Modélisation d'un Code de Calcul à Sorties Fonctionnelles

Benjamin Auder

CEA - UPMC

2 juin 2010

Thèse depuis 02/2008

Directeur de thèse : Gérard Biau (UPMC)

Encadrant : Bertrand Iooss (EDF, ex CEA)

Suivi labo. CEA : Michel Marquès

Contexte industriel

Cadre : durée de vie des cuves.

$\longrightarrow$ Diverses séquences d'accidents envisagées.

But : estimer leurs probabilités.

Contexte industriel

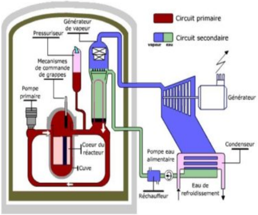
Cadre : durée de vie des cuves.

→ Diverses séquences d'accidents envisagées.

But : estimer leurs probabilités.

Méthodologie

Modélisation



Contexte industriel

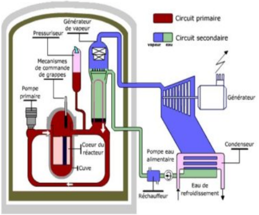
Cadre : durée de vie des cuves.

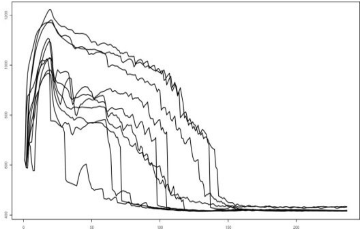
$\rightarrow$ Diverses séquences d'accidents envisagées.

But : estimer leurs probabilités.

Méthodologie

Modélisation $\longrightarrow$ Simulation





Contexte industriel

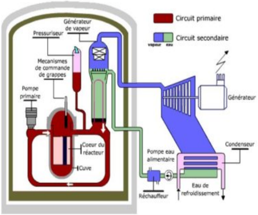
Cadre : durée de vie des cuves.

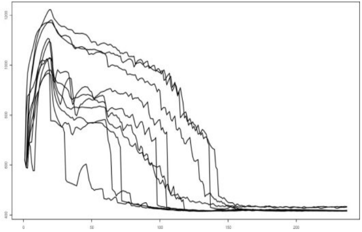
$\longrightarrow$ Diverses séquences d'accidents envisagées.

But : estimer leurs probabilités.

Méthodologie

Modélisation $\longrightarrow$ Simulation $\longrightarrow$ Calculs.





$\rightarrow$ Analyse de sensibilité, propagation d'incertitudes ..etc.

Contexte industriel

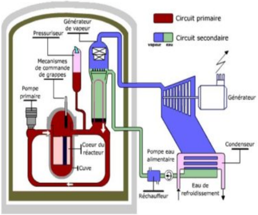
Cadre : durée de vie des cuves.

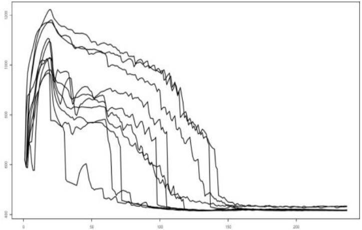
$\longrightarrow$ Diverses séquences d'accidents envisagées.

But : estimer leurs probabilités.

Méthodologie

Modélisation $\longrightarrow$ Simulation $\longrightarrow$ Calculs.





$\rightarrow$ 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

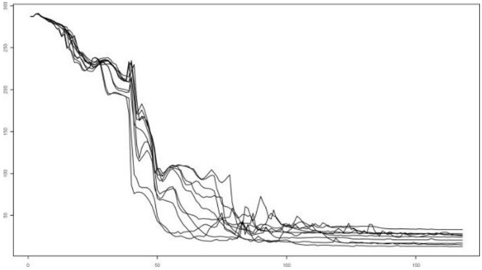


FIG.: Transitoires de température.

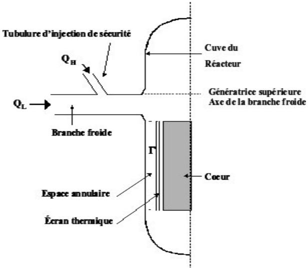


FIG.: Zone modélisée

Au CEA - DER/SESI/LSMR ...

Code thermo-hydraulique CATHARE coûteux en temps

- "boîte noire" ;

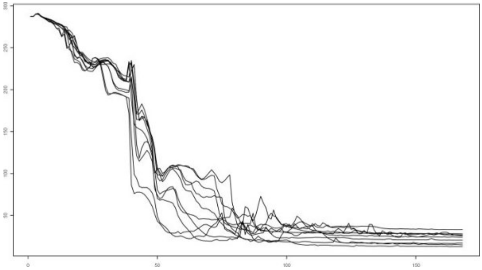


FIG.: Transitoires de température.

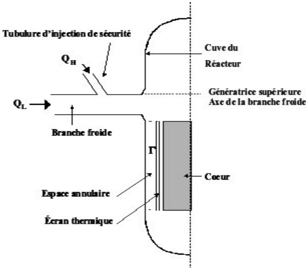


FIG.: Zone modélisée

Au CEA - DER/SESI/LSMR ...

Code thermo-hydraulique CATHARE coûteux en temps

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

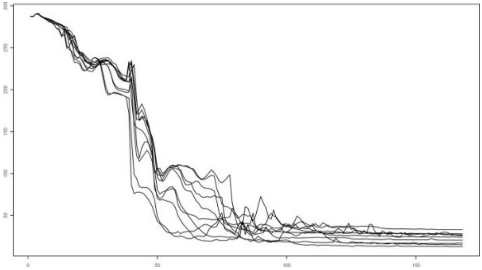


FIG.: Transitoires de température.

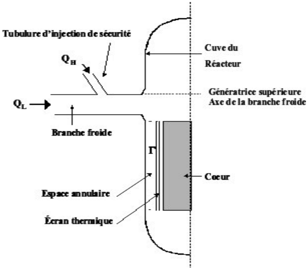


FIG.: Zone modélisée

Au CEA - DER/SESI/LSMR ...

Code thermo-hydraulique CATHARE coûteux en temps

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

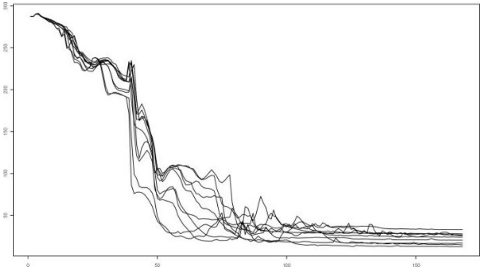


FIG.: Transitoires de température.

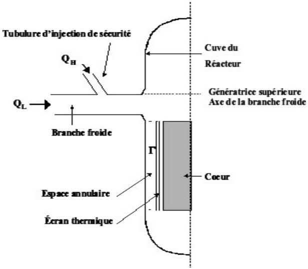


FIG.: Zone modélisée

$\boxed{\text{"Accélérer" } \textit{l'exécution du code CATHARE}}$

Solution ?

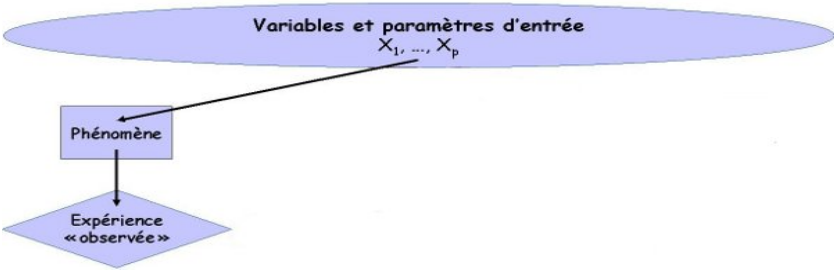


FIG.: Point de départ : phénomène physique.

marges d'incertitudes sur le phénomène

Solution ?

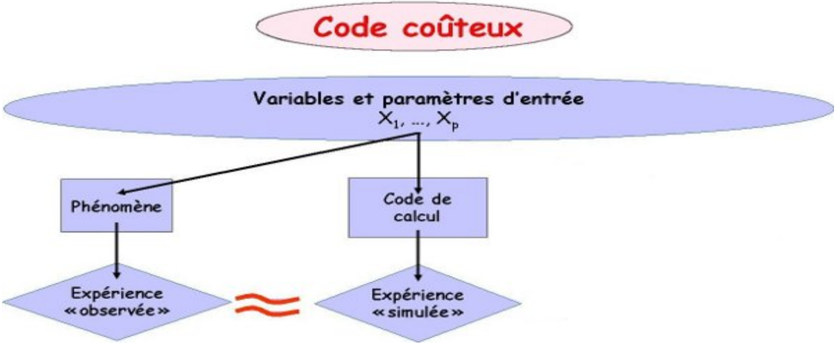


FIG.: Code de calcul = simulation du phénomène.

marges d'incertitudes sur le phénomène
nécessitent beaucoup de résultats de code

Solution : métamodèle

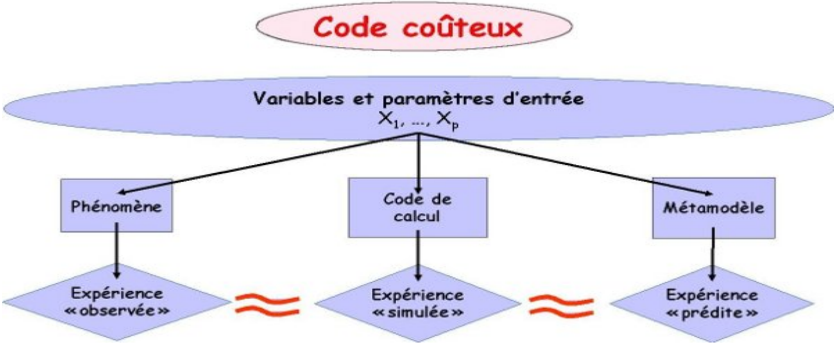


FIG.: Métamodèle = accélération des simulations.

	marges d'incertitudes sur le phénomène
nécessitent	beaucoup de résultats de code
obtenus avec	un métamodèle (= modèle du code).

Solution : métamodèle

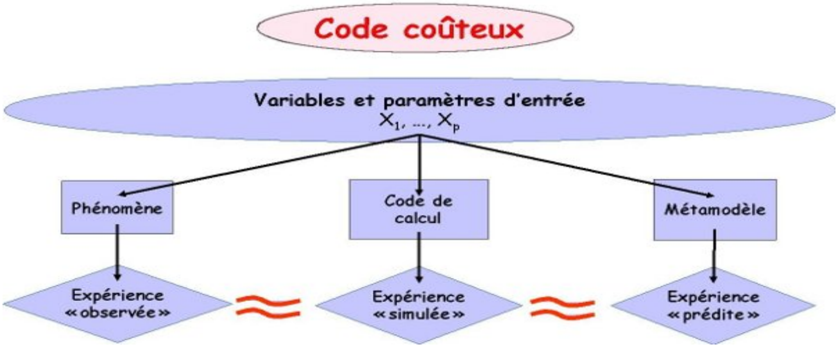


FIG.: Métamodèle = accélération des simulations.

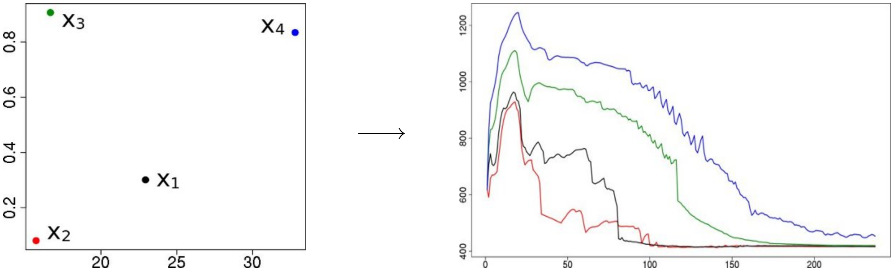
	marges d'incertitudes sur le phénomène
nécessitent	beaucoup de résultats de code
obtenus avec	un métamodèle (= modèle du code).

$\boxed{\textit{Construire un modèle du code CATHARE}}$

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.



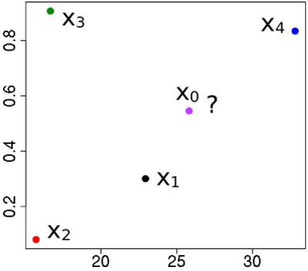
Reformulation finale

n couples (x_i, y_i) connus :

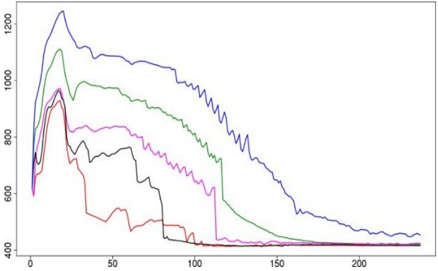
- Entrées $x_i \in \mathbb{R}^p$ = état initial du système physique ;
- Sorties $y_i \in \mathcal{C}([a,b], \mathbb{R})$ = évolution des paramètres.

Objectif = prédiction de données fonctionnelles :

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



$\longrightarrow$



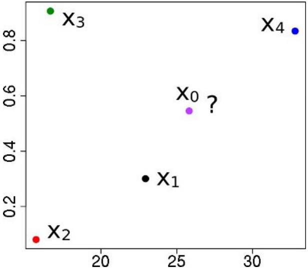
Reformulation finale

n couples (x_i, y_i) connus :

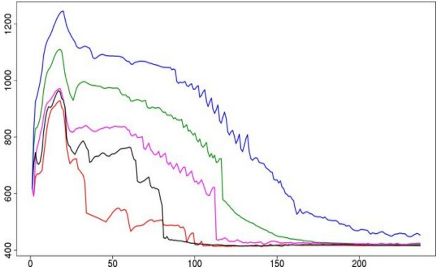
- Entrées $x_i \in \mathbb{R}^p$ = état initial du système physique ;
- Sorties $y_i \in \mathcal{C}([a, b], \mathbb{R})$ = évolution des paramètres.

Objectif = prédiction de données fonctionnelles :

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



$\longrightarrow$

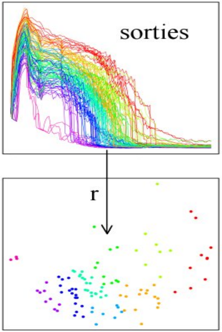


$$\boxed{\textit{Apprentissage statistique "régression" } \mathbb{R}^p \rightarrow \mathcal{C}([a, b], \mathbb{R})}$$

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

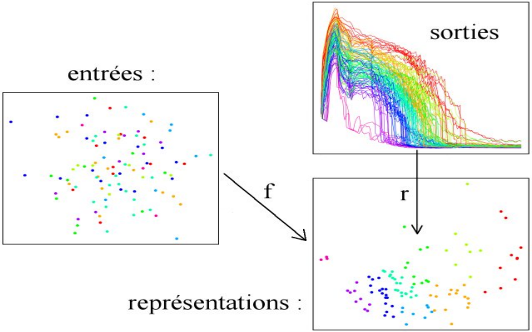
Comment se ramener au cas "simple" $y_i \in \mathbb{R}^d$?

1. réduction de la dimension :
   $r : \mathcal{C}([a, b], \mathbb{R}) \longrightarrow \mathbb{R}^d$ (représentation) ;



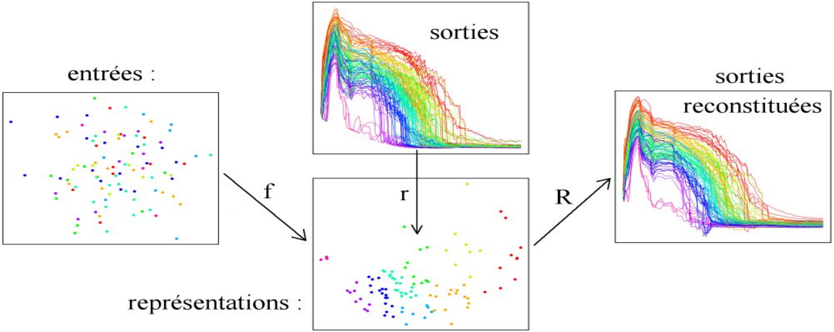
Comment se ramener au cas "simple" $y_i \in \mathbb{R}^d$?

1. réduction de la dimension :
   $r : \mathcal{C}([a, b], \mathbb{R}) \longrightarrow \mathbb{R}^d$ (représentation) ;
2. apprentissage statistique classique :
   $f : \mathbb{R}^p$ (entrées) $\longrightarrow \mathbb{R}^d$ (sorties réduites) ;



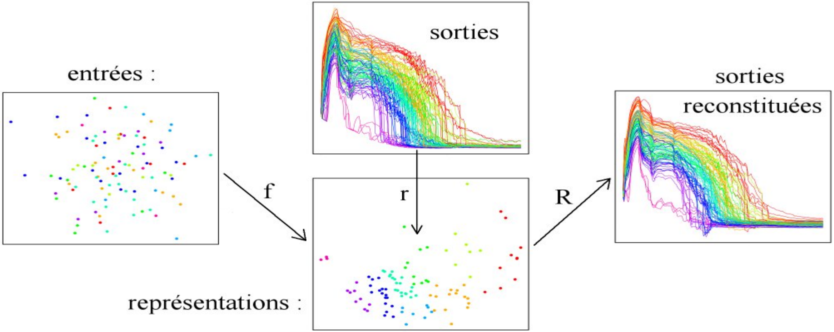
Comment se ramener au cas "simple" $y_i \in \mathbb{R}^d$?

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



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

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

2 Réduction de la dimension des sorties

3 Classification des (entrées-)sorties

4 Applications

1 Exemple introductif

2 Réduction de la dimension des sorties

3 Classification des (entrées-)sorties

4 Applications

Données brutes

→ rien de très clair . . .

a priori 6 dimensions

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.5091564	1.577159	-0.675	2.183015	1.868054	1.1105867
[2,]	0.7788008	1.064494	-0.250	2.635077	1.280696	2.8141909
[3,]	0.6537698	1.197966	-0.425	2.486904	1.510319	1.8096824
[4,]	0.5627049	1.391838	-0.575	2.314497	1.722600	1.3368171
[5,]	0.6872893	1.150993	-0.375	2.535796	1.442348	2.0258229
[6,]	0.7595721	1.078558	-0.275	2.618038	1.311992	2.6167830
[7,]	0.6376282	1.224460	-0.450	2.460703	1.544910	1.7151039
[8,]	0.5220458	1.525771	-0.650	2.216842	1.831594	1.1625144
[9,]	0.6065307	1.284025	-0.500	2.405079	1.615146	1.5472631
[10,]	0.9512294	1.002503	-0.050	2.714887	1.051249	6.6648135
[11,]	0.5769498	1.353238	-0.550	2.345561	1.686554	1.4023176
[12,]	0.5352614	1.477904	-0.625	2.250074	1.795166	1.2173085
[13,]	0.8607080	1.022755	-0.150	2.687929	1.161182	3.9819186
[14,]	0.9277435	1.005641	-0.075	2.710651	1.077808	5.7096872
[15,]	0.4168620	2.150338	-0.875	1.898372	2.154467	0.7707595
[16,]	0.4065697	2.247908	-0.900	1.861923	2.188742	0.7352086
[17,]	0.8394570	1.031099	-0.175	2.677080	1.190184	3.6126043
[18,]	0.7046881	1.130319	-0.350	2.558376	1.409025	2.1504059
[19,]	0.8187308	1.040811	-0.200	2.664634	1.219779	3.3034914
[20,]	0.6703200	1.173511	-0.400	2.511954	1.476122	1.9128022
[21,]	0.4965853	1.632316	-0.700	2.148655	1.904497	1.0612462
[22,]	0.9048374	1.010050	-0.100	2.704736	1.104987	4.9916472
[23,]	0.7985162	1.051928	-0.225	2.650621	1.249954	3.0406655
[24,]	0.4493290	1.896481	-0.800	2.007132	2.049009	0.8852389
[25,]	0.7225274	1.111405	-0.325	2.579642	1.376176	2.2886031
[26,]	0.9753099	1.000625	-0.025	2.717433	1.025312	7.9996666
[27,]	0.3678794	2.718282	-1.000	1.716526	2.319777	0.6033518
[28,]	0.5488116	1.433329	-0.600	2.282647	1.758819	1.2752877
[29,]	0.7408182	1.094174	-0.300	2.599545	1.343825	2.4429793
[30,]	0.4274149	2.059576	-0.850	1.934760	2.119712	0.8075374
[31,]	0.3867410	2.465760	-0.950	1.789047	2.255599	0.6673863
[32,]	0.4607038	1.823258	-0.775	2.043004	2.013181	0.9264261
[33,]	0.3771924	2.587326	-0.975	1.752719	2.288054	0.6349357
[34,]	0.5915554	1.317354	-0.525	2.375776	1.650722	1.4722768
[35,]	0.4723666	1.755055	-0.750	2.078588	1.977115	0.9693713
[36,]	0.8824969	1.015748	-0.125	2.697155	1.132780	4.4315410
[37,]	0.4382350	1.975112	-0.825	1.971031	2.084539	0.8456556
[38,]	0.6218851	1.253106	-0.475	2.433408	1.579863	1.6279430
[39,]	0.4843246	1.691516	-0.725	2.113826	1.940868	1.0142469
[40,]	0.3965314	2.352844	-0.925	1.825465	2.222473	0.7007821

Données brutes

→ rien de très clair . . .

a priori 6 dimensions

représentation en dimension réduite ?

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.5091564	1.577159	-0.675	2.183015	1.868054	1.1105867
[2,]	0.7788008	1.064494	-0.250	2.635077	1.280696	2.8141909
[3,]	0.6537698	1.197966	-0.425	2.486904	1.510319	1.8096824
[4,]	0.5627049	1.391838	-0.575	2.314497	1.722600	1.3368171
[5,]	0.6872893	1.150993	-0.375	2.535796	1.442348	2.0258229
[6,]	0.7595721	1.078558	-0.275	2.618038	1.311992	2.6167830
[7,]	0.6376282	1.224460	-0.450	2.460703	1.544910	1.7151039
[8,]	0.5220458	1.525771	-0.650	2.216842	1.831594	1.1625144
[9,]	0.6065307	1.284025	-0.500	2.405079	1.615146	1.5472631
[10,]	0.9512294	1.002503	-0.050	2.714887	1.051249	6.6648135
[11,]	0.5769498	1.353238	-0.550	2.345561	1.686554	1.4023176
[12,]	0.5352614	1.477904	-0.625	2.250074	1.795166	1.2173085
[13,]	0.8607080	1.022755	-0.150	2.687929	1.161182	3.9819186
[14,]	0.9277435	1.005641	-0.075	2.710651	1.077808	5.7096872
[15,]	0.4168620	2.150338	-0.875	1.898372	2.154467	0.7707595
[16,]	0.4065697	2.247908	-0.900	1.861923	2.188742	0.7352086
[17,]	0.8394570	1.031099	-0.175	2.677080	1.190184	3.6126043
[18,]	0.7046881	1.130319	-0.350	2.558376	1.409025	2.1504059
[19,]	0.8187308	1.040811	-0.200	2.664634	1.219779	3.3034914
[20,]	0.6703200	1.173511	-0.400	2.511954	1.476122	1.9128022
[21,]	0.4965853	1.632316	-0.700	2.148655	1.904497	1.0612462
[22,]	0.9048374	1.010050	-0.100	2.704736	1.104987	4.9916472
[23,]	0.7985162	1.051928	-0.225	2.650621	1.249954	3.0406655
[24,]	0.4493290	1.896481	-0.800	2.007132	2.049009	0.8852389
[25,]	0.7225274	1.111405	-0.325	2.579642	1.376176	2.2886031
[26,]	0.9753099	1.000625	-0.025	2.717433	1.025312	7.9996666
[27,]	0.3678794	2.718282	-1.000	1.716526	2.319777	0.6033518
[28,]	0.5488116	1.433329	-0.600	2.282647	1.758819	1.2752877
[29,]	0.7408182	1.094174	-0.300	2.599545	1.343825	2.4429793
[30,]	0.4274149	2.059576	-0.850	1.934760	2.119712	0.8075374
[31,]	0.3867410	2.465760	-0.950	1.789047	2.255599	0.6673863
[32,]	0.4607038	1.823258	-0.775	2.043004	2.013181	0.9264261
[33,]	0.3771924	2.587326	-0.975	1.752719	2.288054	0.6349357
[34,]	0.5915554	1.317354	-0.525	2.375776	1.650722	1.4722768
[35,]	0.4723666	1.755055	-0.750	2.078588	1.977115	0.9693713
[36,]	0.8824969	1.015748	-0.125	2.697155	1.132780	4.4315410
[37,]	0.4382350	1.975112	-0.825	1.971031	2.084539	0.8456556
[38,]	0.6218851	1.253106	-0.475	2.433408	1.579863	1.6279430
[39,]	0.4843246	1.691516	-0.725	2.113826	1.940868	1.0142469
[40,]	0.3965314	2.352844	-0.925	1.825465	2.222473	0.7007821

1^{ere} approche : ACP

FIG.: Décomposition ACP à une composante : insuffisant

1^{ere} approche : ACP

FIG.: Décomposition ACP à une composante : insuffisant

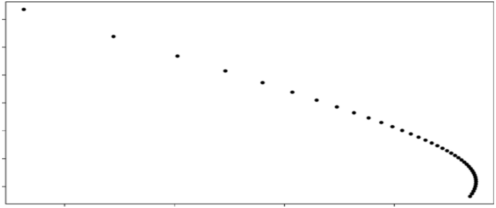


FIG.: Décomposition ACP à deux composantes : "OK"

1^{ere} approche : ACP

FIG.: Décomposition ACP à une composante : insuffisant

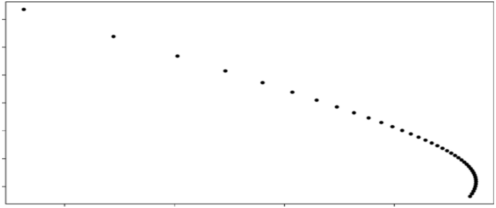


FIG.: Décomposition ACP à deux composantes : "OK"

"Donc" deux dimensions ?

Révélation

réordonnancement

$\Rightarrow$ une structure émerge

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.3678794	2.718282	-1.000	1.716526	2.319777	0.6033518
[2,]	0.3771924	2.587326	-0.975	1.752719	2.288054	0.6349357
[3,]	0.3867410	2.465760	-0.950	1.789047	2.255599	0.6673863
[4,]	0.3965314	2.352844	-0.925	1.825465	2.222473	0.7007821
[5,]	0.4065697	2.247908	-0.900	1.861923	2.188742	0.7352086
[6,]	0.4168620	2.150338	-0.875	1.898372	2.154467	0.7707595
[7,]	0.4274149	2.059576	-0.850	1.934760	2.119712	0.8075374
[8,]	0.4382350	1.975112	-0.825	1.971031	2.084539	0.8456556
[9,]	0.4493290	1.896481	-0.800	2.007132	2.049009	0.8852389
[10,]	0.4607038	1.823258	-0.775	2.043004	2.013181	0.9264261
[11,]	0.4723666	1.755055	-0.750	2.078588	1.977115	0.9693713
[12,]	0.4843246	1.691516	-0.725	2.113826	1.940868	1.0142469
[13,]	0.4965853	1.632316	-0.700	2.148655	1.904497	1.0612462
[14,]	0.5091564	1.577159	-0.675	2.183015	1.868054	1.1105867
[15,]	0.5220458	1.525771	-0.650	2.216842	1.831594	1.1625144
[16,]	0.5352614	1.477904	-0.625	2.250074	1.795166	1.2173085
[17,]	0.5488116	1.433329	-0.600	2.282647	1.758819	1.2752877
[18,]	0.5627049	1.391838	-0.575	2.314497	1.722600	1.3368171
[19,]	0.5769498	1.353238	-0.550	2.345561	1.686554	1.4023176
[20,]	0.5915554	1.317354	-0.525	2.375776	1.650722	1.4722768
[21,]	0.6065307	1.284025	-0.500	2.405079	1.615146	1.5472631
[22,]	0.6218851	1.253106	-0.475	2.433408	1.579863	1.6279430
[23,]	0.6376282	1.224460	-0.450	2.460703	1.544910	1.7151039
[24,]	0.6537698	1.197966	-0.425	2.486904	1.510319	1.8096824
[25,]	0.6703200	1.173511	-0.400	2.511954	1.476122	1.9128022
[26,]	0.6872893	1.150993	-0.375	2.535796	1.442348	2.0258229
[27,]	0.7046881	1.130319	-0.350	2.558376	1.409025	2.1504059
[28,]	0.7225274	1.111405	-0.325	2.579642	1.376176	2.2886031
[29,]	0.7408182	1.094174	-0.300	2.599545	1.343825	2.4429793
[30,]	0.7595721	1.078558	-0.275	2.618038	1.311992	2.6167830
[31,]	0.7788008	1.064494	-0.250	2.635077	1.280696	2.8141909
[32,]	0.7985162	1.051928	-0.225	2.650621	1.249954	3.0406655
[33,]	0.8187308	1.040811	-0.200	2.664634	1.219779	3.3034914
[34,]	0.8394570	1.031099	-0.175	2.677080	1.190184	3.6126043
[35,]	0.8607080	1.022755	-0.150	2.687929	1.161182	3.9819186
[36,]	0.8824969	1.015748	-0.125	2.697155	1.132780	4.4315410
[37,]	0.9048374	1.010050	-0.100	2.704736	1.104987	4.9916472
[38,]	0.9277435	1.005641	-0.075	2.710651	1.077808	5.7096872
[39,]	0.9512294	1.002503	-0.050	2.714887	1.051249	6.6648135
[40,]	0.9753099	1.000625	-0.025	2.717433	1.025312	7.9996666

Révélation

réordonnancement

$\Rightarrow$ une structure émerge

fonction génératrice :

$$t \mapsto (e^{-t}, e^{t^2}, -t, e^{\cos t}, e^{\sin t}, \frac{1}{\tan t + 0.1}),$$

aux points

$$t = 1/40, 2/40, \ldots, 1.$$

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0.3678794	2.718282	-1.000	1.716526	2.319777	0.6033518
[2,]	0.3771924	2.587326	-0.975	1.752719	2.288054	0.6349357
[3,]	0.3867410	2.465760	-0.950	1.789047	2.255599	0.6673863
[4,]	0.3965314	2.352844	-0.925	1.825465	2.222473	0.7007821
[5,]	0.4065697	2.247908	-0.900	1.861923	2.188742	0.7352086
[6,]	0.4168620	2.150338	-0.875	1.898372	2.154467	0.7707595
[7,]	0.4274149	2.059576	-0.850	1.934760	2.119712	0.8075374
[8,]	0.4382350	1.975112	-0.825	1.971031	2.084539	0.8456556
[9,]	0.4493290	1.896481	-0.800	2.007132	2.049009	0.8852389
[10,]	0.4607038	1.823258	-0.775	2.043004	2.013181	0.9264261
[11,]	0.4723666	1.755055	-0.750	2.078588	1.977115	0.9693713
[12,]	0.4843246	1.691516	-0.725	2.113826	1.940868	1.0142469
[13,]	0.4965853	1.632316	-0.700	2.148655	1.904497	1.0612462
[14,]	0.5091564	1.577159	-0.675	2.183015	1.868054	1.1105867
[15,]	0.5220458	1.525771	-0.650	2.216842	1.831594	1.1625144
[16,]	0.5352614	1.477904	-0.625	2.250074	1.795166	1.2173085
[17,]	0.5488116	1.433329	-0.600	2.282647	1.758819	1.2752877
[18,]	0.5627049	1.391838	-0.575	2.314497	1.722600	1.3368171
[19,]	0.5769498	1.353238	-0.550	2.345561	1.686554	1.4023176
[20,]	0.5915554	1.317354	-0.525	2.375776	1.650722	1.4722768
[21,]	0.6065307	1.284025	-0.500	2.405079	1.615146	1.5472631
[22,]	0.6218851	1.253106	-0.475	2.433408	1.579863	1.6279430
[23,]	0.6376282	1.224460	-0.450	2.460703	1.544910	1.7151039
[24,]	0.6537698	1.197966	-0.425	2.486904	1.510319	1.8096824
[25,]	0.6703200	1.173511	-0.400	2.511954	1.476122	1.9128022
[26,]	0.6872893	1.150993	-0.375	2.535796	1.442348	2.0258229
[27,]	0.7046881	1.130319	-0.350	2.558376	1.409025	2.1504059
[28,]	0.7225274	1.111405	-0.325	2.579642	1.376176	2.2886031
[29,]	0.7408182	1.094174	-0.300	2.599545	1.343825	2.4429793
[30,]	0.7595721	1.078558	-0.275	2.618038	1.311992	2.6167830
[31,]	0.7788008	1.064494	-0.250	2.635077	1.280696	2.8141909
[32,]	0.7985162	1.051928	-0.225	2.650621	1.249954	3.0406655
[33,]	0.8187308	1.040811	-0.200	2.664634	1.219779	3.3034914
[34,]	0.8394570	1.031099	-0.175	2.677080	1.190184	3.6126043
[35,]	0.8607080	1.022755	-0.150	2.687929	1.161182	3.9819186
[36,]	0.8824969	1.015748	-0.125	2.697155	1.132780	4.4315410
[37,]	0.9048374	1.010050	-0.100	2.704736	1.104987	4.9916472
[38,]	0.9277435	1.005641	-0.075	2.710651	1.077808	5.7096872
[39,]	0.9512294	1.002503	-0.050	2.714887	1.051249	6.6648135
[40,]	0.9753099	1.000625	-0.025	2.717433	1.025312	7.9996666

En une dimension

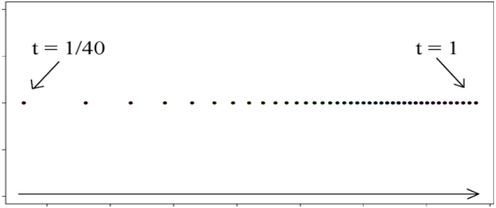


FIG.: Réduction de dimension non linéaire : une composante suffit.

En une dimension

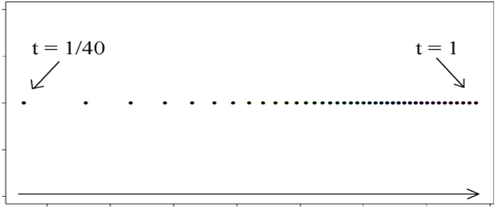


FIG.: Réduction de dimension non linéaire : une composante suffit.

Hypothèse

Les courbes en sortie ont une structure de variété

. . .éventuellement non linéaire

$\Rightarrow$ *décomposition sur une base = sous-optimal.*

1 Exemple introductif

2 Réduction de la dimension des sorties

3 Classification des (entrées-)sorties

4 Applications

Dimension réduite "pas à pas"

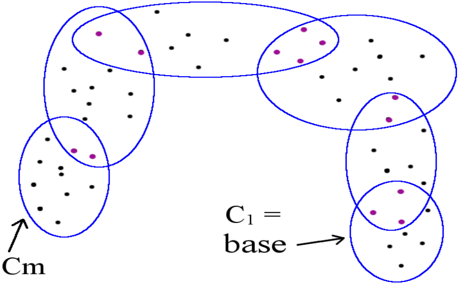
But : transformer y_i en $z_i \in \mathbb{R}^d$ avec d "petit".

Dimension réduite "pas à pas"

But : transformer y_i en $z_i \in \mathbb{R}^d$ avec d "petit".

Local PCA Manifold Learning (Zhan et al., 2008) ;

Clé : "Traversal Sequence of Local Neighborhoods"



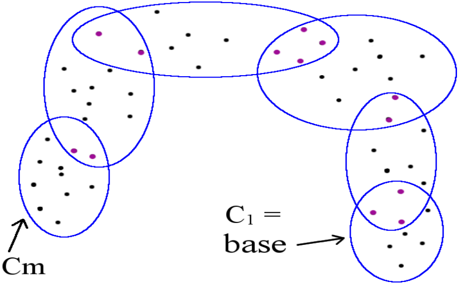
1. Suite $C_1, \ldots, C_m \subset \{1, \ldots, n\}$ recouvrant $\{1, \ldots, n\}$, avec des intersections $\cup_{j<i} C_i \cap C_j$ non vides $\forall i > 1$;

Dimension réduite "pas à pas"

But : transformer y_i en $z_i \in \mathbb{R}^d$ avec d "petit".

Local PCA Manifold Learning (Zhan et al., 2008) ;

Clé : "Traversal Sequence of Local Neighborhoods"



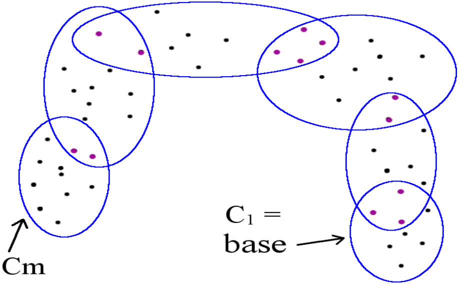
1. Suite $C_1, \ldots, C_m \subset \{1, \ldots, n\}$ recouvrant $\{1, \ldots, n\}$, avec des intersections $\cup_{j<i} C_i \cap C_j$ non vides $\forall i > 1$;
2. ACP locale sur chaque $C_j \rightarrow$ coordonnées z_i' locales ;

Dimension réduite "pas à pas"

But : transformer y_i en $z_i \in \mathbb{R}^d$ avec d "petit".

Local PCA Manifold Learning (Zhan et al., 2008) ;

Clé : "Traversal Sequence of Local Neighborhoods"



1. Suite $C_1, \ldots, C_m \subset \{1, \ldots, n\}$ recouvrant $\{1, \ldots, n\}$, avec des intersections $\cup_{j<i} C_i \cap C_j$ non vides $\forall i > 1$;
2. ACP locale sur chaque $C_j \rightarrow$ coordonnées z_i' locales ;
3. Transformations affines $\rightarrow$ coordonnées z_i globales ; $\hookrightarrow$ *principe :* optimiser la matrice de transformation sur l'intersection, puis appliquer sur la "cellule" suivante.

Estimation des coordonnées globales (étapes 2−3)

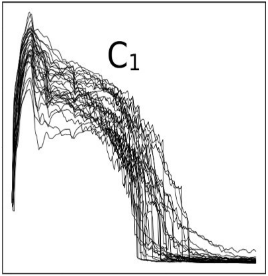


FIG.: données

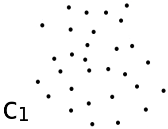


FIG.: coord. locales

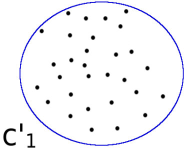


FIG.: coord. globales

1. ACP sur $C_1 \longrightarrow$ coord. locales c_1 ;
   on pose $c_1' = c_1$ (coord. globales) ;

Estimation des coordonnées globales (étapes 2–3)

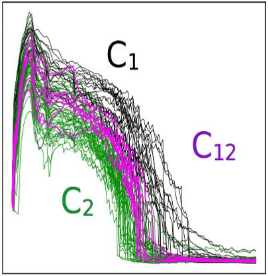


FIG.: données

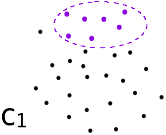


FIG.: coord. locales

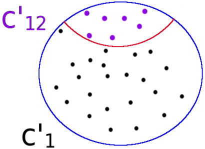


FIG.: coord. globales

1. ACP sur $C_1 \longrightarrow$ coord. locales c_1 ;
   on pose $c'_1 = c_1$ (coord. globales) ;
2. c'_1 contient les coord. globales de $C_1 \cap C_2$, notées c'_{12} ;

Estimation des coordonnées globales (étapes 2–3)

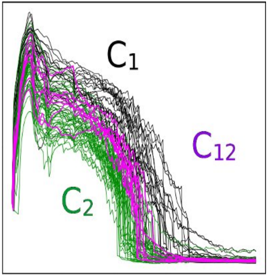


FIG.: données

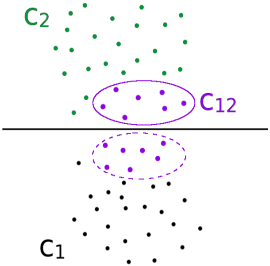


FIG.: coord. locales

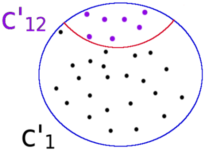


FIG.: coord. globales

1. ACP sur $C_1 \longrightarrow$ coord. locales c_1 ;
   on pose $c_1' = c_1$ (coord. globales) ;
2. c_1' contient les coord. globales de $C_1 \cap C_2$, notées c_{12}' ;
3. ACP sur $C_2 \longrightarrow$ coord. locales c_2 ;

Estimation des coordonnées globales (étapes 2–3)

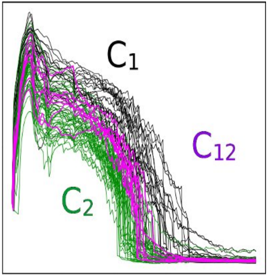


FIG.: données

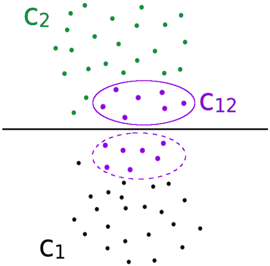


FIG.: coord. locales

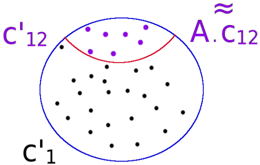


FIG.: coord. globales

1. ACP sur $C_1 \longrightarrow$ coord. locales c_1 ;
   on pose $c_1' = c_1$ (coord. globales) ;
2. c_1' contient les coord. globales de $C_1 \cap C_2$, notées c_{12}' ;
3. ACP sur $C_2 \longrightarrow$ coord. locales c_2 ;
4. recherche d'une transformation affine A telle que $A\, c_{12} \simeq c_{12}'$;

Estimation des coordonnées globales (étapes 2–3)

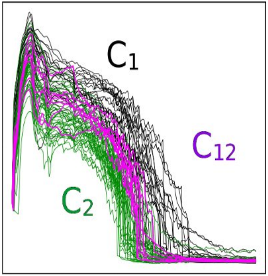


FIG.: données

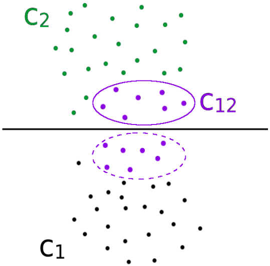


FIG.: coord. locales

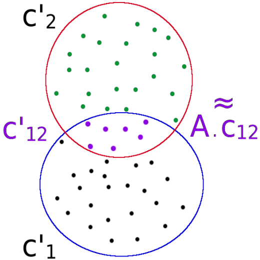


FIG.: coord. globales

1. ACP sur $C_1 \longrightarrow$ coord. locales c_1 ;
   on pose $c_1' = c_1$ (coord. globales) ;
2. c_1' contient les coord. globales de $C_1 \cap C_2$, notées c_{12}' ;
3. ACP sur $C_2 \longrightarrow$ coord. locales c_2 ;
4. recherche d'une transformation affine A telle que $A\,c_{12} \simeq c_{12}'$;
5. application de A sur les points de $C_2 \setminus C_1 \longrightarrow c_2'$;
   retour en 2. avec c_2' ... etc.

Exemples

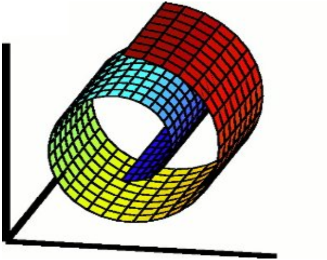


FIG.: Swissroll, 400 points 3D

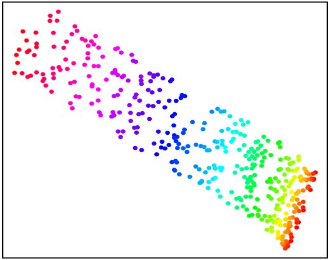


FIG.: Représentation LpcaML

Exemples

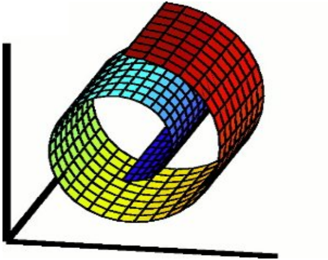


FIG.: Swissroll, 400 points 3D

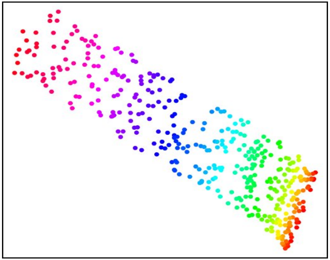


FIG.: Représentation LpcaML

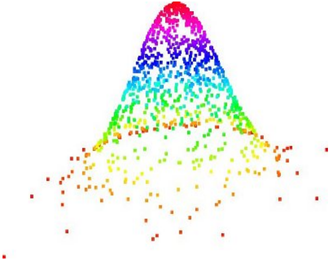


FIG.: Gaussienne, 1000 points 3D

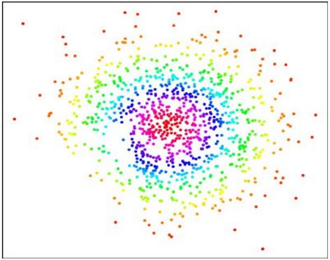


FIG.: Représentation LpcaML

1 Exemple introductif

2 Réduction de la dimension des sorties

3 Classification des (entrées-)sorties

4 Applications

Motivations

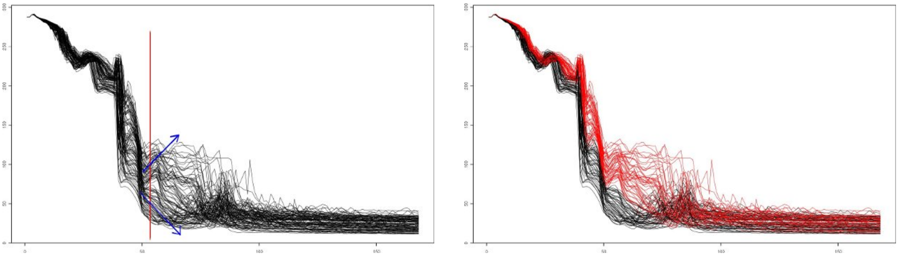


Fig.: 100 transitoires de température

$\longrightarrow$ Différents types de comportements physiques.

Motivations

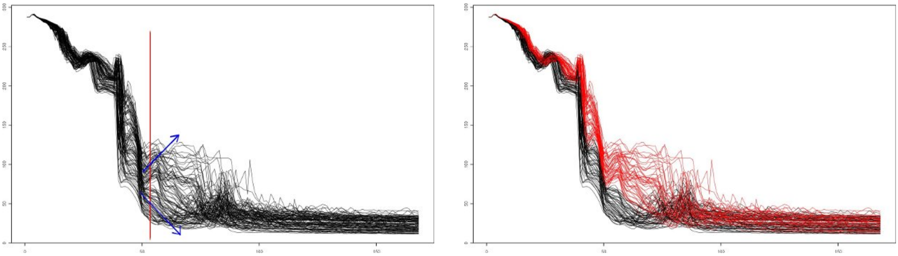


FIG.: 100 transitoires de température

→ Différents types de comportements physiques.

Regroupement des courbes aux caractéristiques similaires

⇒ meilleure modélisation dans chaque cluster.

(Entrées-)Sorties dans un graphe

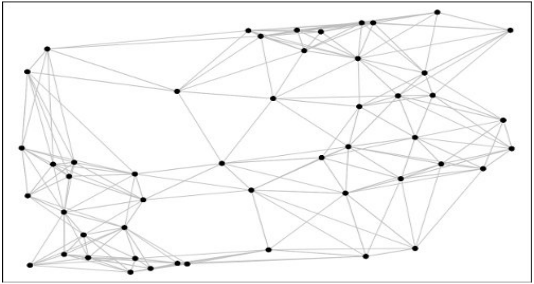


FIG.: Représentation des (entrées-)sorties dans un graphe ;

sommets = données, arêtes = distances.

(Entrées-)Sorties dans un graphe

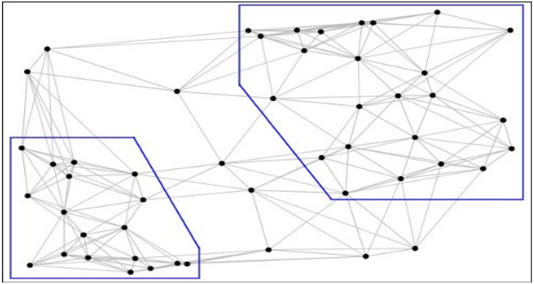


Fig.: En bleu : deux clusters (visuels) ;

(Entrées-)Sorties dans un graphe

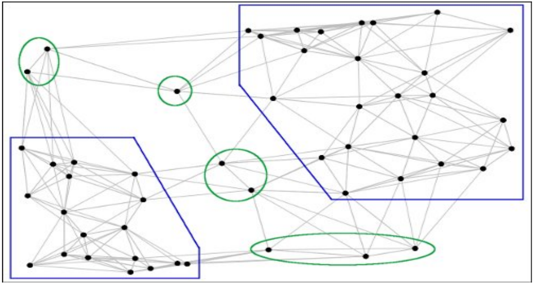


FIG.: En vert : points "isolés" ;

(Entrées-)Sorties dans un graphe

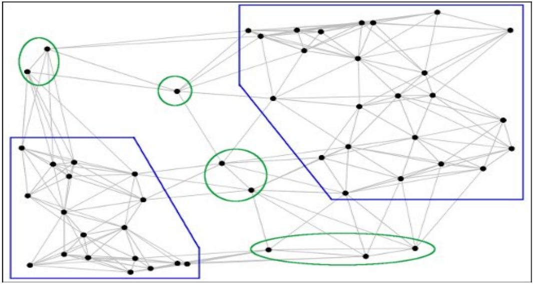


FIG.: En vert : points "isolés" ;

Méthode

1. Calcul des distances "commute-time".
2. Clustering hiérarchique basé sur ces distances.

Résumé du modèle

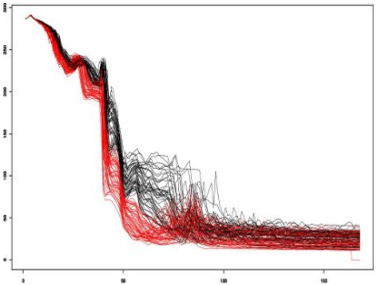


FIG.: 100 transitoires de température en sortie (cathare)

1. Classification non supervisée des n courbes y_i en K clusters C_j

Résumé du modèle

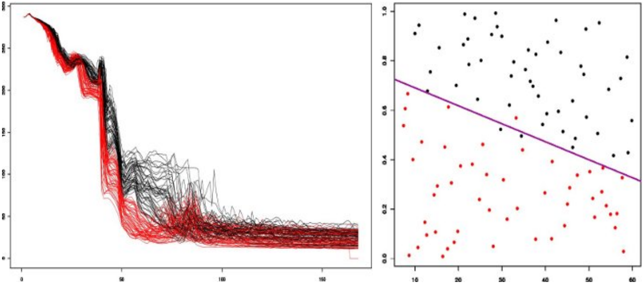


FIG.: g. à d. : sorties fonctionnelles, scatterplot entrées 1 - 4

1. Classification non supervisée des n courbes y_i en K clusters C_j + classification supervisée des entrées x_i.

Résumé du modèle

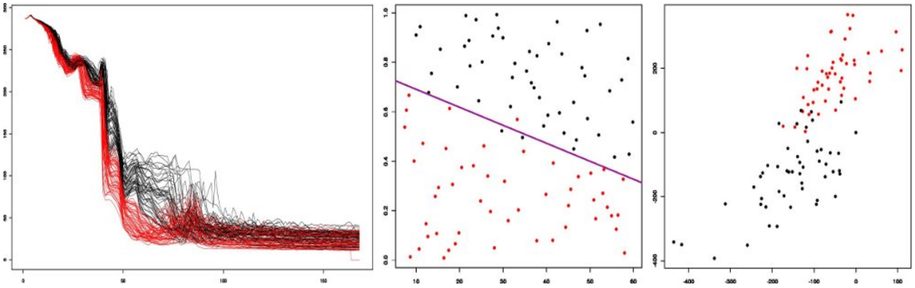


FIG.: g. à d. : sorties fonc., entrées 1 - 4, représentation 2D des sorties

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 $\mathbb{R}^d$;

Résumé du modèle

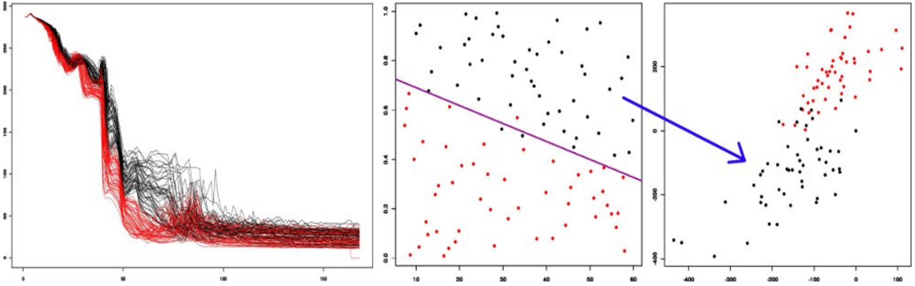


FIG.: g. à d. : sorties fonc., entrées 1 - 4, rep. 2D des sorties

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 : $f(x_i) \simeq z_i$;

Résumé du modèle

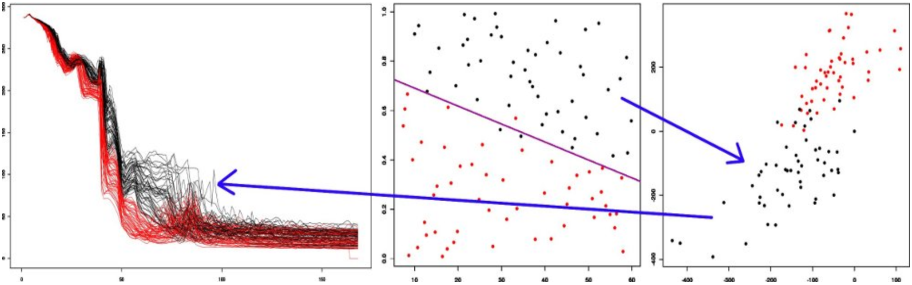


FIG.: g. à d. : sorties fonc., entrées 1 - 4, rep. 2D des sorties

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 : $f(x_i) \simeq z_i$;
   3. apprentissage d'une fonction de reconstruction : $R(z_i) \simeq y_i$.

1 Exemple introductif

2 Réduction de la dimension des sorties

3 Classification des (entrées-)sorties

4 Applications

Étape de validation

Données :

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

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

Étape de validation

Données :

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

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

Mesure "absolue" puis relative de l'erreur ponctuelle

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

Étape de validation

Données :

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

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

Mesure "absolue" puis relative de l'erreur ponctuelle

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

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

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

$\simeq 1 \Rightarrow$ modèle parfait.

Test I - température "facile"

100 simulations,
4 dimensions en entrée,
168 points de discrétisation.

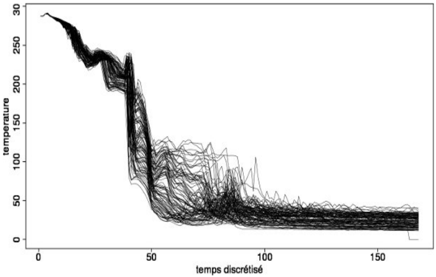


FIG.: Les 100 sorties du code

Test I - température "facile"

100 simulations,
4 dimensions en entrée,
168 points de discrétisation.

validation croisée
leave-one-out :

MSE à g., Q_2 à d. ; $\boxed{d = 4}$

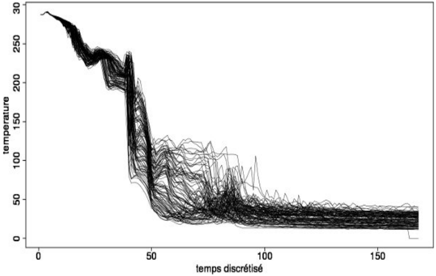


FIG.: Les 100 sorties du code

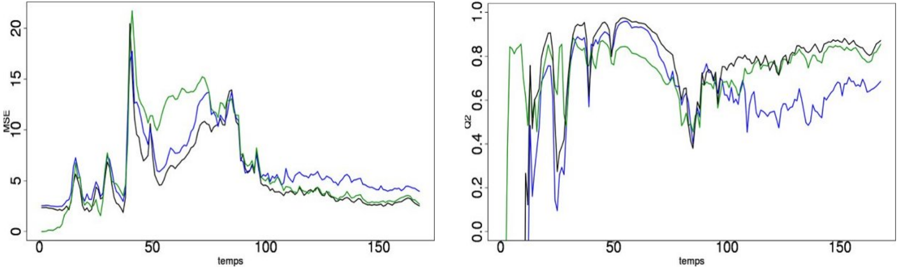


FIG.: Noir : ACP fonctionnelle ; bleu : LPcaML ; vert : Nadaraya-Watson.

Courbes prédites

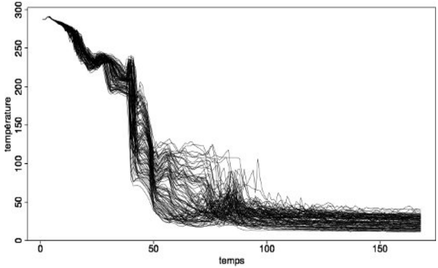


FIG.: Courbes d'origine

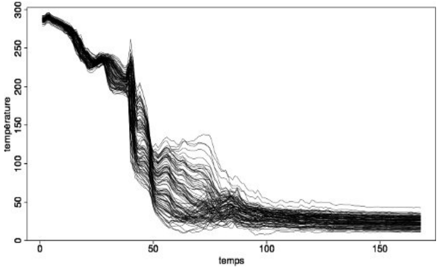


FIG.: ACPF

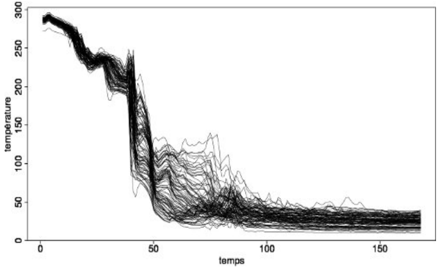


FIG.: LPcaML

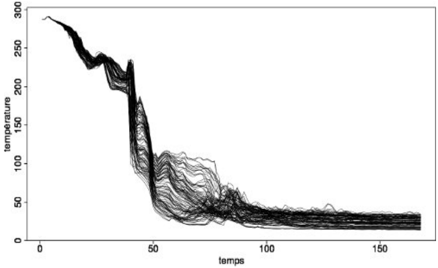


FIG.: kNN (N-W)

5 courbes "représentatives"

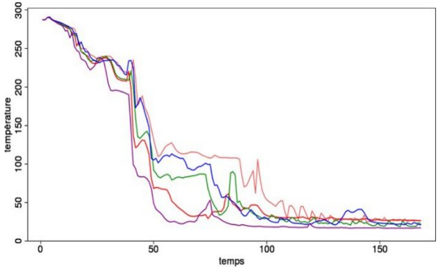


FIG.: Courbes d'origine

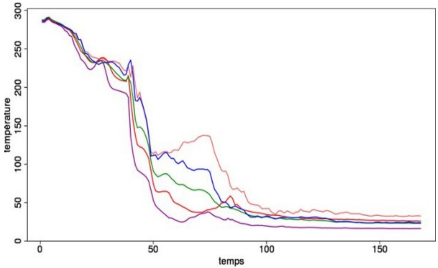


FIG.: ACPF

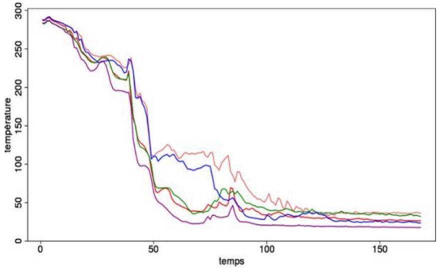


FIG.: LPcaML

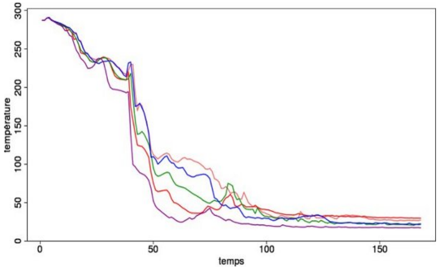


FIG.: kNN (N-W)

Test II - température "difficile"

600 simulations,
11 dimensions en entrée,
1039 points de discrétisation.

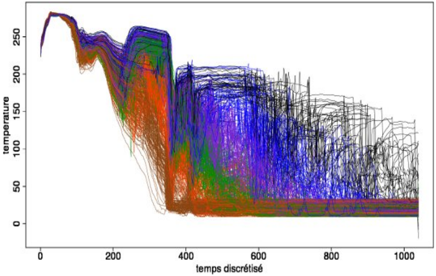


FIG.: Les 600 sorties du code

Test II - température "difficile"

600 simulations,
11 dimensions en entrée,
1039 points de discrétisation.

validation croisée
leave-one-out :

MSE à g., Q_2 à d. ; $\boxed{d = 7}$

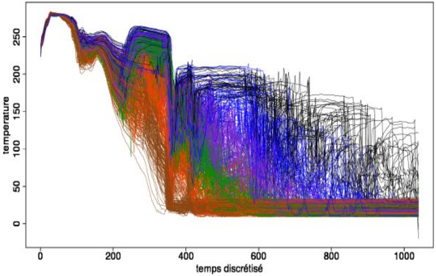


FIG.: Les 600 sorties du code

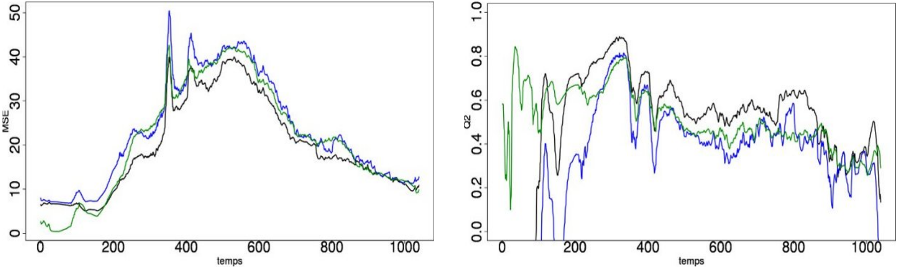


FIG.: Noir : ACP fonctionnelle ; bleu : LPcaML ; vert : Nadaraya-Watson.

Courbes prédites

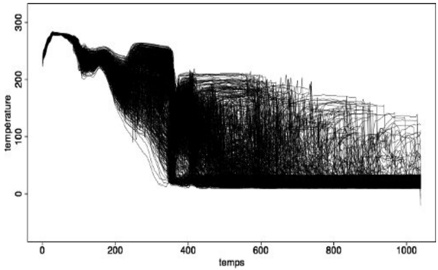


FIG.: Courbes d'origine

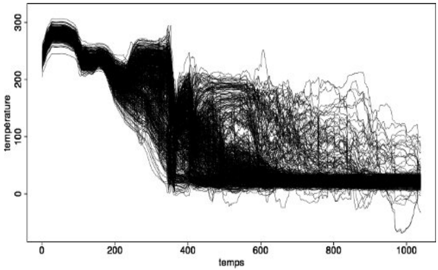


FIG.: ACPF

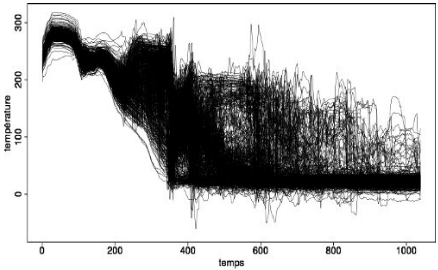


FIG.: LPcaML

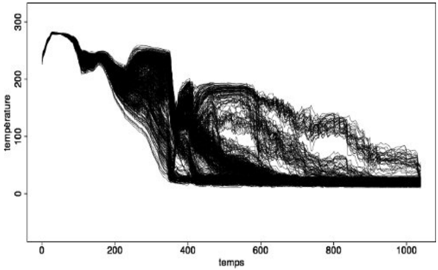


FIG.: kNN (N-W)

10 courbes "représentatives"

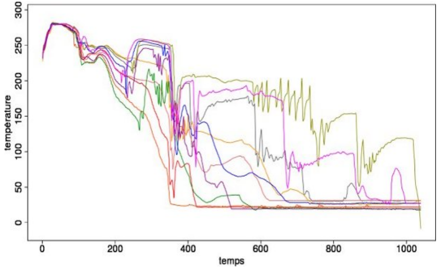


FIG.: Courbes d'origine

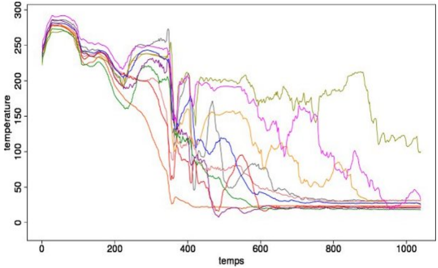


FIG.: ACPF

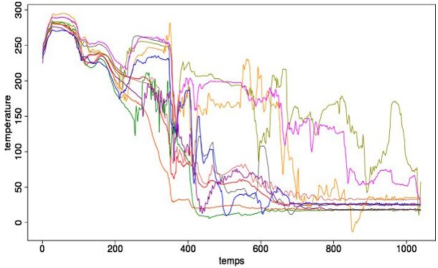


FIG.: LPcaML

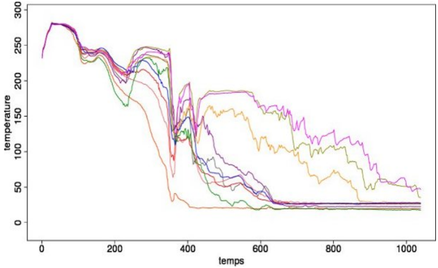


FIG.: kNN (N-W)

Conclusion et perspectives

Modèle assez satisfaisant par rapport aux objectifs industriels.

$\Rightarrow$ aide au projet DDVCV (Durée De Vie des CuVes).

Conclusion et perspectives

Modèle assez satisfaisant par rapport aux objectifs industriels.

$\Rightarrow$ aide au projet DDVCV (Durée De Vie des CuVes).

Réduction de dimension

Linéaire et non-linéaire : complémentaires.

Conclusion et perspectives

Modèle assez satisfaisant par rapport aux objectifs industriels.

$\Rightarrow$ aide au projet DDVCV (Durée De Vie des CuVes).

Réduction de dimension

Linéaire et non-linéaire : complémentaires.

Cadre fonctionnel non linéaire : preuves de convergence ? ...
...encore quelques paramètres à "optimiser" automatiquement.

Conclusion et perspectives

Modèle assez satisfaisant par rapport aux objectifs industriels.

$\Rightarrow$ aide au projet DDVCV (Durée De Vie des CuVes).

Réduction de dimension

Linéaire et non-linéaire : complémentaires.

Cadre fonctionnel non linéaire : preuves de convergence ? . . .
. . .encore quelques paramètres à "optimiser" automatiquement.

Recherches futures : courbes, surfaces principales "fonctionnelles".

Exemple de surface principale en $2D$: