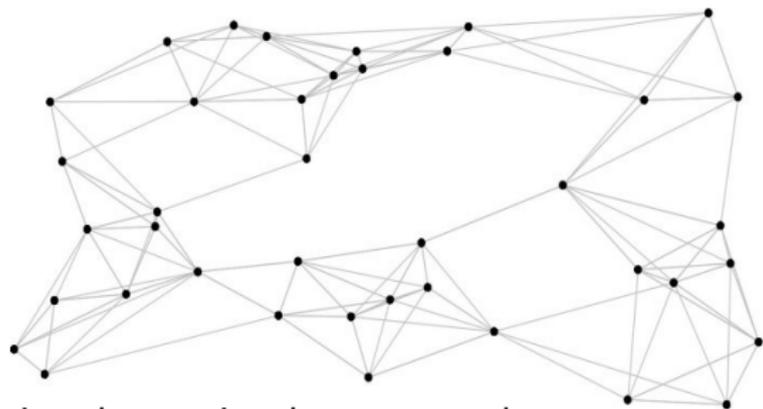


FIG.: Représentation des sorties dans un graphe ;
sommets = données, arêtes = distances.

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

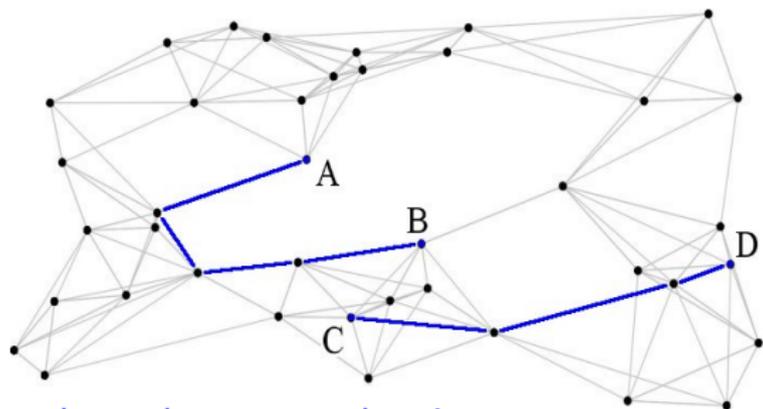


FIG.: En bleu : deux plus courts chemins ;

y_j se connecte à y_i 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)

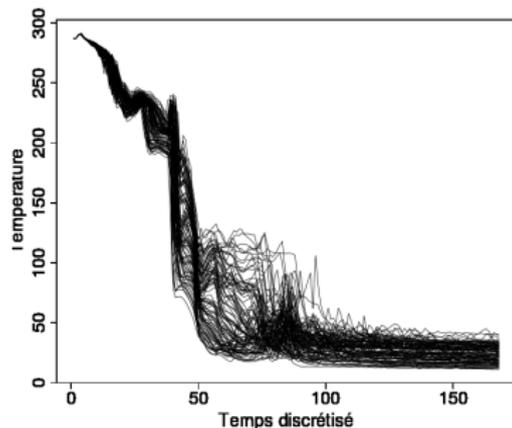


FIG.: 100 courbes CATHARE I

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**)

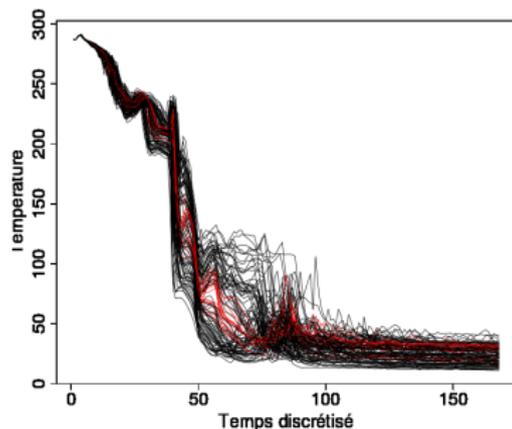


FIG.: 15 courbes au voisinage de y_0

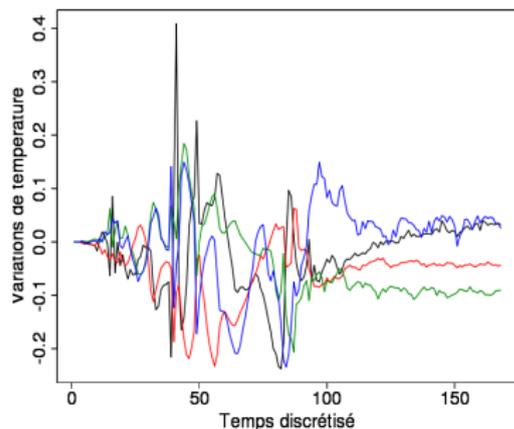


FIG.: Base locale en y_0 ($d = 4$)

Étape locale II

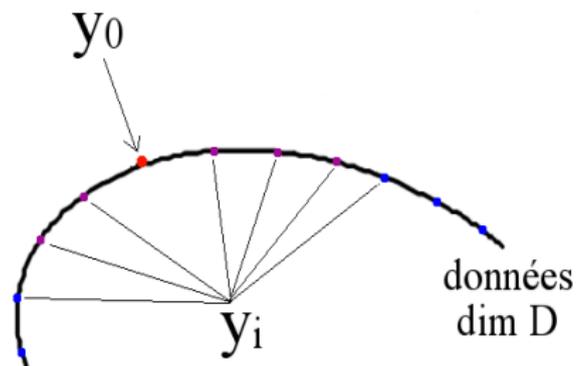


FIG.: Courbe origine y_0

Étape locale II

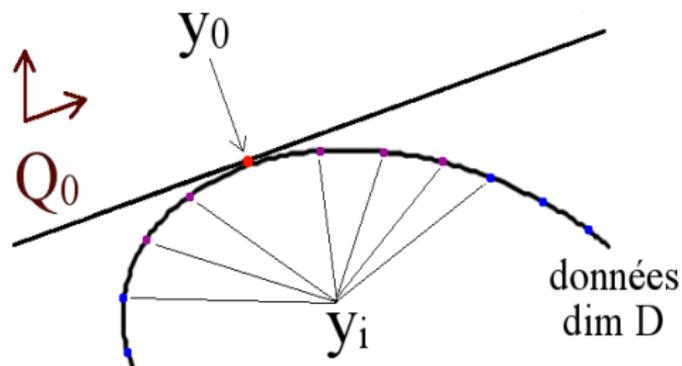


FIG.: Espace tangent en y_0 + base locale Q_0

Étape locale II

Coordonnées réduites z_i calculées par projection sur Q_0 ,

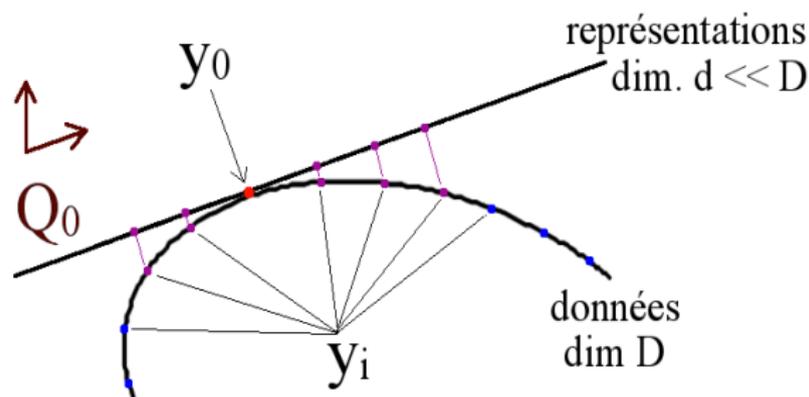


FIG.: Coordonnées locales z_i sur l'espace tangent

Étape locale II

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

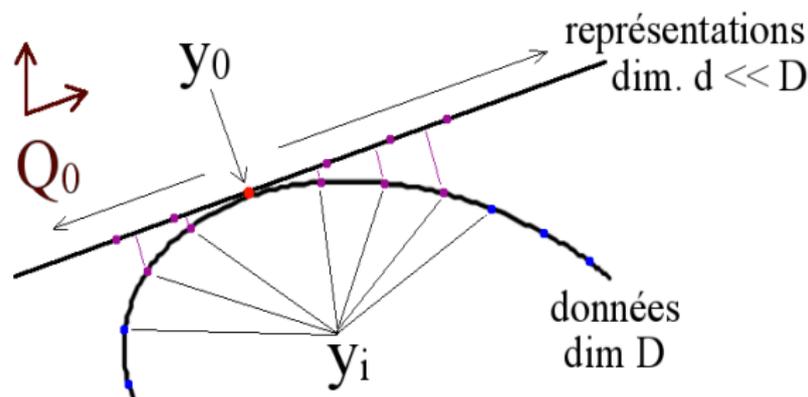


FIG.: Normalisation des coordonnées z_i

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

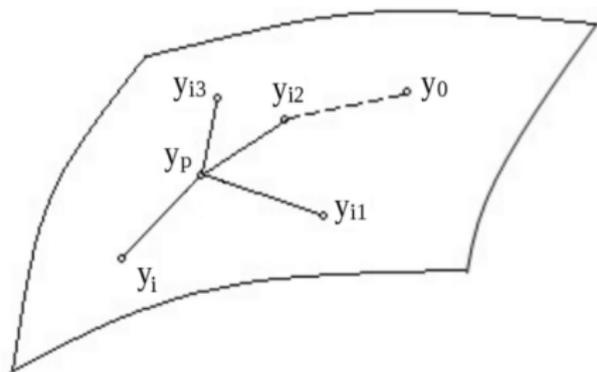


FIG.: Données y_i en dim. D

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 $\|y_i - y_p\| = \|z_i - z_p\|$.

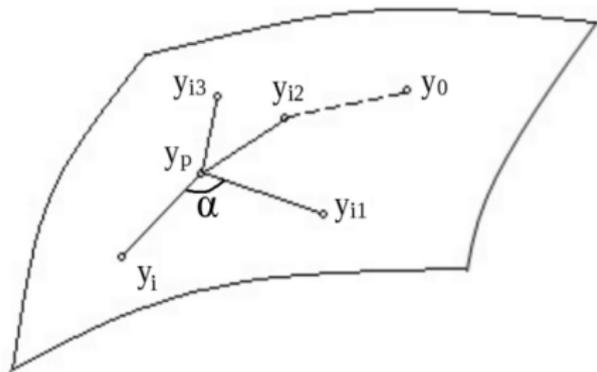


FIG.: Données y_i en dim. D

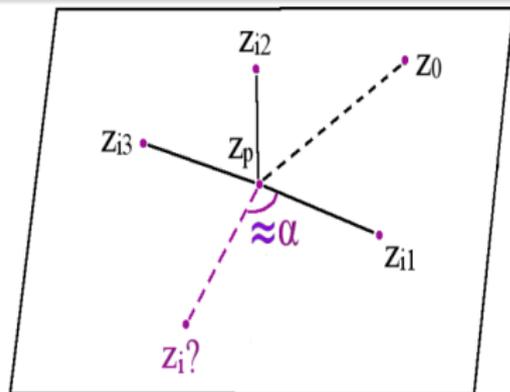


FIG.: $z_i = r(y_i)$ en dim. $d \ll D$

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

Méthode

$z \in \mathbb{R}^d$; courbe correspondante $\in \mathcal{C}([a, b])$?

Appliquer l'algorithme "à l'envers" :

Méthode

$z \in \mathbb{R}^d$; courbe correspondante $\in \mathcal{C}([a, b])$?

Appliquer l'algorithme "à l'envers" :

- 1 $PPV(z)$ = plus proches voisins de z dans $\{z_i\}_{i=1..n}$;
- 2 si les représentations des voisins de y_0 (origine) sont majoritaires, alors z est la représentation sur Q_0 de la fonction cherchée ;

Méthode

$z \in \mathbb{R}^d$; courbe correspondante $\in \mathcal{C}([a, b])$?

Appliquer l'algorithme "à l'envers" :

- 1 $PPV(z) =$ plus proches voisins de z dans $\{z_i\}_{i=1..n}$;
- 2 si les représentations des voisins de y_0 (origine) sont majoritaires, alors z est la représentation sur Q_0 de la fonction cherchée ;
- 3 sinon :

Méthode

$z \in \mathbb{R}^d$; courbe correspondante $\in \mathcal{C}([a, b])$?

Appliquer l'algorithme "à l'envers" :

- 1 $PPV(z)$ = plus proches voisins de z dans $\{z_i\}_{i=1..n}$;
- 2 si les représentations des voisins de y_0 (origine) sont majoritaires, alors z est la représentation sur Q_0 de la fonction cherchée ;
- 3 sinon :

Inversion de l'étape "y loin de y_0 "

- 1 ACP locale au voisinage de y_α avec $\alpha = \arg \min_{i=1..n} \|z - z_i\|$;
- 2 remplacer les $PPV(y_\alpha)$ par leurs projections d -dimensionnelles sur la base ACP, notées $w_i, i = 1..k$;

Méthode

$z \in \mathbb{R}^d$; courbe correspondante $\in \mathcal{C}([a, b])$?

Appliquer l'algorithme "à l'envers" :

- 1 $PPV(z)$ = plus proches voisins de z dans $\{z_i\}_{i=1..n}$;
- 2 si les représentations des voisins de y_0 (origine) sont majoritaires, alors z est la représentation sur Q_0 de la fonction cherchée ;
- 3 sinon :

Inversion de l'étape "y loin de y_0 "

- 1 ACP locale au voisinage de y_α avec $\alpha = \arg \min_{i=1..n} \|z - z_i\|$;
- 2 remplacer les $PPV(y_\alpha)$ par leurs projections d -dimensionnelles sur la base ACP, notées $w_i, i = 1..k$;
- 3 appliquer l'algorithme précédent avec $y_{(i)} \leftarrow z_{(i)}$ et $z_{(i)} \leftarrow w_{(i)}$.

Exemple 1

Swissroll fonctionnel :

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

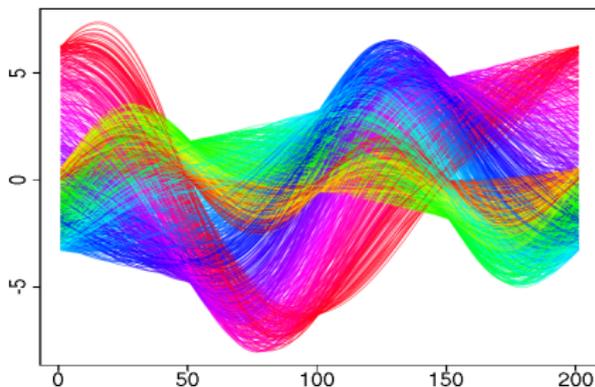


FIG.: Swissroll fonctionnel

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

TAB.: MSE moyenne en fonction
de d

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

TAB.: MSE moyenne en fonction de d

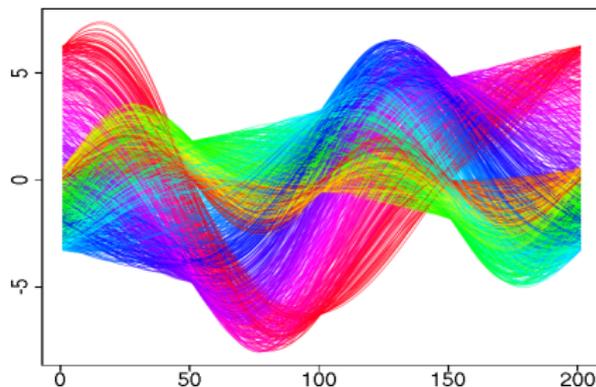


FIG.: Swissroll fonctionnel

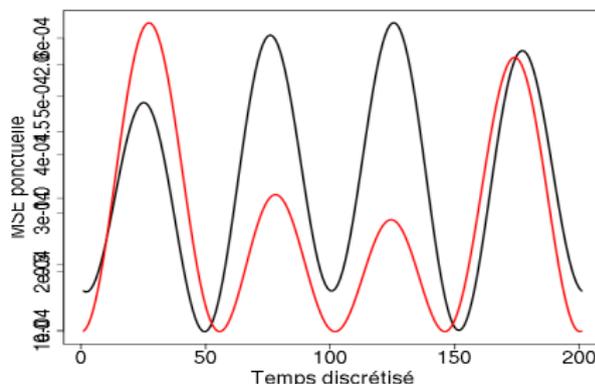


FIG.: MSE RML (noir) et ACP (rouge) pour $d = 2$

Exemple II

\mathcal{T} = paramètres des n courbes (aléatoires dans $[0, 2\pi]$).

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

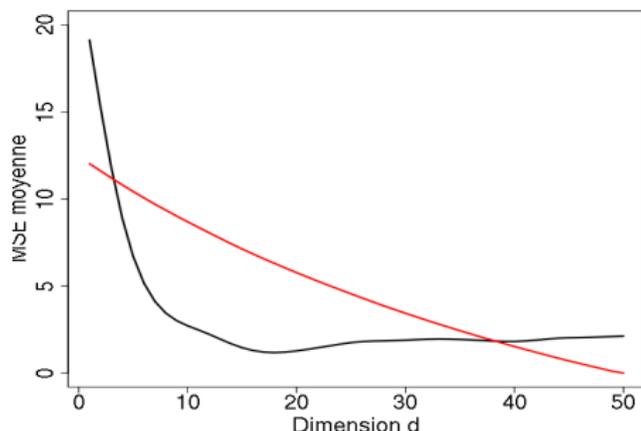


FIG.: MSE RML (noir) et ACP (rouge) en fonction de d

Points de croisement :

- 1 entre $d = 2$ et $d = 3$;
- 2 entre $d = 38$ et $d = 39$.

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

Motivations

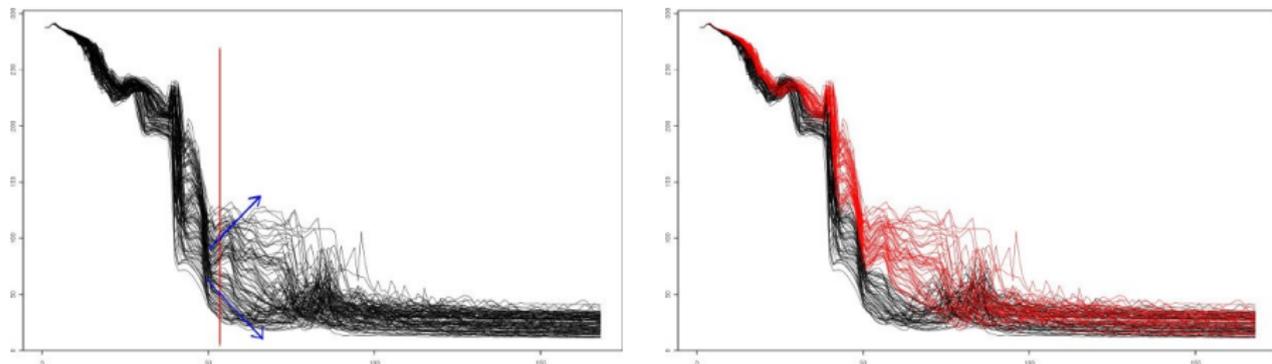


FIG.: 100 transitoires de température

→ Différents types de comportements physiques.

Regroupement des courbes aux caractéristiques similaires

⇒ meilleure modélisation dans chaque cluster.

Motivations

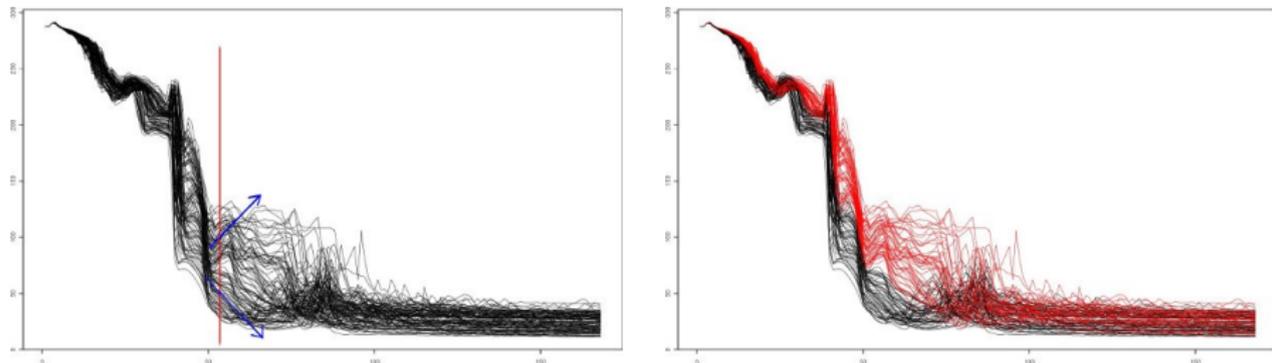


FIG.: 100 transitoires de température

→ Différents types de comportements physiques.

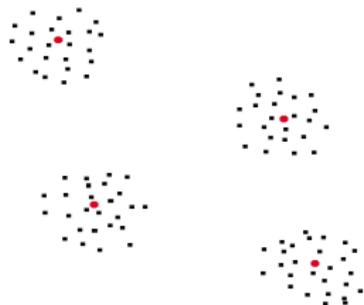
Regroupement des courbes aux caractéristiques similaires

⇒ meilleure modélisation dans chaque cluster.

Classification des sorties *puis* des entrées.

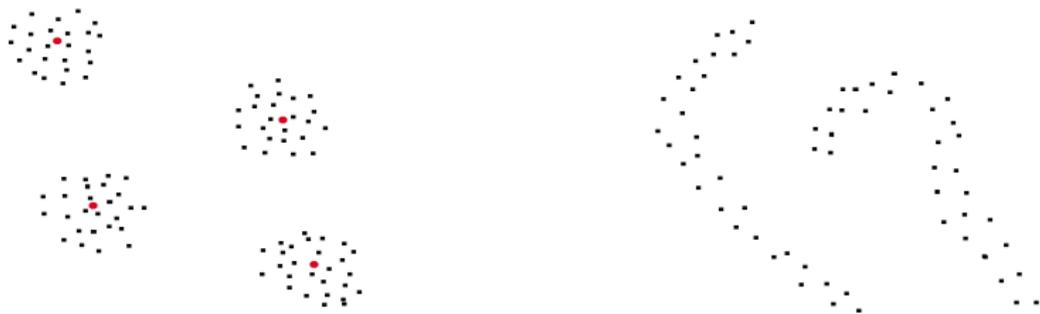
Types de méthodes

- Regroupement autour de k centres : (variantes du) k -means, ...



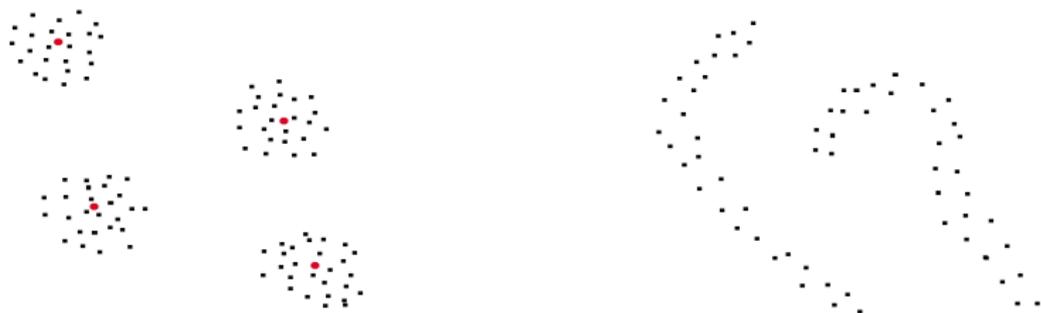
Types de méthodes

- Regroupement autour de k centres : (variantes du) k -means, ...
- Regroupement sans prototypes de clusters ;
 - ▶ algorithmes **dans un graphe** : clustering spectral, coloriage, ...
 - ▶ sans graphe support : **hiérarchique**, évolutionnaire, ...



Types de méthodes

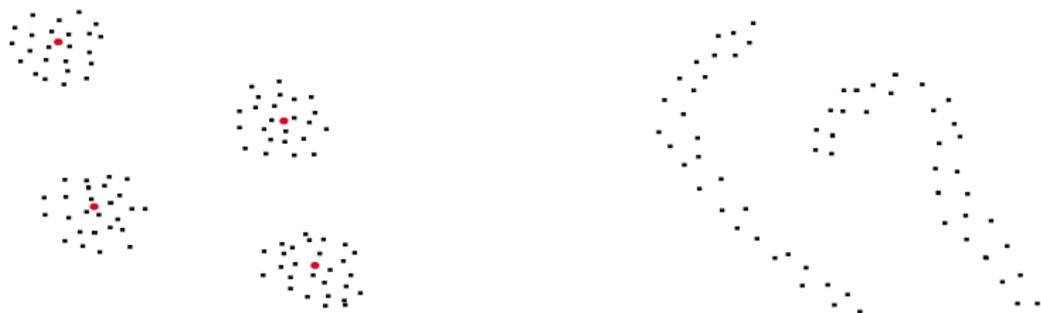
- Regroupement autour de k centres : (variantes du) k -means, ...
- Regroupement sans prototypes de clusters ;
 - ▶ algorithmes dans un graphe : clustering spectral, coloriage, ...
 - ▶ sans graphe support : hiérarchique, évolutionnaire, ...



Choix : CAH avec distances calculées dans un graphe.

Types de méthodes

- Regroupement autour de k centres : (variantes du) k -means, ...
- Regroupement sans prototypes de clusters ;
 - ▶ algorithmes **dans un graphe** : clustering spectral, coloriage, ...
 - ▶ sans graphe support : **hiérarchique**, évolutionnaire, ...



Choix : CAH avec distances calculées dans un graphe.

Courbes : possible prise en compte du décalage "temporel", et/ou de la continuité.

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

Fonctionnement

Algorithme déjà conçu pour une matrice de distances.

Principe : partant de n clusters, les fusionner petit à petit en fonction des distances relatives.

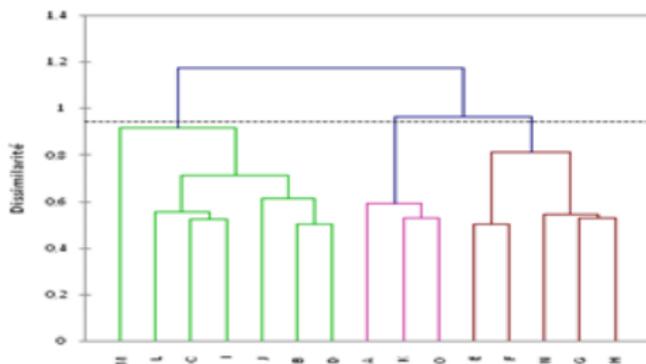


FIG.: Dendrogramme résultat d'un clustering hiérarchique.

Fonctionnement

Algorithme déjà conçu pour une matrice de distances.

Principe : partant de n clusters, les fusionner petit à petit en fonction des distances relatives.

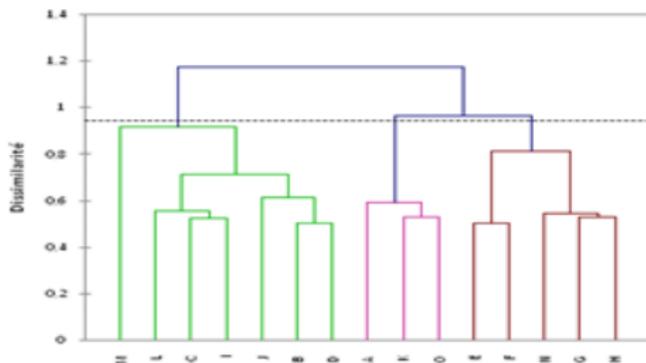


FIG.: Dendrogramme résultat d'un clustering hiérarchique.

Résultat : suite de partitions emboîtées

$(A_i)_{i=1..n}$ avec $A_{i+1} \prec A_i$; A_k contient k groupes.

Comparaisons inter-clusters

- 1 Définition d'une distance d dans \mathbb{R}^p ou $\mathcal{C}([a, b])$;
- 2 Définition d'une distance entre les clusters à partir de d .

Comparaisons inter-clusters

- 1 Définition d'une distance d dans \mathbb{R}^p ou $\mathcal{C}([a, b])$;
- 2 Définition d'une distance entre les clusters à partir de d .

1 : distance L_2 pour l'instant.

2 : distance de Ward, définie par

$$d_W(A, B) = \sqrt{\frac{|A||B|}{(|A| + |B|)}} d(g_A, g_B)$$

→ minimise la variation d'inertie

⇒ favorise les partitions "stables"

Exemple CH + distance de Ward

240 courbes dans 3 clusters :

$$\beta_1 \sim \mathcal{U}(1, 2), \beta_2 \sim \mathcal{U}(0, 1), \beta_3 \sim \mathcal{U}(0.4, 1); \alpha_{1,2,3} = 3\beta_{1,2,3}$$

$$\gamma_1 = \sqrt{4 - \beta_1^2}, \gamma_2 = \sqrt{1 - \beta_2^2}, \gamma_3 = 3\sqrt{1 - \beta_3^2} + 3$$

$$f_{\alpha,\beta,\gamma}(t) = \left(\frac{\sin \alpha t}{t} + e^{-\beta t}\right) \cos \gamma t$$

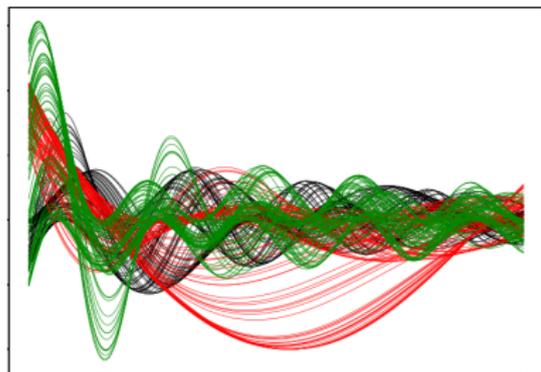


FIG.: Vraie partition

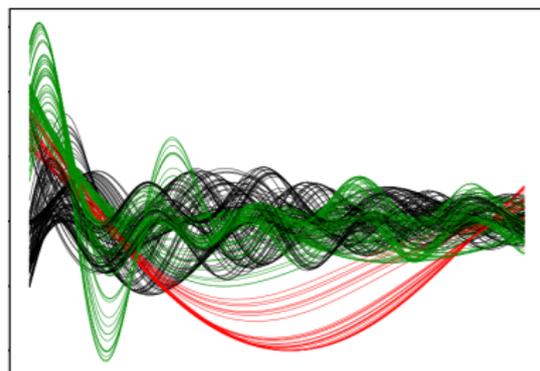


FIG.: Partition obtenue

Précision : 0.47

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- **Distance "Commute-Time"**
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

(Entrées-)Sorties dans un graphe

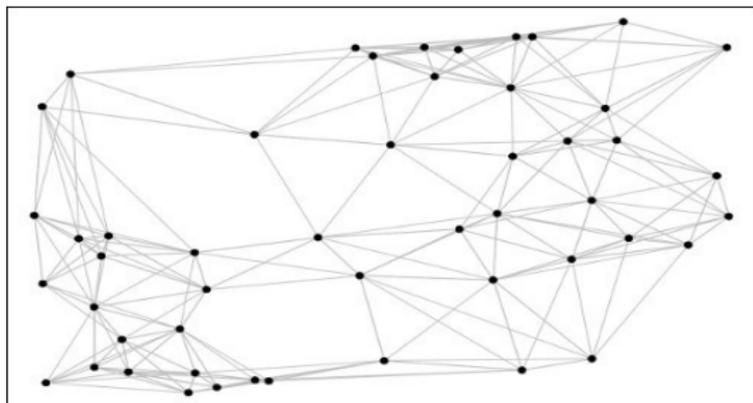


FIG.: Représentation des (entrées-)sorties dans un graphe ;
sommets = données, arêtes = distances.

(Entrées-)Sorties dans un graphe

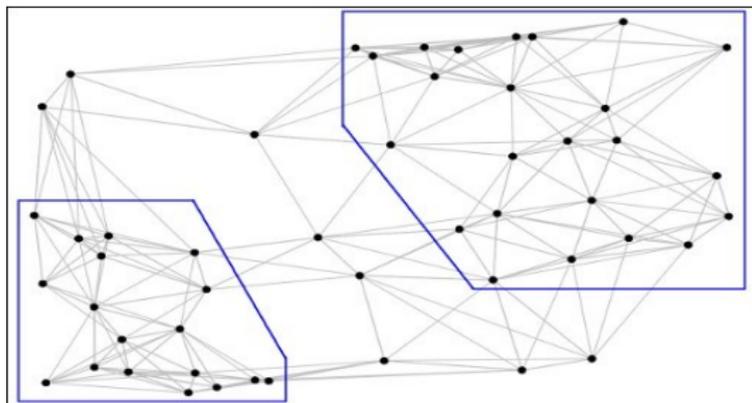


FIG.: En bleu : deux clusters (visuels) ;

(Entrées-)Sorties dans un graphe

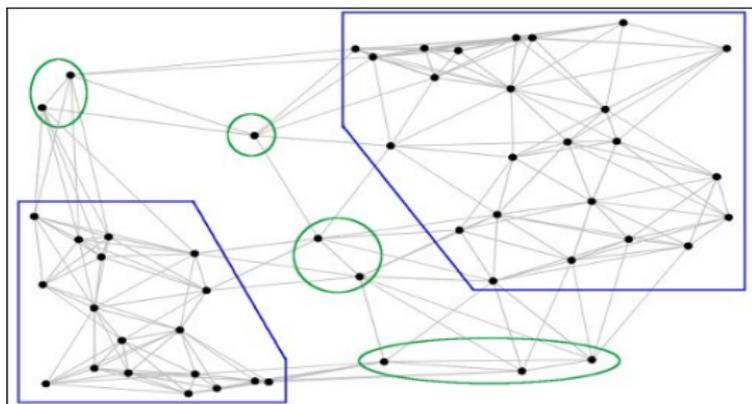


FIG.: En vert : points "isolés" ;

(Entrées-)Sorties dans un graphe

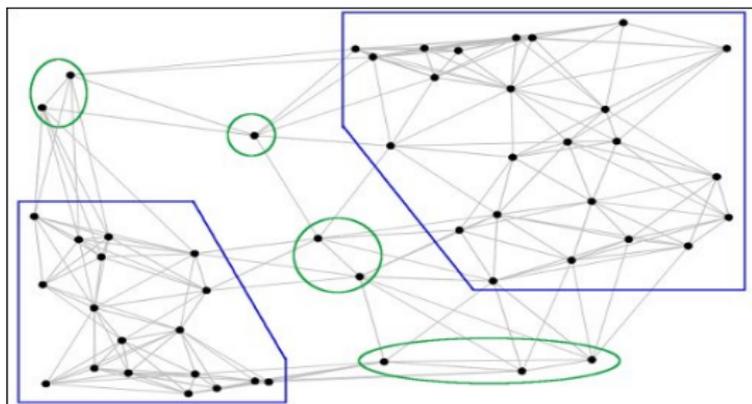


FIG.: En vert : points "isolés" ;

Définition : $d(i, j)$ = temps moyen d'aller-retour entre i et j

(Entrées-)Sorties dans un graphe

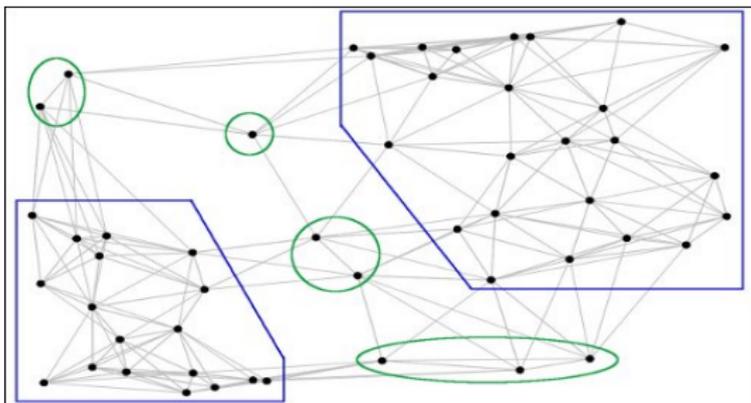


FIG.: En vert : points "isolés" ;

Définition : $d(i, j)$ = temps moyen d'aller-retour entre i et j

Avantages/inconvénients

- + sépare mieux les clusters, rendant la classification plus facile
- nécessite le choix du nombre de voisins en chaque point

Parcours sur le graphe

"Il y a beaucoup de courts chemins de A à B "

= "Une marche aléatoire issue de A reviendra vite vers A ".

Parcours sur le graphe

"Il y a beaucoup de courts chemins de A à B "

= "Une marche aléatoire issue de A reviendra vite vers A ".

Marche aléatoire dans le graphe de voisinage

Transition de y_i vers y_j avec probabilité

$$\mathbb{P}(i \rightarrow j) = \begin{cases} \alpha e^{-\frac{\|y_i - y_j\|^2}{\sigma_i^2}} & \text{si } y_j \text{ voisin de } y_i, \\ 0 & \text{sinon.} \end{cases}$$

σ_i : paramètre local dépendant de k et ...

Expression de la distance CT

Principal résultat (M. Saerens et al. 2004)

Temps moyen d'aller-retour entre A et B

=

carré de la distance euclidienne entre A et B dans un espace de représentation à N dimensions.

Expression de la distance CT

Principal résultat (M. Saerens et al. 2004)

Temps moyen d'aller-retour entre A et B
=
carré de la distance euclidienne entre A et B dans un espace de représentation à N dimensions.

Calcul facile :

- P = matrice de transition, $P_{ij} = \mathbb{P}(i \rightarrow j)$;
- L laplacien du graphe : $L = I - P$.

$$d_{CT}^2(y_i, y_j) \propto L_{ii}^+ + L_{jj}^+ - 2L_{ij}^+,$$

avec L^+ pseudo-inverse de L .

Exemple CH + distance ECT

240 courbes dans 3 clusters :

$$\beta_1 \sim \mathcal{U}(1, 2), \beta_2 \sim \mathcal{U}(0, 1), \beta_3 \sim \mathcal{U}(0.4, 1); \alpha_{1,2,3} = 3\beta_{1,2,3}$$

$$\gamma_1 = \sqrt{4 - \beta_1^2}, \gamma_2 = \sqrt{1 - \beta_2^2}, \gamma_3 = 3\sqrt{1 - \beta_3^2} + 3$$

$$f_{\alpha,\beta,\gamma}(t) = \left(\frac{\sin \alpha t}{t} + e^{-\beta t}\right) \cos \gamma t$$

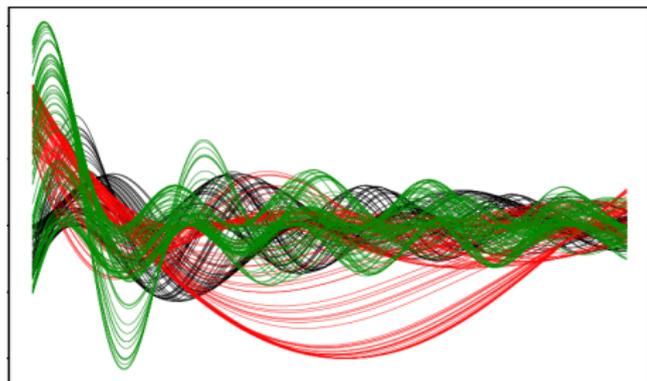


FIG.: Vraie partition = partition obtenue

Précision : 1

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

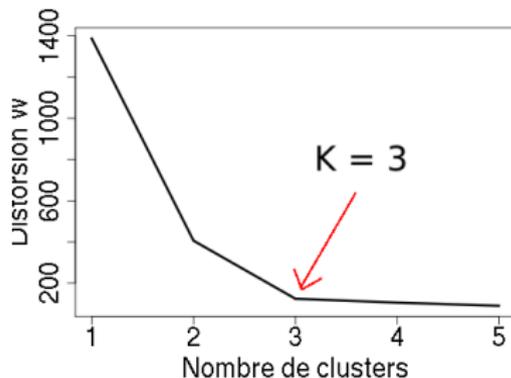
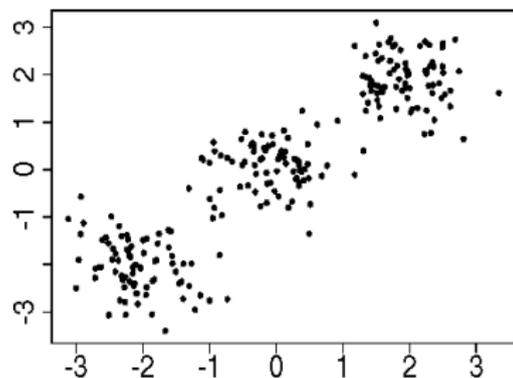
2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

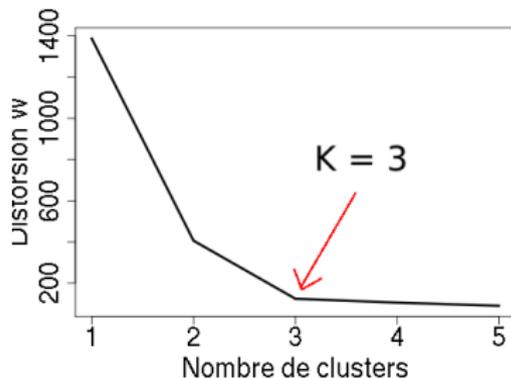
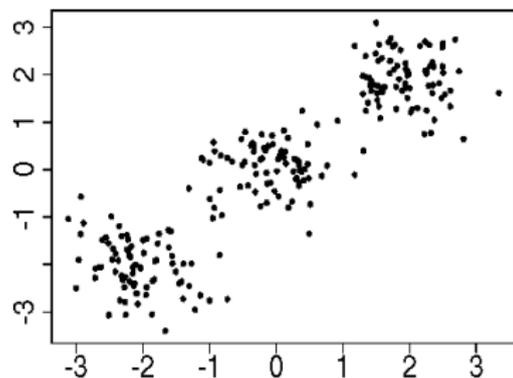
- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

Aperçu des approches existantes



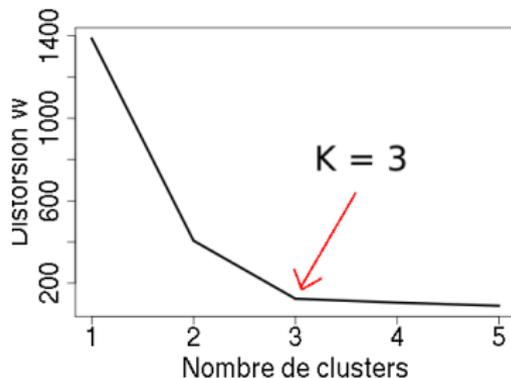
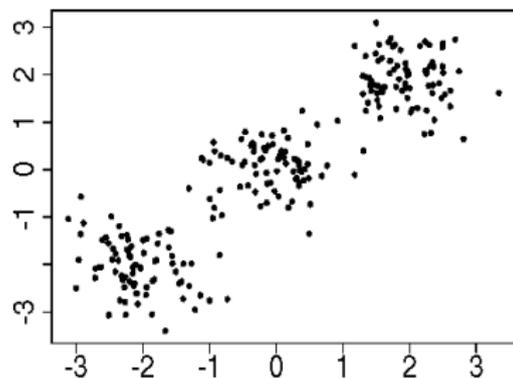
- Méthodes basées sur la structure des données..
 - ▶ imposant une forme aux groupes (diverses formules ...);
 - ▶ ne supposant pas de formes prédéfinies (analyse du spectre du laplacien, kernel k -means + log-vraisemblance ...)

Aperçu des approches existantes



- Méthodes basées sur la structure des données..
 - ▶ imposant une forme aux groupes (diverses formules ...);
 - ▶ ne supposant pas de formes prédéfinies (analyse du spectre du laplacien, kernel k -means + log-vraisemblance ...)
- Méthodes basées sur la stabilité des partitions
 - ▶ sans classification supervisée : *trop général*
 - ▶ avec classification supervisée : permet d'utiliser les entrées

Aperçu des approches existantes

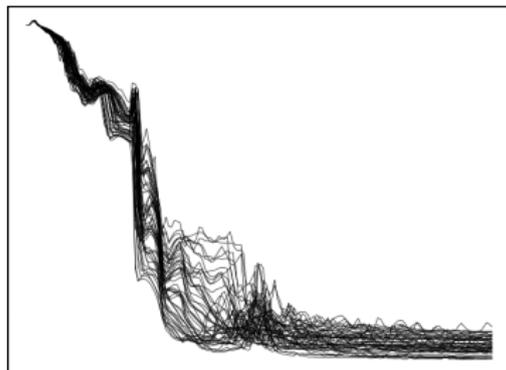
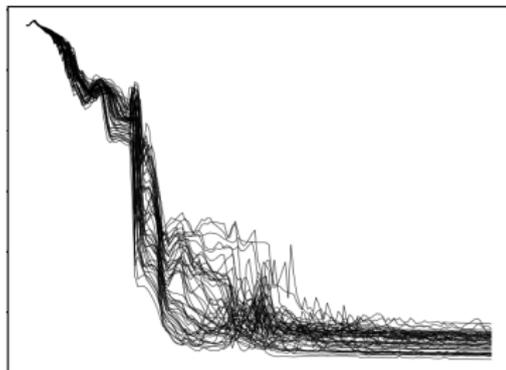


- Méthodes basées sur la structure des données..
 - ▶ imposant une forme aux groupes (diverses formules ...);
 - ▶ ne supposant pas de formes prédéfinies (analyse du spectre du laplacien, kernel k -means + log-vraisemblance ...)
- Méthodes basées sur la stabilité des partitions
 - ▶ sans classification supervisée : *trop général*
 - ▶ avec classification supervisée : permet d'utiliser les entrées

Approche 1 : Roth et al. 2002

Algorithme : N boucles, K augmentant

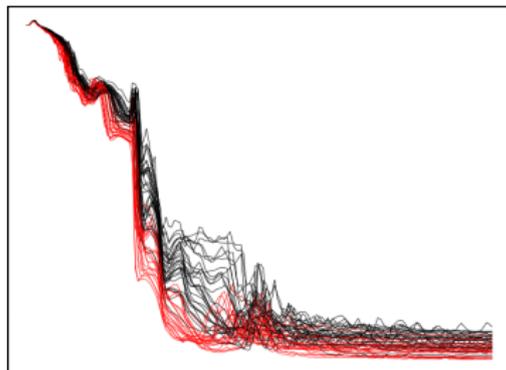
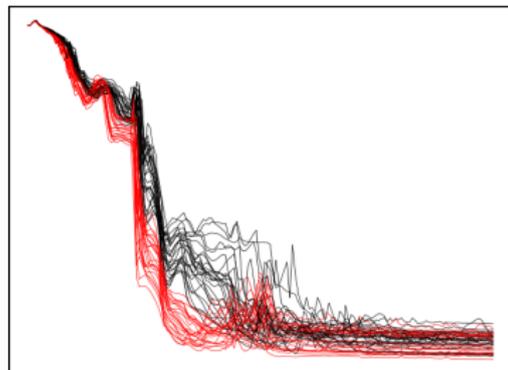
- 1 $\tau \leftarrow 0$;
- 2 division des données en deux parties égales $\mathcal{P}_1, \mathcal{P}_2$;



Approche 1 : Roth et al. 2002

Algorithme : N boucles, K augmentant

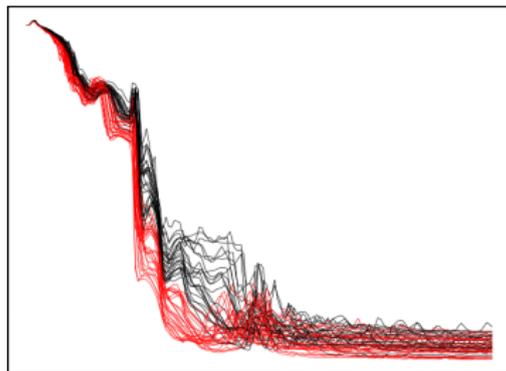
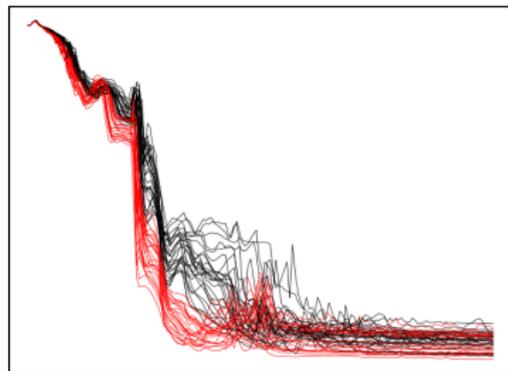
- 1 $\tau \leftarrow 0$;
- 2 division des données en deux parties égales $\mathcal{P}_1, \mathcal{P}_2$;
- 3 clustering de \mathcal{P}_1 et $\mathcal{P}_2 \rightarrow$ partitions P_1, P_2 ;



Approche 1 : Roth et al. 2002

Algorithme : N boucles, K augmentant

- 1 $\tau \leftarrow 0$;
- 2 division des données en deux parties égales $\mathcal{P}_1, \mathcal{P}_2$;
- 3 clustering de \mathcal{P}_1 et $\mathcal{P}_2 \rightarrow$ partitions P_1, P_2 ;
- 4 classification supervisée de \mathcal{P}_1 puis prédiction des labels de $\mathcal{P}_2 \rightarrow$ partition P'_2 ;
- 5 $\tau \leftarrow \tau + \frac{2}{n} \sum_{i=1..n/2} \mathbb{1}_{P_2[i]=P'_2[i]}$;



Approche 2 : basée sur Dudoit & Fridlyand 2002

Algorithme : N boucles, K augmentant

- 1 classer les données dans P_{ref} (partition de référence) ;
- 2 sous-échantillonner en deux jeux d'"entraînement" E et E' ;
- 3 clustering \rightarrow partitions \mathcal{E} et \mathcal{E}' , s'intersectant en I et I' ;
si (en moyenne) trop différentes sur l'intersection, **stop** ;

Approche 2 : basée sur Dudoit & Fridlyand 2002

Algorithme : N boucles, K augmentant

- 1 classer les données dans P_{ref} (partition de référence) ;
- 2 sous-échantillonner en deux jeux d'"entraînement" E et E' ;
- 3 clustering \rightarrow partitions \mathcal{E} et \mathcal{E}' , s'intersectant en I et I' ;
si (en moyenne) trop différentes sur l'intersection, **stop** ;
- 4 diviser les données en jeu d'entraînement E et de test T ;
 $P =$ étiquettes de T dans P_{ref} ;
- 5 classification supervisée $E \mapsto$ labels des groupes, puis prédiction
des étiquettes de $T = P'$;
- 6 si (en moyenne) P trop différente de P' , **stop** ;

Retourner la dernière partition P_{ref} obtenue.

Clustering des entrées

Étape 1 : partitionnement des courbes.

- 1 partition des n courbes en sortie = P_0
(méthode précédente)
- 2 nombre de clusters = K_0

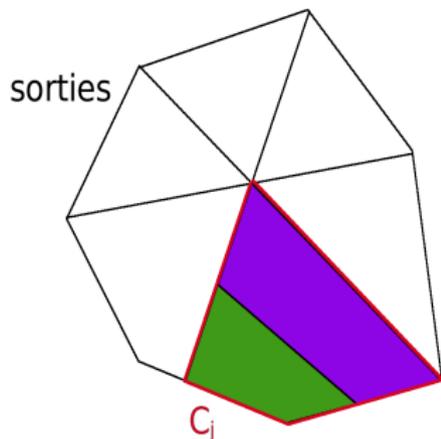


FIG.: Partition P_0 (sorties).

Clustering des entrées

Étape 2 : pour chaque cluster C_j du regroupement P_0 ,

- 1 partition des n_j vecteurs en entrée = P_j
(méthode précédente)
- 2 division de C_j en α clusters ; $K_0 \leftarrow K_0 + \alpha - 1$

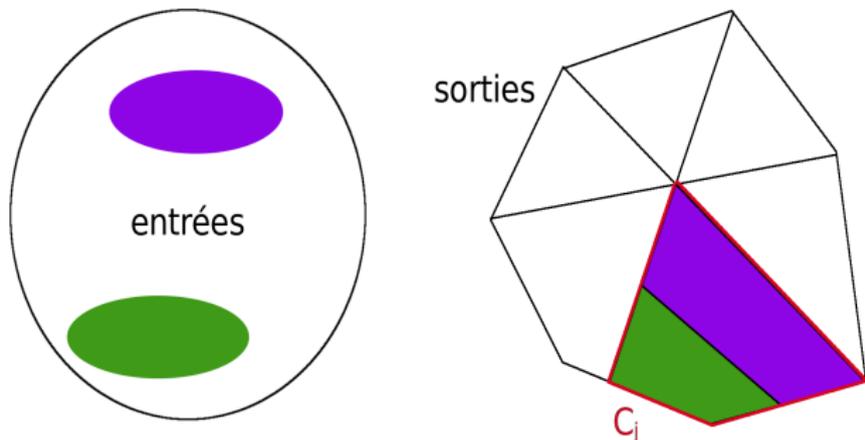


FIG.: Cluster C_j divisé en deux (entrées) et partition P_0 (sorties).

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- **Méthodologie**
- Jeu de données CATHARE I
- Jeu de données CATHARE II

Chronologie

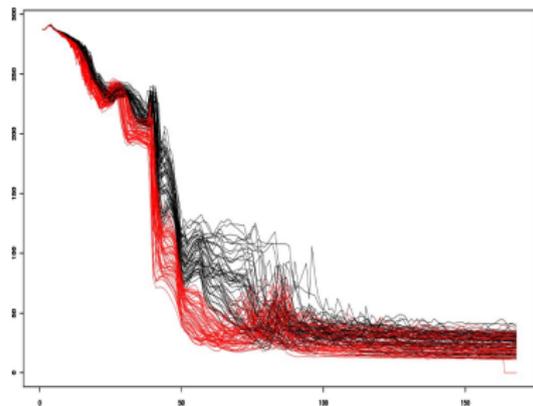


FIG.: 100 transitoires de température en sortie (cathare)

- 1 Classification non supervisée des N courbes y_i en k clusters C_j

Chronologie

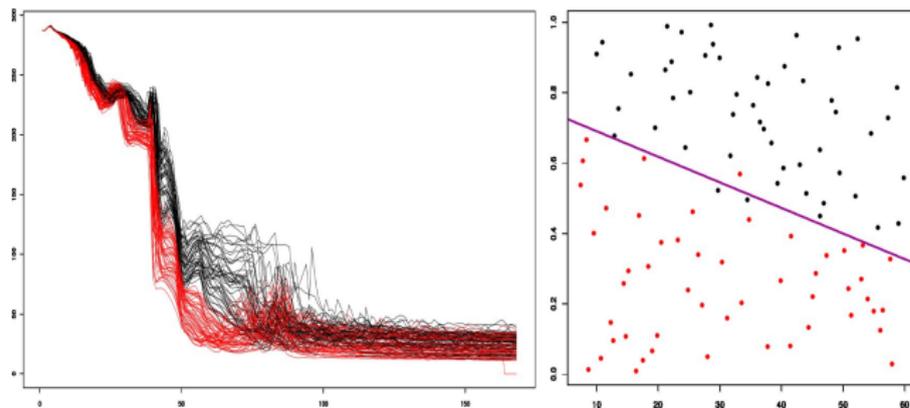


FIG.: g. à d. : sorties fonctionnelles, scatterplot entrées 1 - 4

- 1 Classification non supervisée des N courbes y_i en k clusters C_j
+ classification supervisée des entrées x_i .

Chronologie

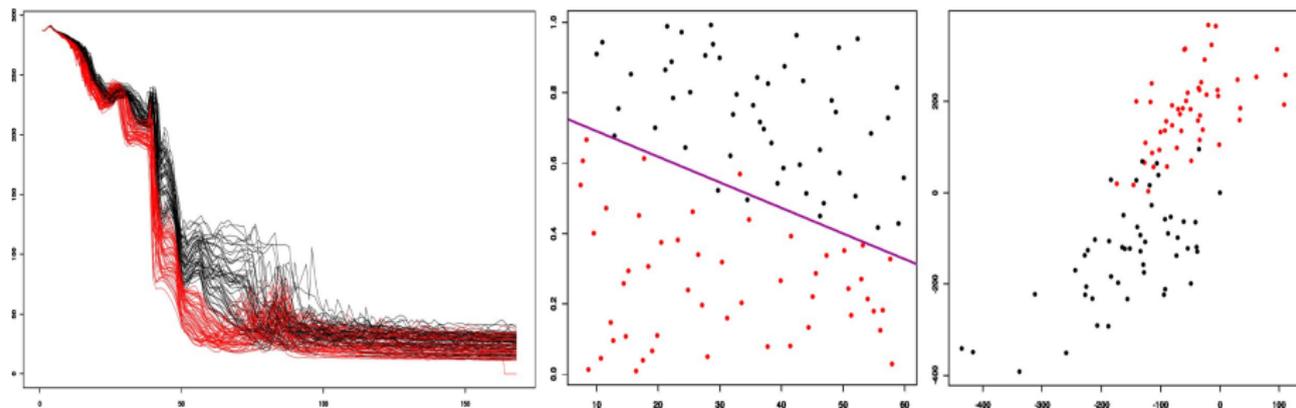


FIG.: g. à d. : sorties fonc., entrées 1 - 4, représentation 2D des sorties

- 1 Classification non supervisée des N courbes y_i en k clusters C_j
+ classification supervisée des entrées x_i .
- 2 Pour chaque cluster C_j ,
 - 1 réduction de la dimension : $r(y_i) = z_i$ représente y_i dans \mathbb{R}^d ;

Chronologie

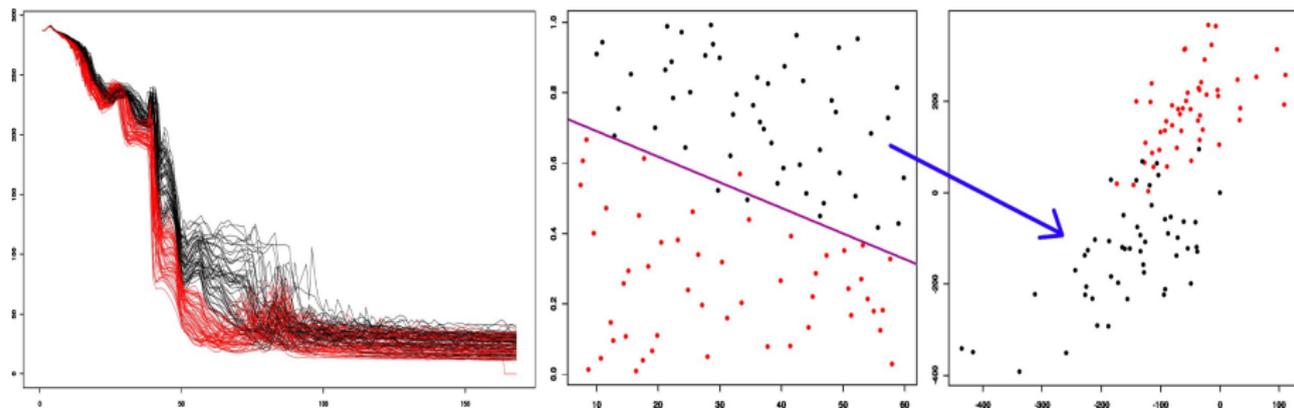


FIG.: g. à d. : sorties fonc., entrées 1 - 4, rep. 2D des sorties

- 1 Classification non supervisée des N courbes y_i en k clusters C_j
+ classification supervisée des entrées x_i .
- 2 Pour chaque cluster C_j ,
 - 1 réduction de la dimension : $r(y_i) = z_i$ représente y_i dans \mathbb{R}^d ;
 - 2 apprentissage d'une fonction de régression : $f(x_i) \simeq z_i$;

Chronologie

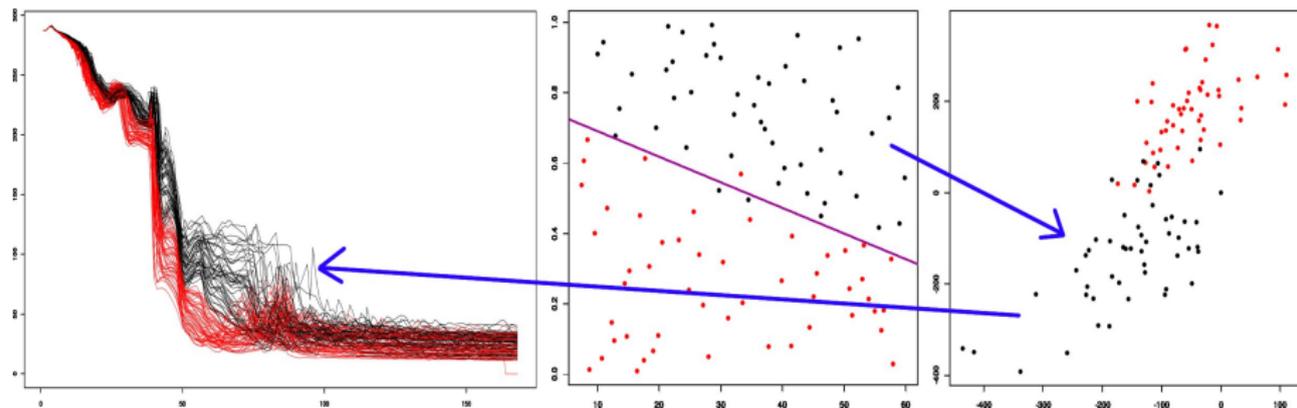


FIG.: g. à d. : sorties func., entrées 1 - 4, rep. 2D des sorties

- 1 Classification non supervisée des N courbes y_i en k clusters C_j
+ classification supervisée des entrées x_i .
- 2 Pour chaque cluster C_j ,
 - 1 réduction de la dimension : $r(y_i) = z_i$ représente y_i dans \mathbb{R}^d ;
 - 2 apprentissage d'une fonction de régression : $f(x_i) \simeq z_i$;
 - 3 apprentissage d'une fonction de reconstruction : $g(z_i) \simeq y_i$.

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.

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- **Jeu de données CATHARE I**
- Jeu de données CATHARE II

Test I - courbes de température CATHARE

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

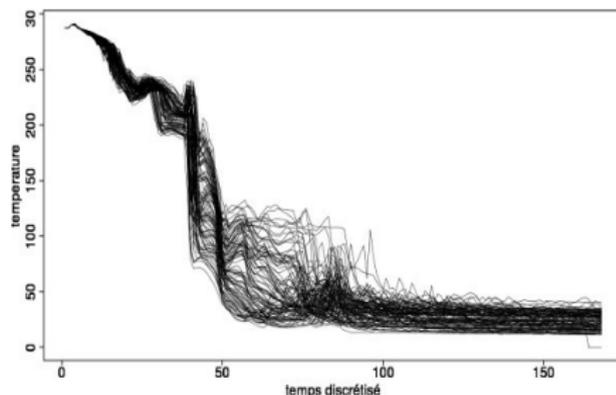


FIG.: Les 100 courbes en sortie

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. ; $d = 4$

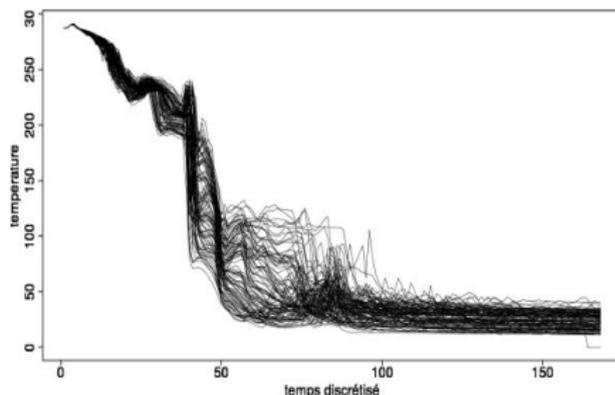


FIG.: Les 100 courbes en sortie

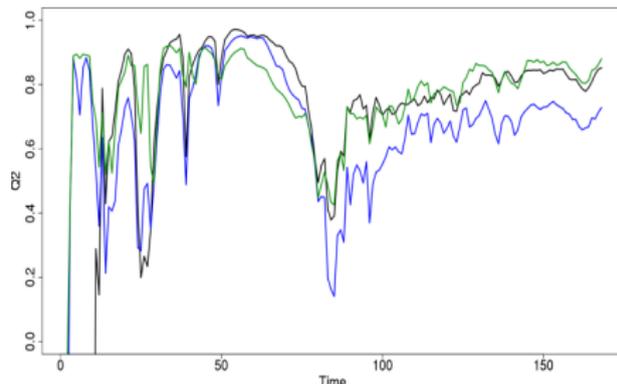
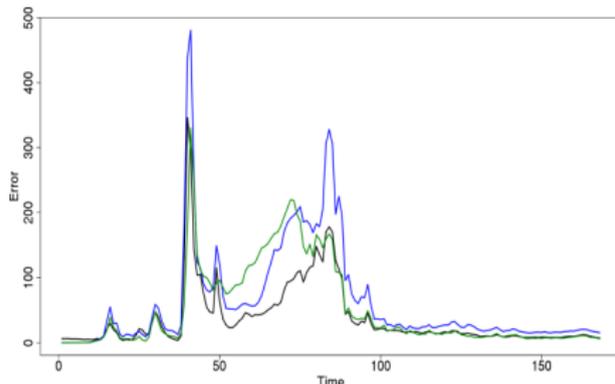


FIG.: Noir : ACP fonctionnelle ; bleu : RML ; vert : Nadaraya-Watson.

5 courbes prédites

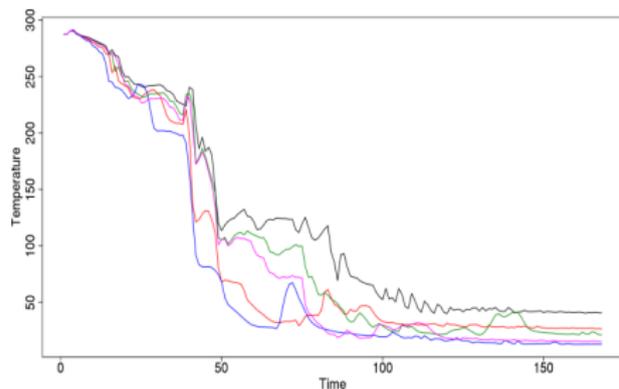


FIG.: Courbes cibles

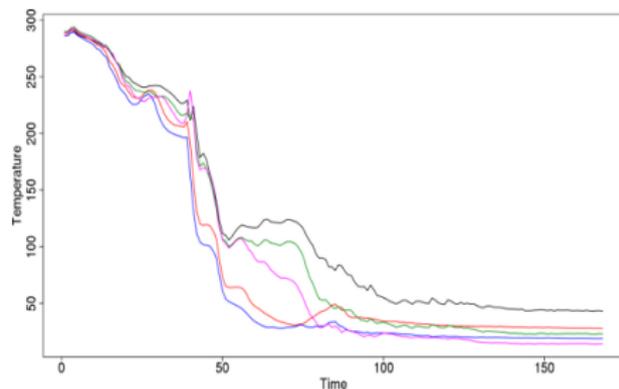


FIG.: ACP fonc.

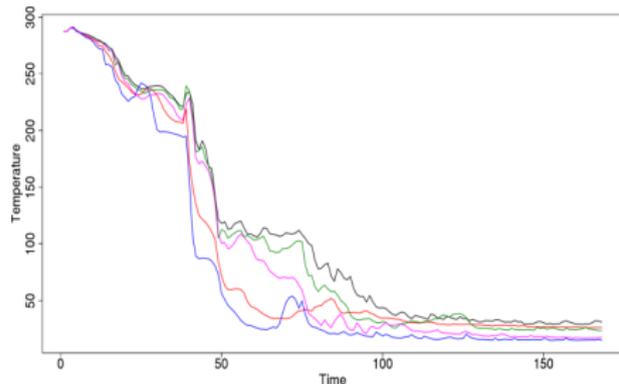


FIG.: RML

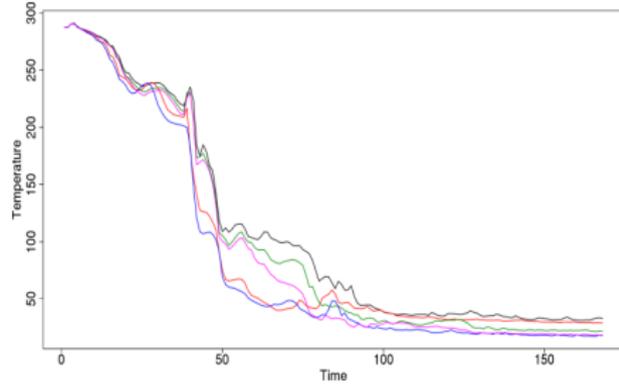


FIG.: Nadaraya-Watson

1 Réduction de la dimension

- ACP fonctionnelle
- Riemannian Manifold Learning (Lin et al. 2006)
- Reconstruction RML

2 Clustering des entrées-sorties

- Classification ascendante hiérarchique
- Distance "Commute-Time"
- Détermination du nombre de groupes

3 Applications

- Méthodologie
- Jeu de données CATHARE I
- Jeu de données CATHARE II

Test II - courbes de température CATHARE

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

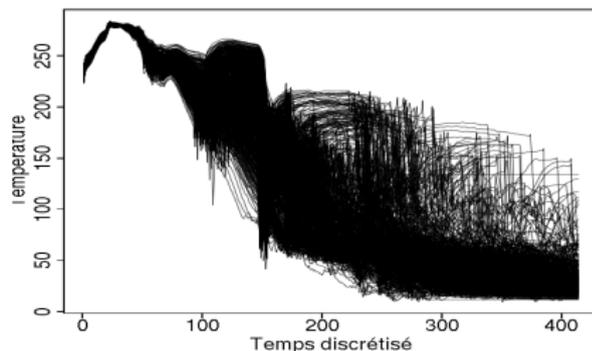


FIG.: Les 600 courbes en sortie

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. ; $d = 7$

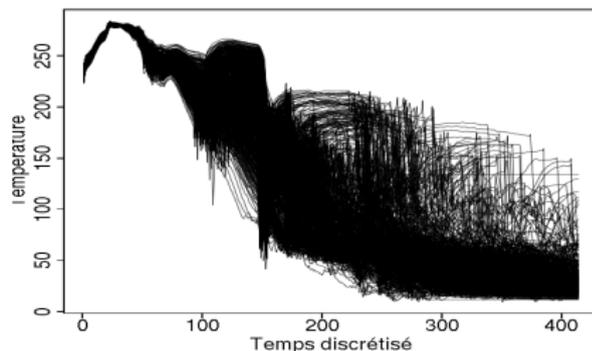


FIG.: Les 600 courbes en sortie

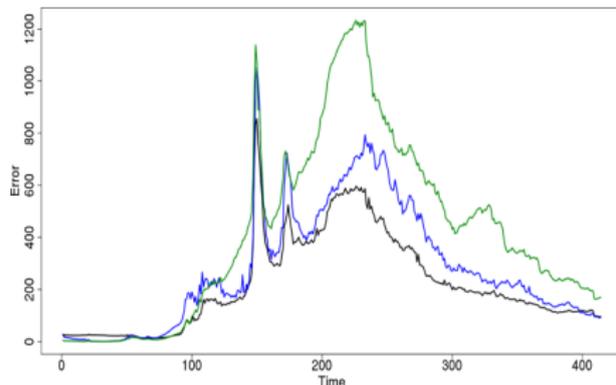
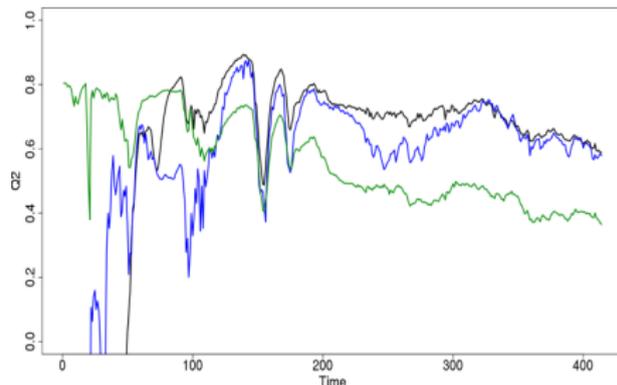


FIG.: Noir : functional PCA ; bleu : RML ; vert : Nadaraya-Watson.



5 courbes prédites

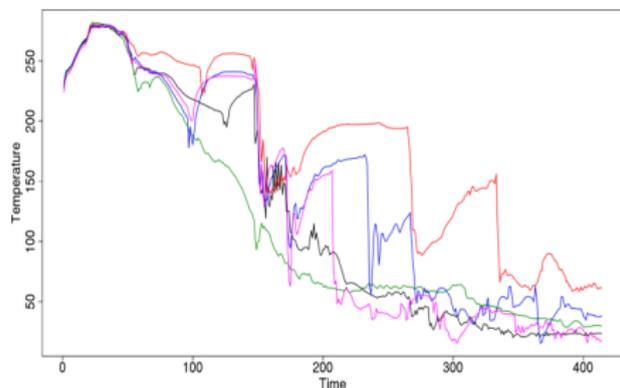


FIG.: Courbes cibles

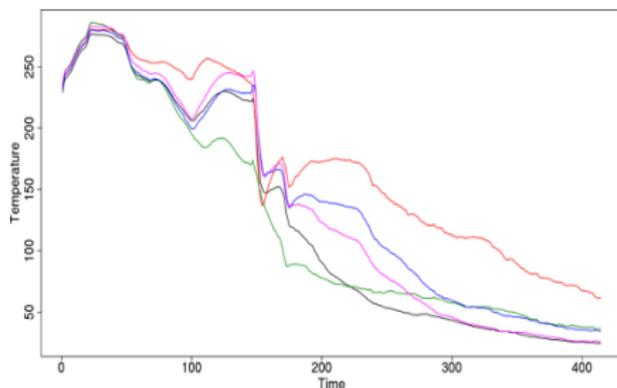


FIG.: ACP fonc.

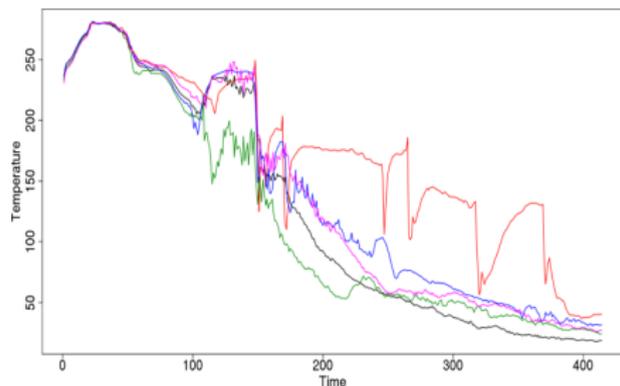


FIG.: RML

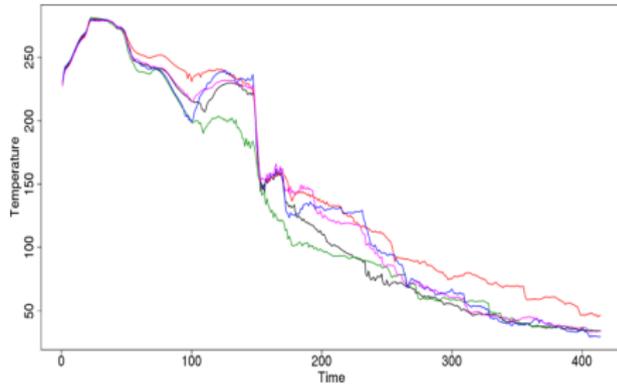


FIG.: Nadaraya-Watson

Conclusion

Modèle assez satisfaisant par rapport aux objectifs industriels.

⇒ aide au projet DDVCV (durée de vie des cuves).

Conclusion

Modèle assez satisfaisant par rapport aux objectifs industriels.

⇒ aide au projet DDVCV (durée de vie des cuves).

Réduction de dimension

ACPF (linéaire) et RML : complémentaires.

Quelques paramètres ne sont pas optimisés automatiquement . . .

Conclusion

Modèle assez satisfaisant par rapport aux objectifs industriels.

⇒ aide au projet DDVCV (durée de vie des cuves).

Réduction de dimension

ACPF (linéaire) et RML : complémentaires.

Quelques paramètres ne sont pas optimisés automatiquement . . .

Clustering

Inutile pour les courbes CATHARE

Distance ECT $>$ distance euclidienne si formes arbitraires.