



# Apprentissage de la topologie de signaux multi-capteurs

20/10/2011

Michaël Aupetit  
Ingénieur Chercheur

CEA, LIST,  
Laboratoire Information, Modèles et Apprentissage  
F-91191 Gif-sur-Yvette, France





## Plan

---



- Contexte
- Notion de topologie
- Analyse *in situ*
- Conclusion



## Plan

---



- **Contexte**
- Notion de topologie
- Analyse *in situ*
- Conclusion



## Contexte

---



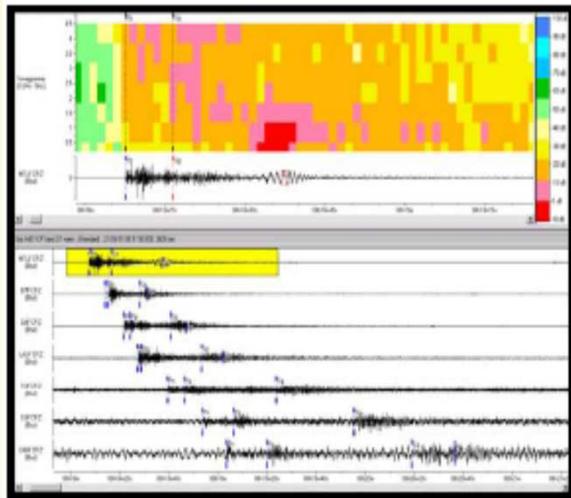
- Objets
  - Systèmes physiques (dynamiques, non linéaires) mesurés par différent capteurs (signaux)
  - Les capteurs fournissent un espace de description des états du système (espace d'état, espace des phases...)
  - Exemples de systèmes physiques
    - Avion, centrale nucléaire, plaques tectoniques, cerveau humain...
- Objectifs
  - **Analyser** : comprendre la structure topologique des états du système au travers des signaux mesurés
  - **Prédire** : utiliser les caractéristiques topologiques extraites pour prédire l'état futur du système



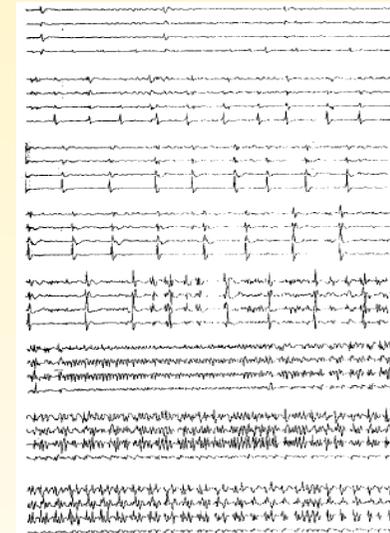
## Contexte



- Exemples de signaux mesurés

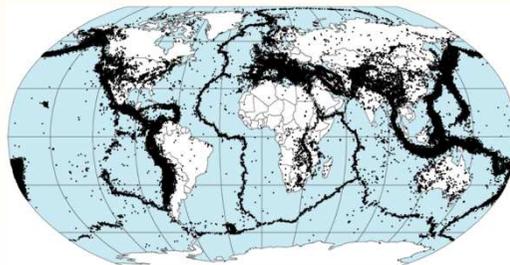


Signaux sismiques [CEA]

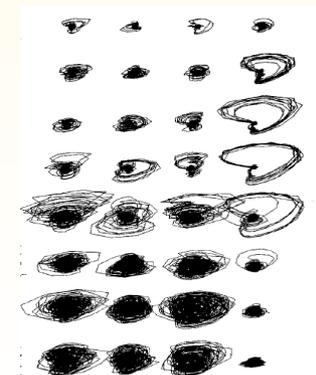


Signaux EEG [Collura et al. 92]

- Exemples de structures topologiques sous-jacentes



Epicentres des séismes [USGS 98]



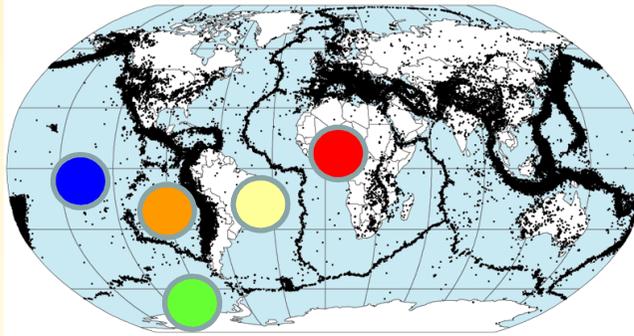
Espace des phases EEG [Collura et al. 92]



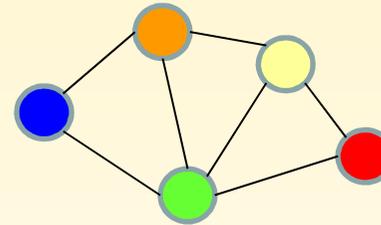
## Contexte



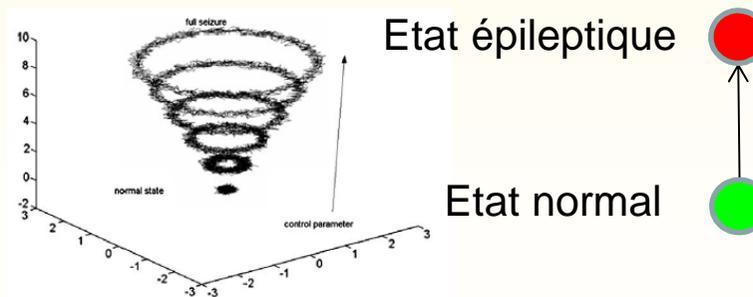
- Intérêt de l'information topologique 1/2
  - Décrire les états et leurs relations de voisinage



Epicentres des séismes [USGS 98]



- Décrire les états et les transitions entre eux



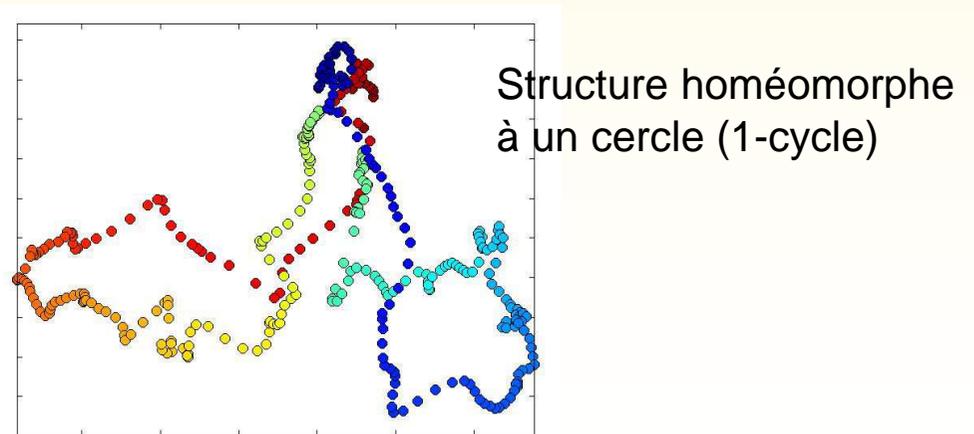
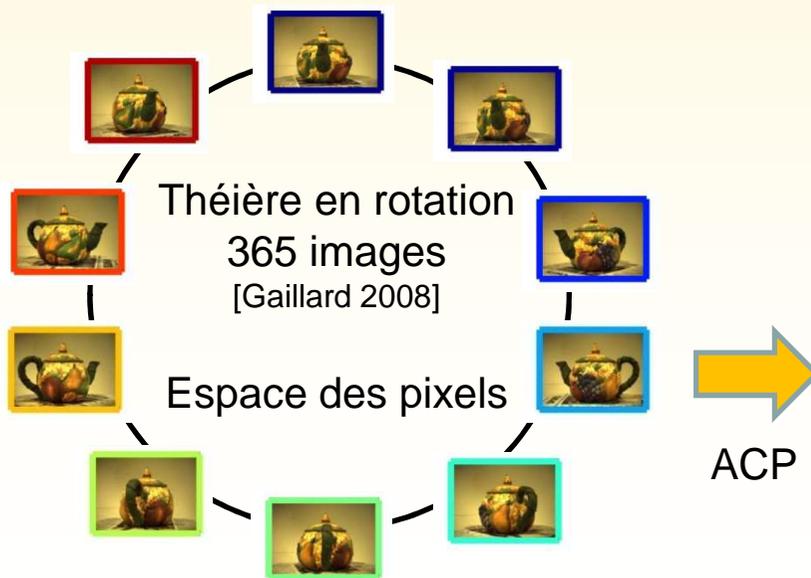
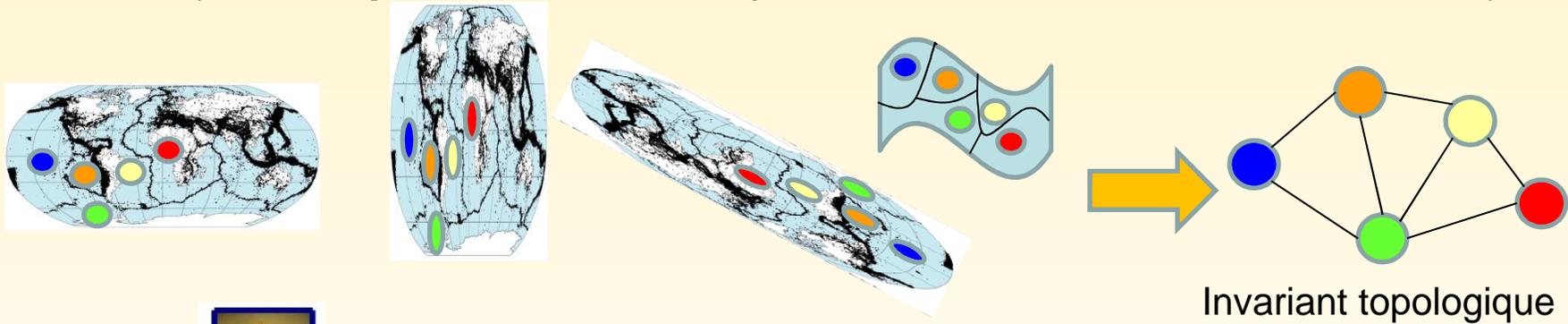
Espace des phases EEG [Kalitzin et al. 2010]



# Contexte



- Intérêt de l'information topologique 2/2
  - **Robuste aux transformations** linéaires et non linéaires  
(homotopies  $\supset$  homéomorphismes  $\supset$  similitudes  $\supset$  isométries)

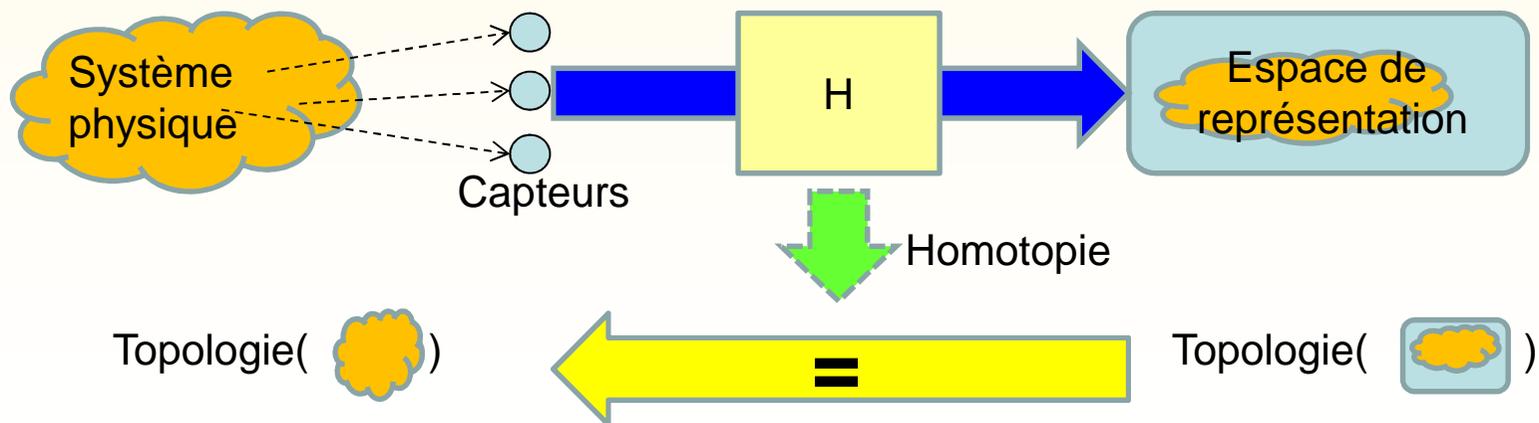




## Contexte



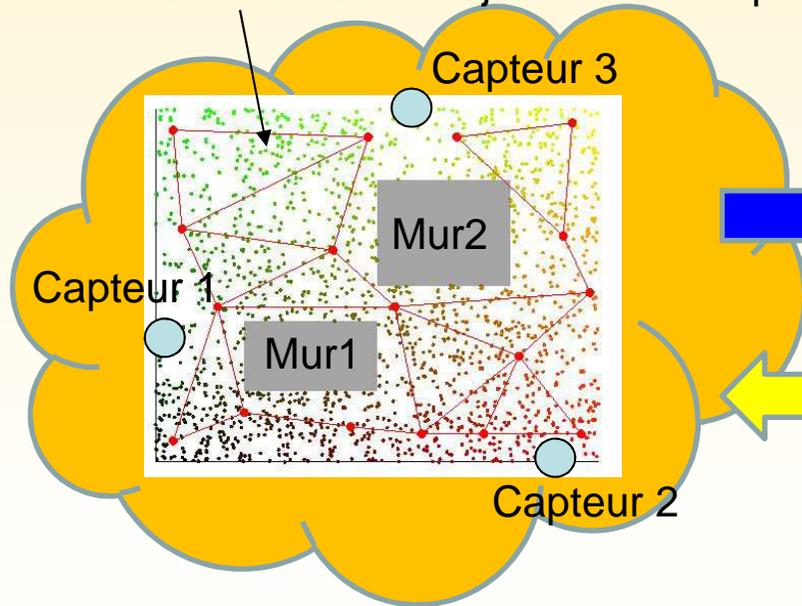
- Conséquence
  - Si la fonction de transfert de la chaîne de mesure réalise une **homotopie**, l'information topologique essentielle (connexité et groupes d'homologie) concernant les états du système, est préservée sous réserve de la qualité de l'échantillonnage (densité suffisante, bruit modéré...)
- Corolaire
  - Pour accéder à la topologie des états du système physique observé, il suffit alors d'extraire la topologie de son image par  $H$  à partir de l'échantillon que son observation génère



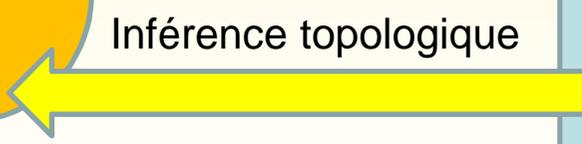
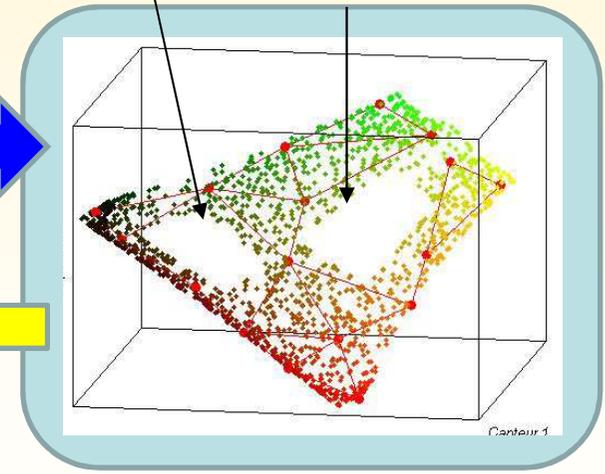
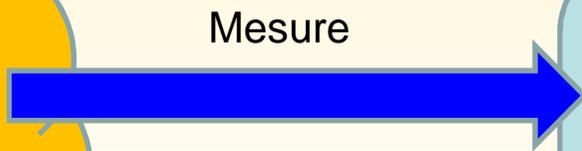


Capteurs

Echantillon de la trajectoire suivie par un robot



Images des murs 1 et 2 dans l'espace des distances aux capteurs

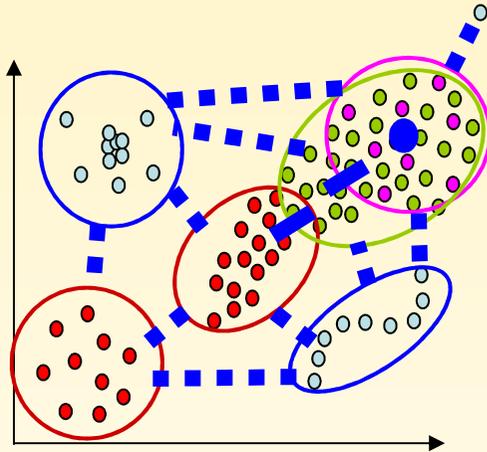




Contexte

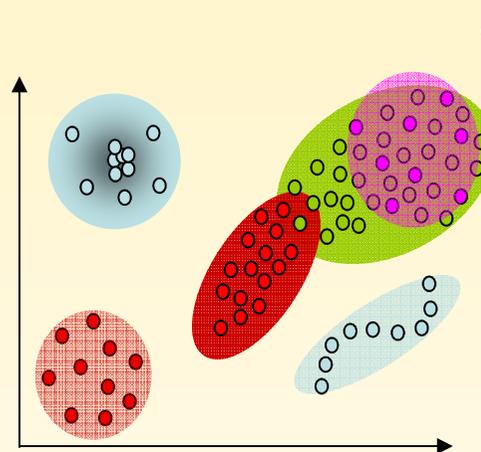


Trois regards complémentaires



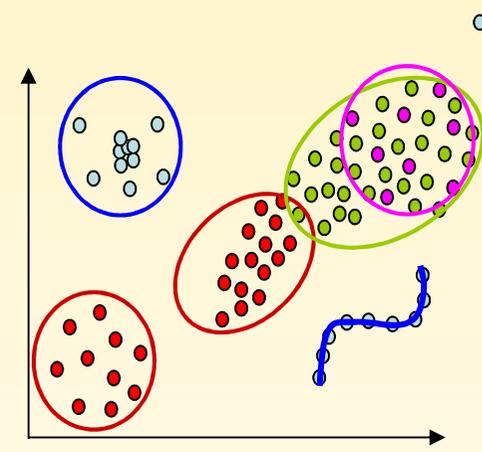
**Topologique**

Caractérise la **structure** de la population



**Statistique**

Caractérise la **densité** de probabilité générée par cette structure



**Géométrique**

Caractérise la **géométrie** de cette structure ou l'échantillon lui-même

**Statistique inférentielle**

Hypothèse a priori sur l'origine des données

On modélise et on infère

**objectivement** (Bayes) ou **subjectivement** (visuellement) des propriétés de la population à partir de l'échantillon

**Statistique descriptive**

Pas d'hypothèse a priori

On quantifie

(min, max, moy, var...) l'échantillon





## Plan

---



- Contexte
- **Notion de topologie**
- Analyse *in situ*
- Conclusion





## Topologie, homéomorphisme et variétés...



- La **Topologie** étudie les **propriétés intrinsèques** d'un objet, **indépendantes du plongement** de cet objet dans un espace « ambiant »
  - intuitivement la dimension intrinsèque et « les » connexités
  - objet = espace compact sous-ensemble de  $\mathbb{R}^d$

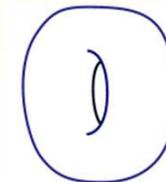
- **Homéomorphisme**

application  $f$  bijective de  $X$  vers  $Y$ , avec  $f$  et  $f^{-1}$  continues.

$X$  et  $Y$  sont **homéomorphes**

$\Leftrightarrow X$  et  $Y$  ont **même topologie**

$\Leftrightarrow X \approx Y$





# Topologie, homéomorphisme et variétés...



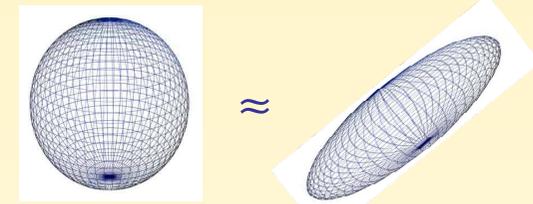
- **d-variété sans bord**

Espace topologique donc chaque point a un voisinage homéomorphe à  $\mathbb{R}^d$ ,  
d est la **dimension intrinsèque**

- **d-variété avec bord**

Espace topologique donc chaque point a un voisinage homéomorphe à  $\mathbb{R}^d$   
et les **points du bord** à  $\mathbb{H}^d = \mathbb{R}^+ \times \mathbb{R}^{d-1}$   
Une d-boule est une variété ayant pour bord une (d-1)-sphère

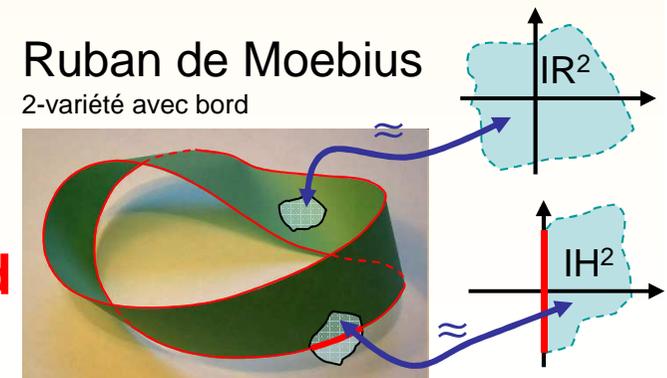
- **bord d'une d-variété**  
= **(d-1)-variété sans bord**



2-sphère  
2-variété sans bord



Tore  
2-variété sans bord





## Classer les objets selon leur topologie... notion d'invariant

---



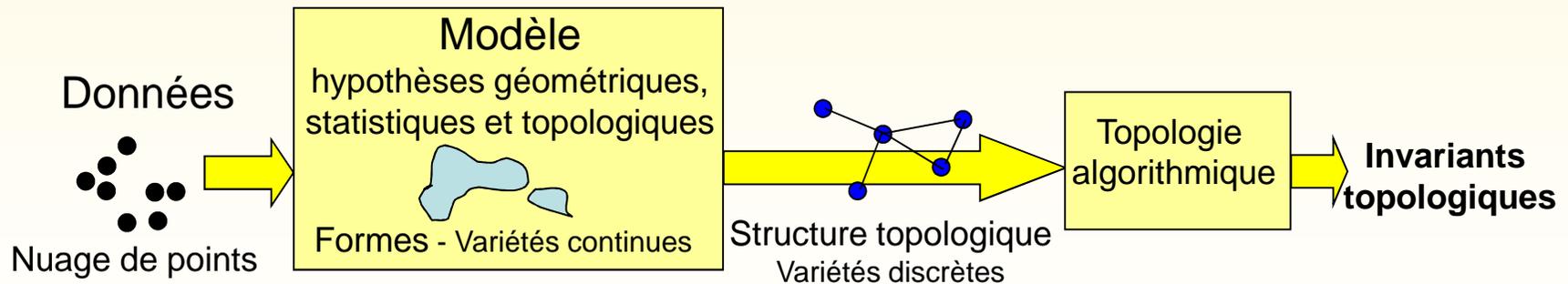
- Invariant
  - **Descripteur** permettant de classer les objets
    - Les objets d'une même classe peuvent différer selon certains aspects (géométrie...), mais le descripteur qui permet de les classer reste *invariant* à ces différences, d'où son nom.
  - Idéalement on cherche à associer à chaque objet des **invariants topologiques** (nombre, groupe...) tels que si les objets ont **même topologie (homéomorphie)**, leurs **invariants** sont **identiques et réciproquement**
  - En pratique, le problème de l'existence dans le cas général, d'invariants topologiques assez puissants pour discriminer deux formes selon l'homéomorphie est **indécidable**.



## Comment calculer les invariants topologiques



- Etant donné deux objets, comment savoir s'ils sont dans la même classe topologique?
- Solution : travailler sur des **objets discrets/combinatoires**, calculables, représentables et manipulables par ordinateur
  - **Graphes**
  - **Complexes simpliciaux**
- Du nuage de points aux invariants topologiques

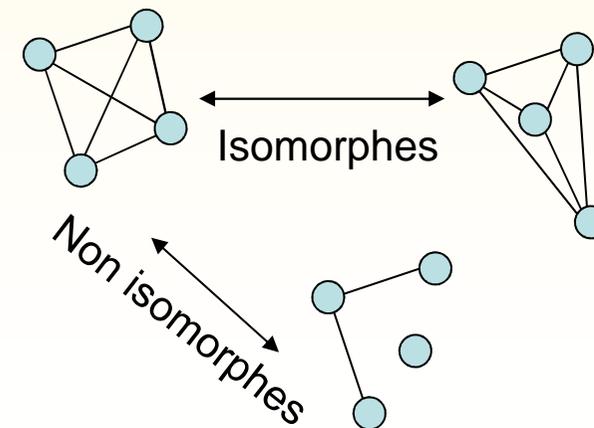
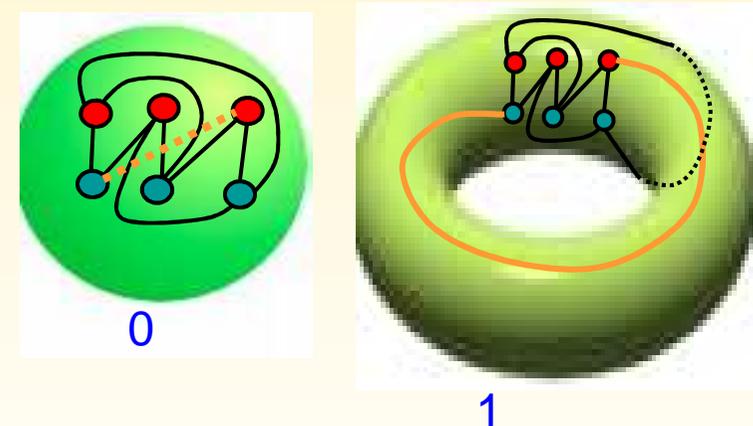
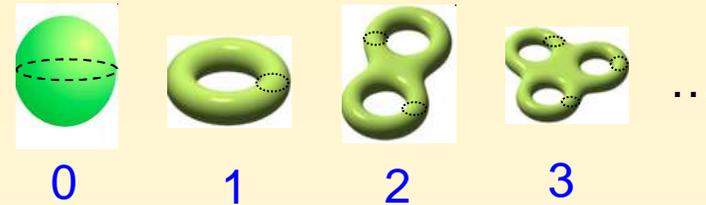




## Exemples d'invariants de graphes



- **Genre** d'une 2-variété (surface) sans bord
  - nombre de tunnels ou anses
- **Genre** d'un graphe
  - plus petit genre de la surface sur laquelle peut-être tracé le graphe sans croisement d'arcs
  - un graphe planaire est de genre 0
- **Isomorphisme**
  - Deux graphes sont isomorphes s'ils ont même structure à une permutation près des sommets
  - Si on les dessine, on peut passer de l'un à l'autre en déplaçant seulement les sommets, sans ajouter ni supprimer d'arc ni de sommet





# Simplexes, complexes, triangulation



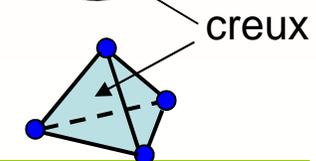
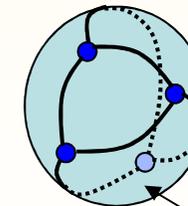
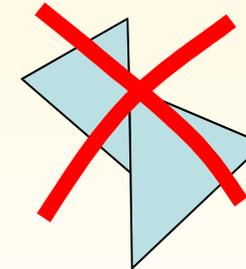
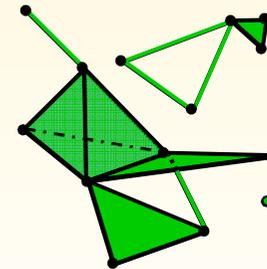
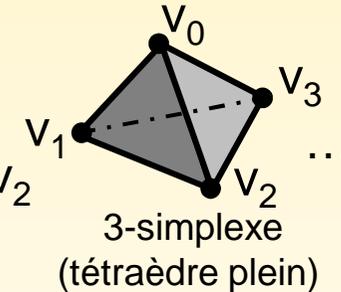
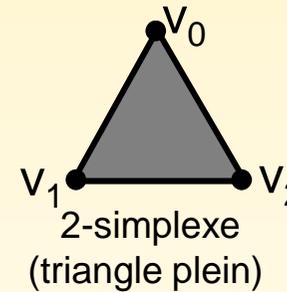
- **k-simplexe**
  - **enveloppe convexe** de  $k+1$  points affinement indépendants de  $\mathbb{R}^d$  ( $d > k-1$ )
  - $k$ -simplexe  $\approx$   $k$ -boule
- **Complexe simplicial  $K$** 
  - Union finie de simplexes telle que
    - Toute face d'un simplexe de  $K$  est dans  $K$
    - L'intersection de deux simplexes de  $K$  est soit vide soit un autre simplexe de  $K$  face commune aux deux
- **Triangulation**
  - La triangulation d'un espace topologique  $X$  est un complexe simplicial  $K$  tel que  $K \approx X$
  - Exemple : Le bord d'un 3-simplexe (un tétraèdre creux) est une triangulation de la 2-sphère (la sphère usuelle)

$V_0$

0-simplexe  
(point)

$V_0$   $V_1$

1-simplexe  
(segment)

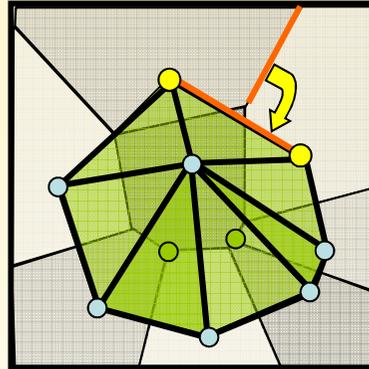




## Exemple de complexe simplicial



- Cellules de Voronoï et complexe simplicial de Delaunay



Graphes de Delaunay  
(dual des cellules de Voronoï)



## Exemples d'invariants de complexes simpliciaux

---



- Nombres de Betti...



## Exemples d'invariants de complexes simpliciaux



- **Nombres de Betti** d'un complexe simplicial  $K$ 
  - $\beta_0(K)$  est le nombre de **composantes connexes** de  $K$
  - $\beta_1(K)$  est le nombre de **tunnels** de  $K$
  - $\beta_2(K)$  est le nombre de **vides** de  $K$
  - ... généralisable en dimension  $k > 2$

2-sphère



$$\begin{aligned}\beta_0(K) &= 1 \\ \beta_1(K) &= 0 \\ \beta_2(K) &= 1\end{aligned}$$

3-boule



$$\begin{aligned}1 \\ 0 \\ 0\end{aligned}$$

2-tore



$$\begin{aligned}1 \\ 2 \\ 1\end{aligned}$$

2-tore plein

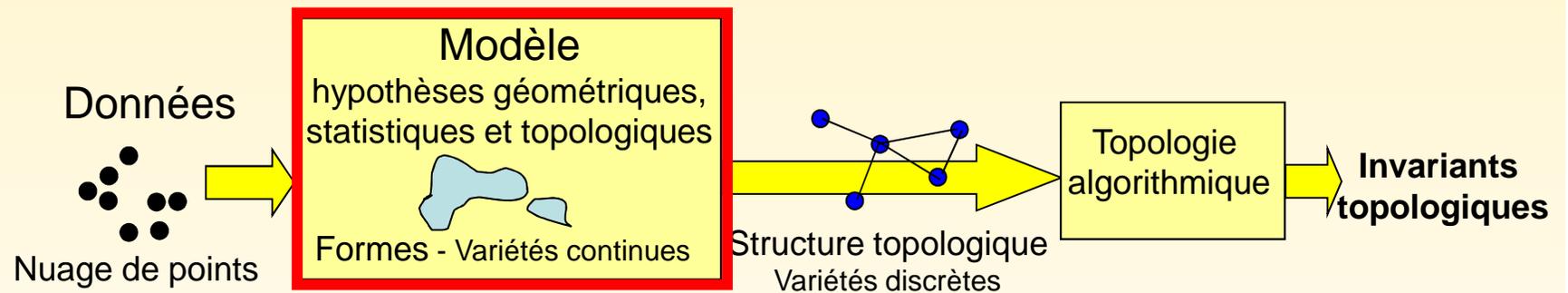


$$\begin{aligned}1 \\ 1 \\ 0\end{aligned}$$



## Comment calculer les invariants topologiques

- Du nuage de points aux invariants topologiques



Nous proposons un modèle statistique...



## Plan

---



- Contexte
- Notion de topologie
- **Analyse *in situ***
- Conclusion



## Analyse *in situ* > Introduction

---



- Comment extraire des informations sur la topologie d'un nuage de points de  $\mathbb{R}^D$  ?
- Solution : un modèle génératif...



## Analyse *in situ* > Hypothèses générales sur la génération des données

---



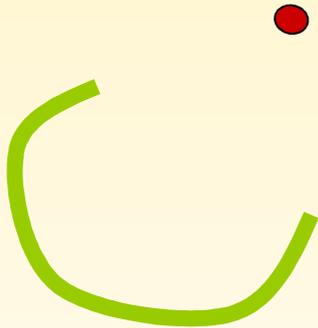
Des **variétés génératrices**  
étiquetées inconnues  
points,





## Analyse *in situ* > Hypothèses générales sur la génération des données

---



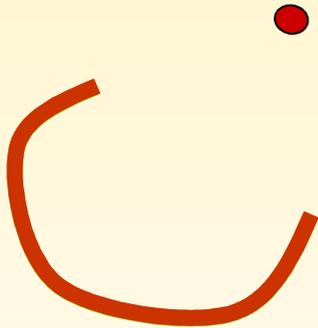
Des **variétés génératrices**  
étiquetées inconnues  
points, courbes





## Analyse *in situ* > Hypothèses générales sur la génération des données

---



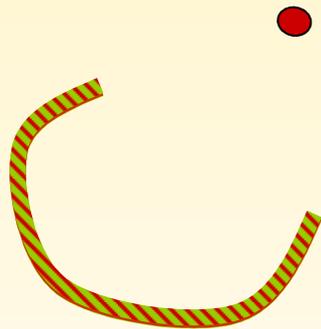
Des **variétés génératrices**  
étiquetées inconnues  
points, courbes superposées





## Analyse *in situ* > Hypothèses générales sur la génération des données

---



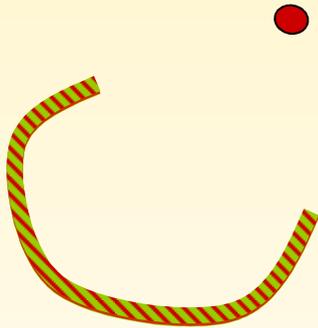
Des **variétés génératrices**  
étiquetées inconnues  
points, courbes superposées



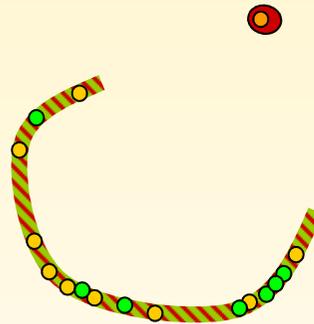


## Analyse *in situ* > Hypothèses générales sur la génération des données

---



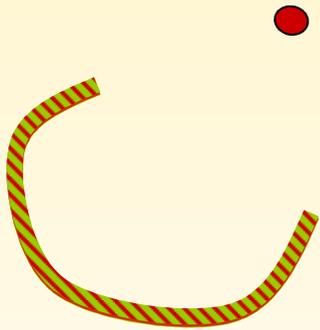
Des **variétés génératrices**  
étiquetées inconnues  
points, courbes superposées



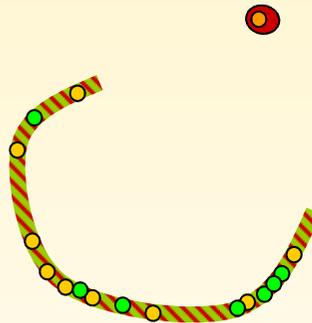
...desquelles sont tirées  
des données avec une  
**fdp** inconnue



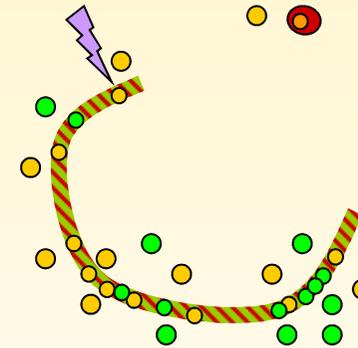
## Analyse *in situ* > Hypothèses générales sur la génération des données



Des **variétés génératrices**  
étiquetées inconnues  
points, courbes superposées



...desquelles sont tirées  
des données avec une  
**fdp** inconnue

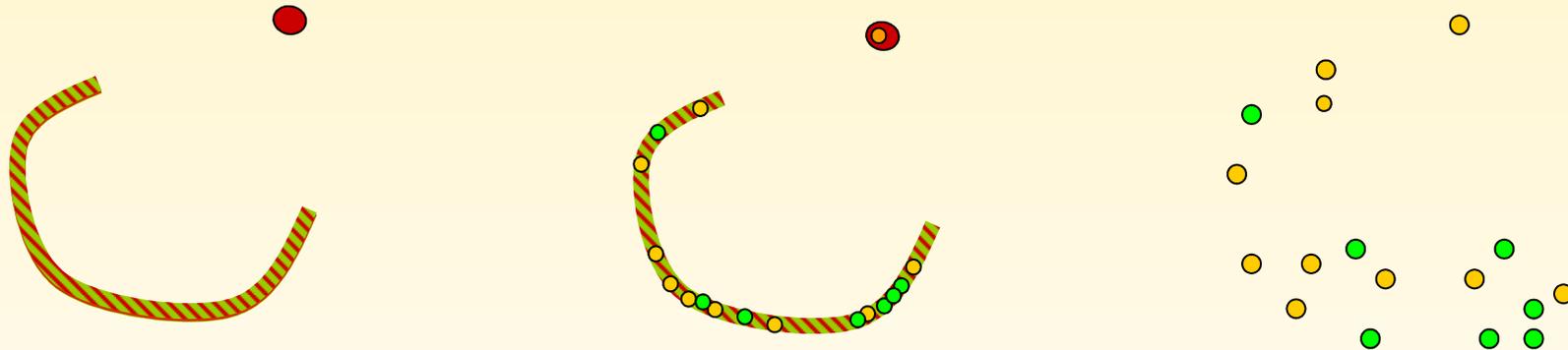


...corrompues par un  
**bruit** de nature inconnue,  
menant aux observation





## Analyse *in situ* > Hypothèses générales sur la génération des données



Des **variétés génératrices**  
étiquetées inconnues  
points, courbes superposées

...desquelles sont tirées  
des données avec une  
**fdp** inconnue

...corrompues avec un  
**bruit** de nature inconnue,  
menant aux observation

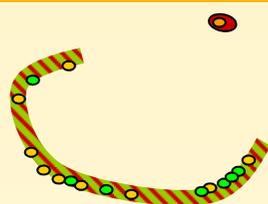
**L'objectif est de modéliser les variétés génératrices  
à partir des données observées  
de manière à en extraire les caractéristiques topologiques**



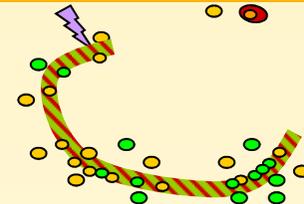
## Analyse *in situ* > Hypothèses simplifiées



Des **variétés génératrices**  
étiquetées inconnues



...desquelles sont tirées  
des données avec une  
**pdf** inconnue



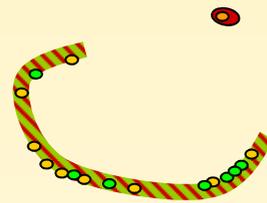
...corrompues avec un  
**bruit** de nature inconnue  
menant aux observations



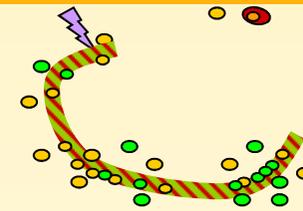
## Analyse *in situ* > Hypothèses simplifiées



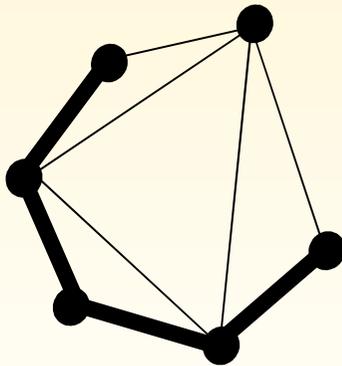
Des **variétés génératrices**  
étiquetées inconnues



...desquelles sont tirées  
des données avec une  
**pdf** inconnue



...corrompues avec un  
**bruit** de nature inconnue  
menant aux observations



La variété est proche du graphe de Delaunay  
engendré par des « prototypes ».

$$p(x, c) = \sum_{j \in J}$$

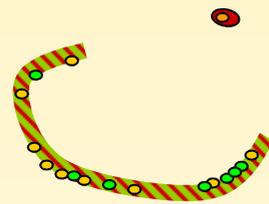




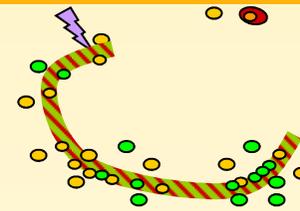
# Analyse *in situ* > Hypothèses simplifiées



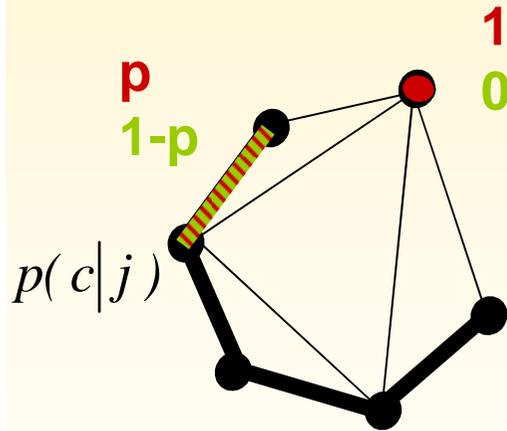
Des **variétés génératrices**  
étiquetées inconnues



...desquelles sont tirées  
des données avec une  
**pdf** inconnue



...corrompues avec un  
**bruit** de nature inconnue  
menant aux observations



La variété est proche du graphe de Delaunay engendré par des « prototypes ».

Les variétés « élémentaires » peuvent générer la classe  $c$  avec une probabilité donnée

$$p(x, c) = \sum_{j \in J} p(c|j)$$

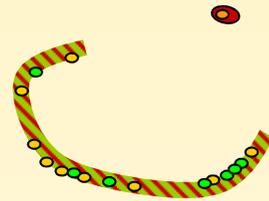




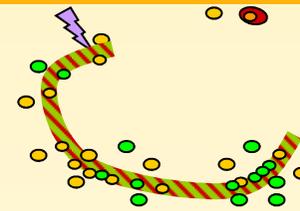
# Analyse *in situ* > Hypothèses simplifiées



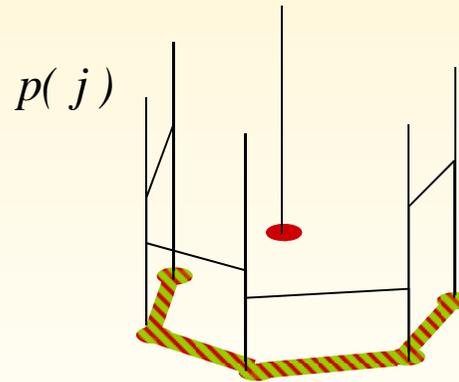
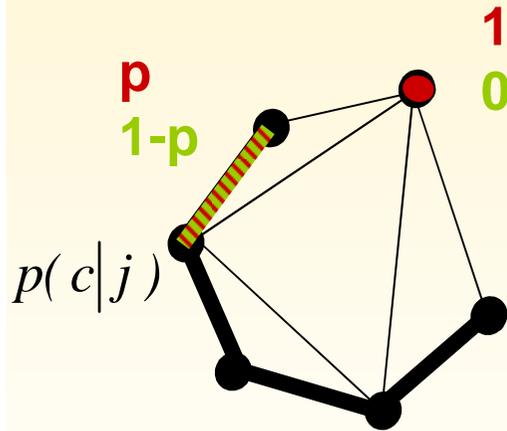
Des **variétés génératrices**  
étiquetées inconnues



...desquelles sont tirées  
des données avec une  
**pdf** inconnue



...corrompues avec un  
**bruit** de nature inconnue  
menant aux observations



La variété est proche du graphe de Delaunay  
engendré par des « prototypes ».

Les variétés « élémentaires » peuvent générer  
la classe  $c$  avec une probabilité donnée

...on associe à chaque variété  
une pdf uniforme et pondérée

$$p(x, c) = \sum_{j \in J} p(j) p(c|j)$$

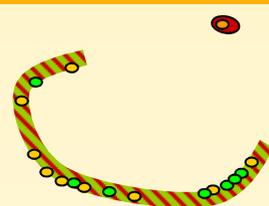




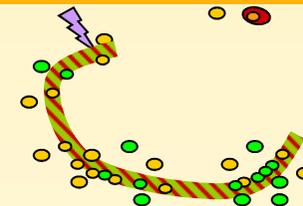
# Analyse *in situ* > Hypothèses simplifiées



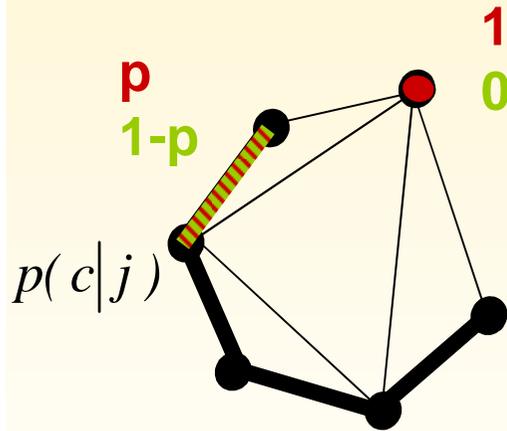
Des **variétés génératrices** étiquetées inconnues



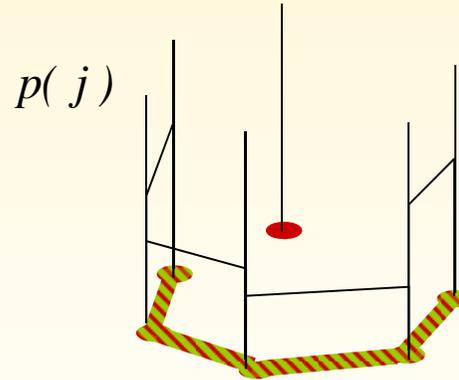
...desquelles sont tirées des données avec une **pdf inconnue**



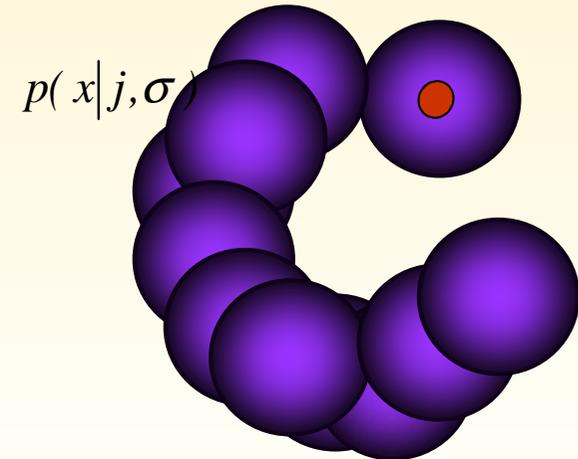
...corrompues avec un **bruit** de nature inconnue menant aux observations



La variété est proche du graphe de Delaunay engendré par des « prototypes ».  
 Les variétés « élémentaires » peuvent générer la classe  $c$  avec une probabilité donnée



...on associe à chaque variété une pdf uniforme et pondérée



...convoluée à une bruit gaussien isovarié

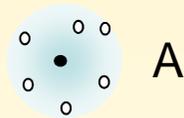
$$p(x, c) = \sum_{j \in J} p(j) p(c|j) p(x|j, \sigma)$$

$$\begin{cases} \sum_j p(j) = 1 & p(j) \geq 0 \\ \sum_c p(c|j) = 1 & p(c|j) \geq 0 \end{cases}$$



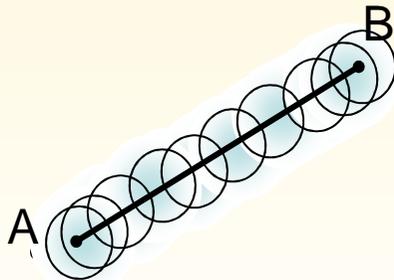


## Comment définir un modèle génératif basé sur des points et des segments ?



*Point Gaussien*

$$p(x|A, \sigma) = (2\pi\sigma^2)^{-\frac{D}{2}} \exp\left(-\frac{(x-A)^2}{2\sigma^2}\right)$$



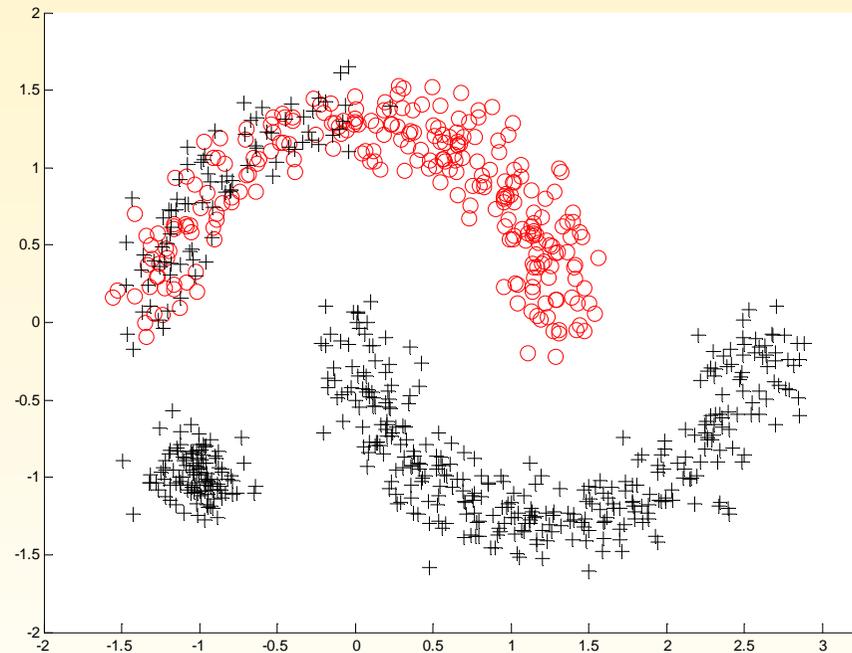
*Segment Gaussien*

$$p(x|[AB], \sigma) = \int_{[AB]} p(x|v, \sigma) dv$$

peut être exprimée à l'aide de la  
fonction « erf »



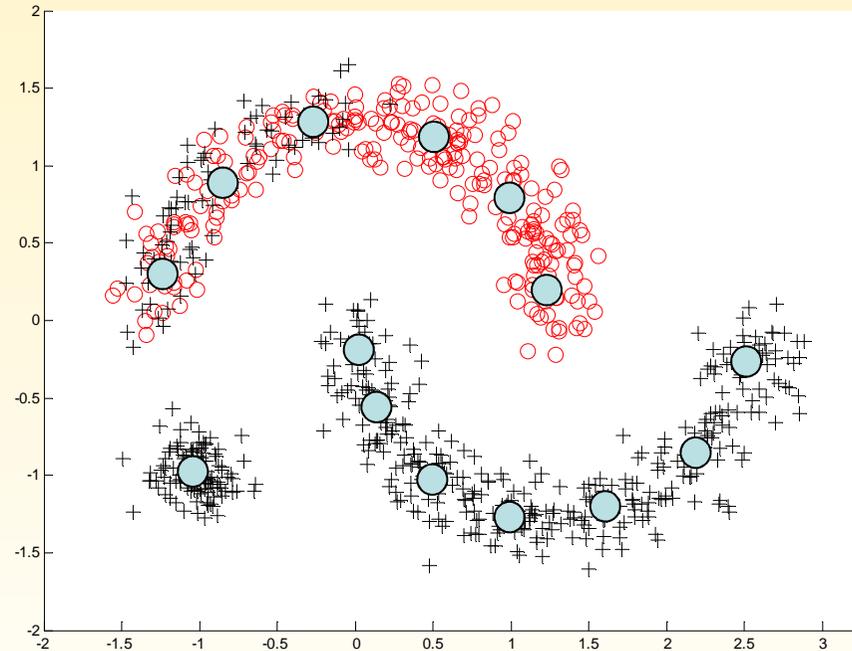
## 1. Initialisation



Placement des  $N$  prototypes avec un modèle de mélange



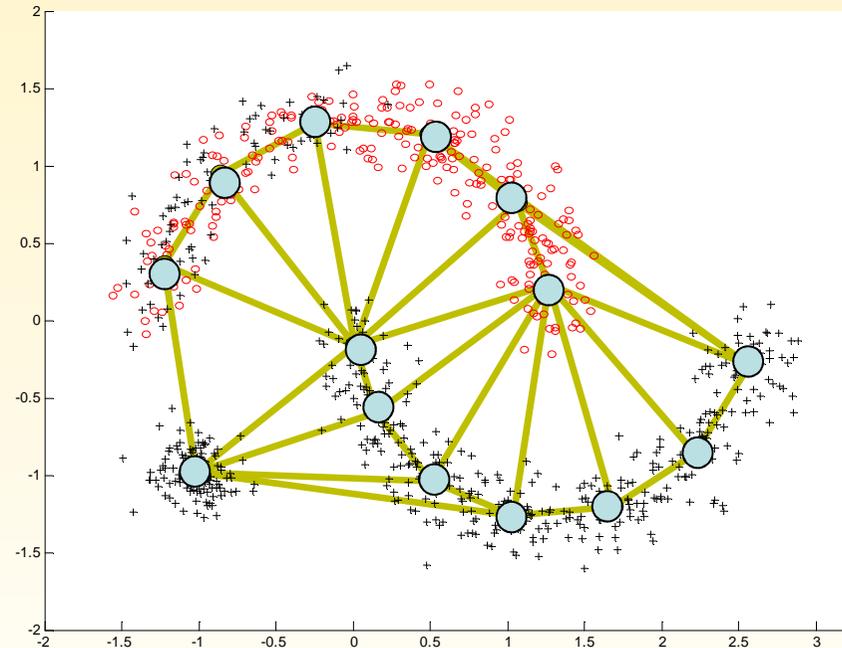
## 1. Initialisation



Placement des  $N$  prototypes avec un modèle de mélange



## 1. Initialisation

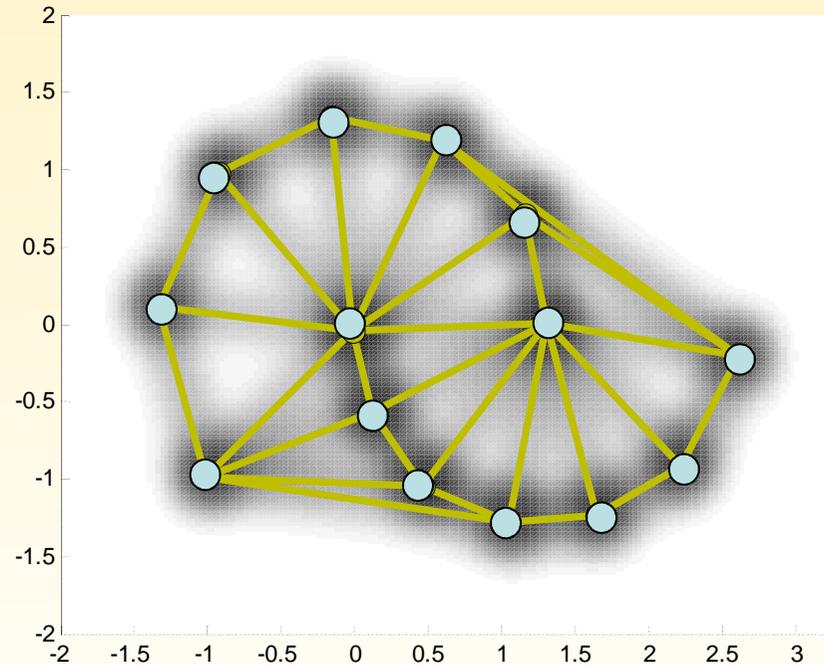


Placement des  $N$  prototypes avec un modèle de mélange

Et construction **du graphe de Delaunay des prototypes  $O(DN^3)$**



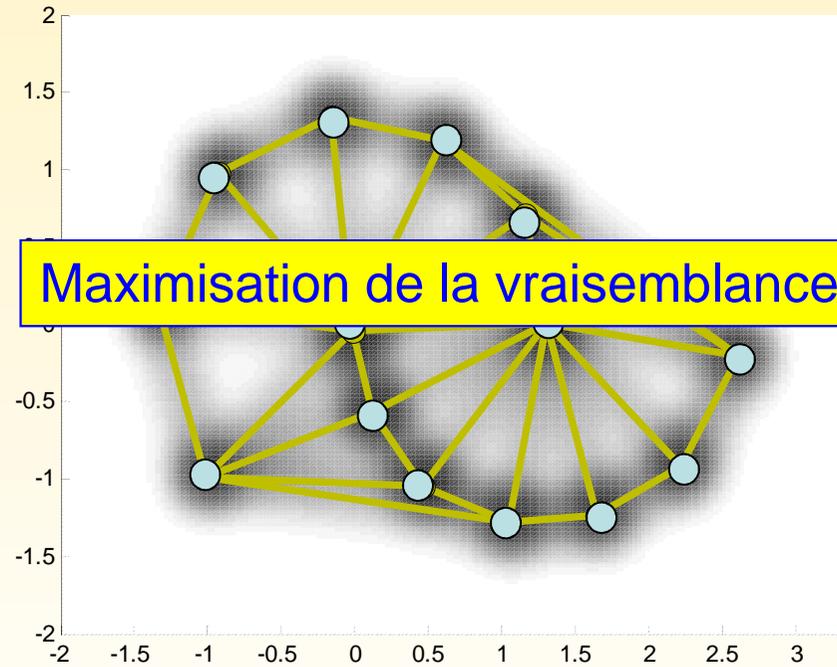
## 1. Initialisation



**Initialisation du graphe génératif  
(équiprobabilité des poids des variétés  
élémentaires et équiprobabilité des classes)**

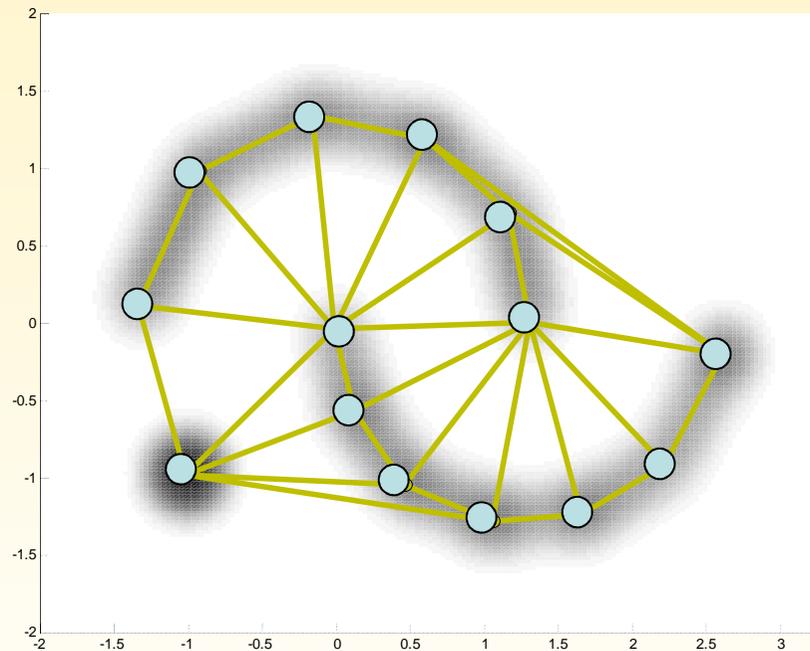


## 2. Apprentissage





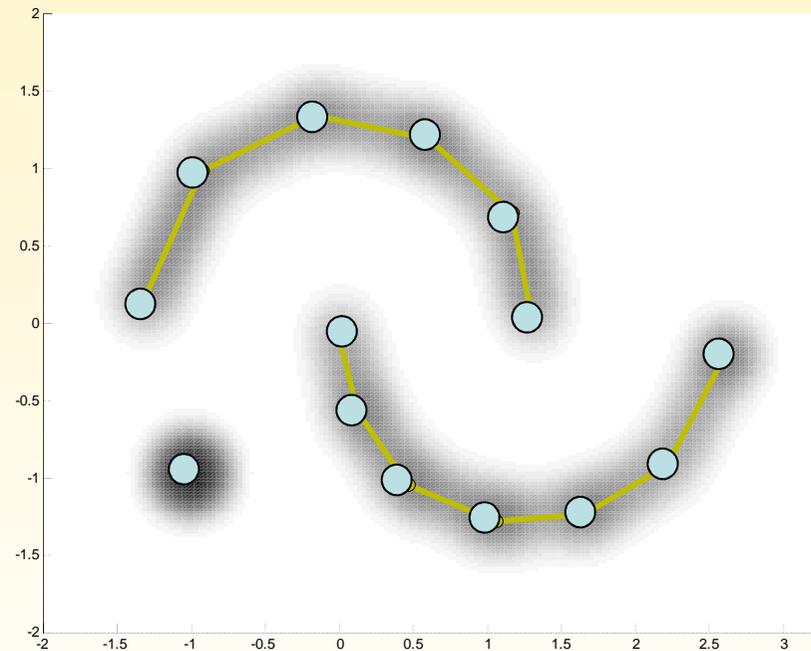
### 3. Après l'apprentissage



Après l'apprentissage, certaines variétés élémentaires ont **un poids nul ou quasi-nul**.



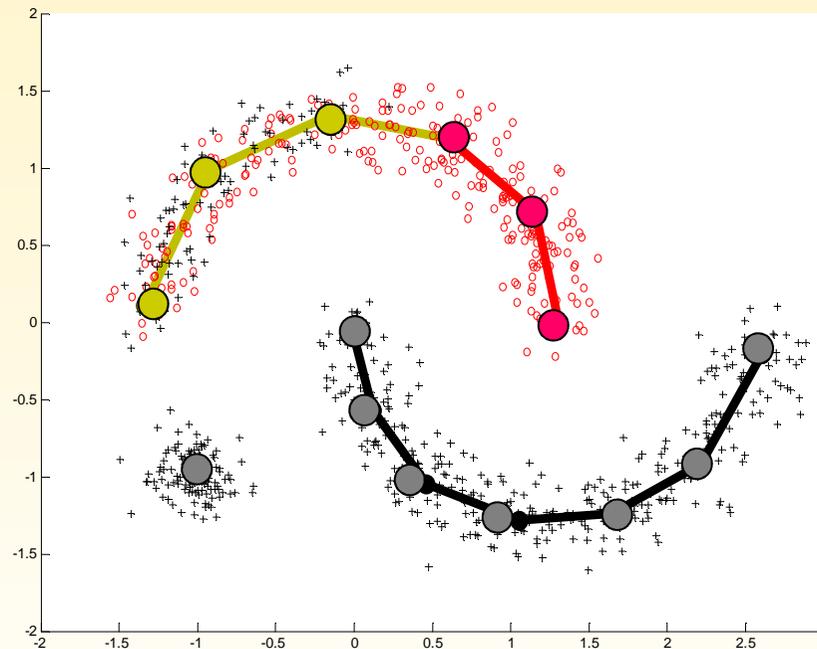
### 3. Après l'apprentissage



Après l'apprentissage, certaines variétés élémentaires ont **un poids nul ou quasi-nul**. Elles n'expliquent pas les données et sont donc **supprimées du graphe initial**



### 3. Après l'apprentissage



Les variétés élémentaires restantes sont associées à leur(s) classe(s)

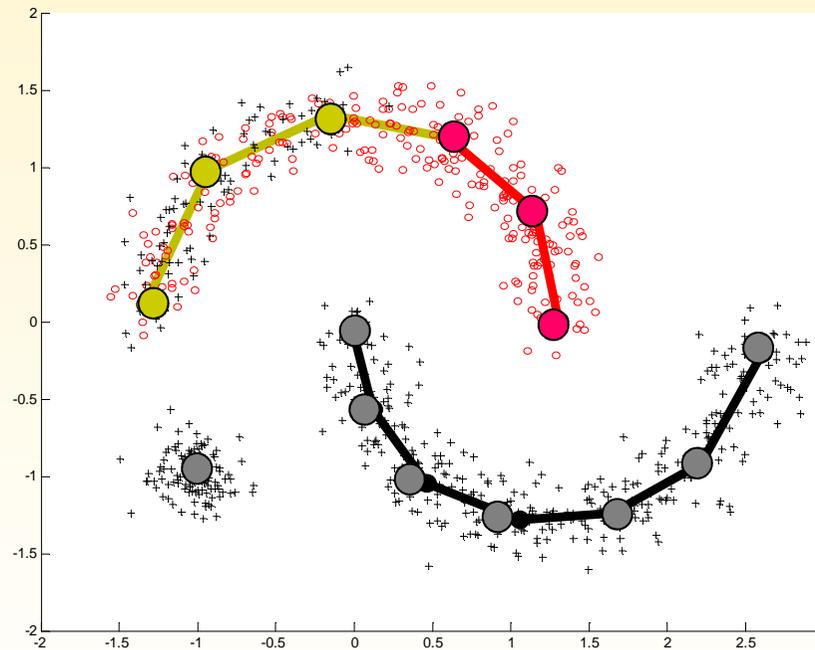
Le critère de pénalisation BIC est utilisé pour sélectionner le bon nombre de prototypes et le bon seuil d'élagage



## Analyse *in situ* > Le graphe des classes



- On construit un graphe synthétisant l'information topologique

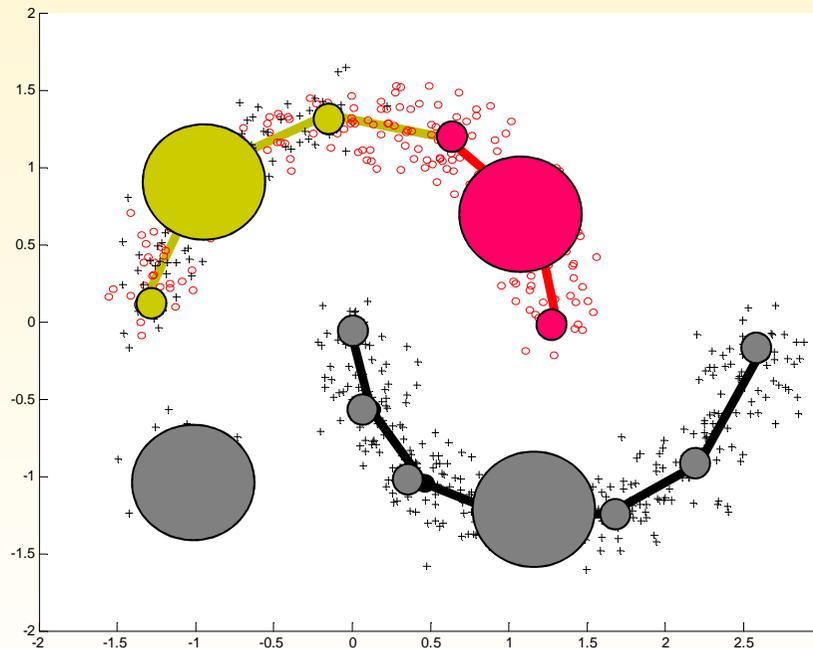




## Analyse *in situ* > Le graphe des classes



- On construit un graphe synthétisant l'information topologique



Connexité intra-classe et inter-classe :  
Un sommet par composante connexe uni-classe

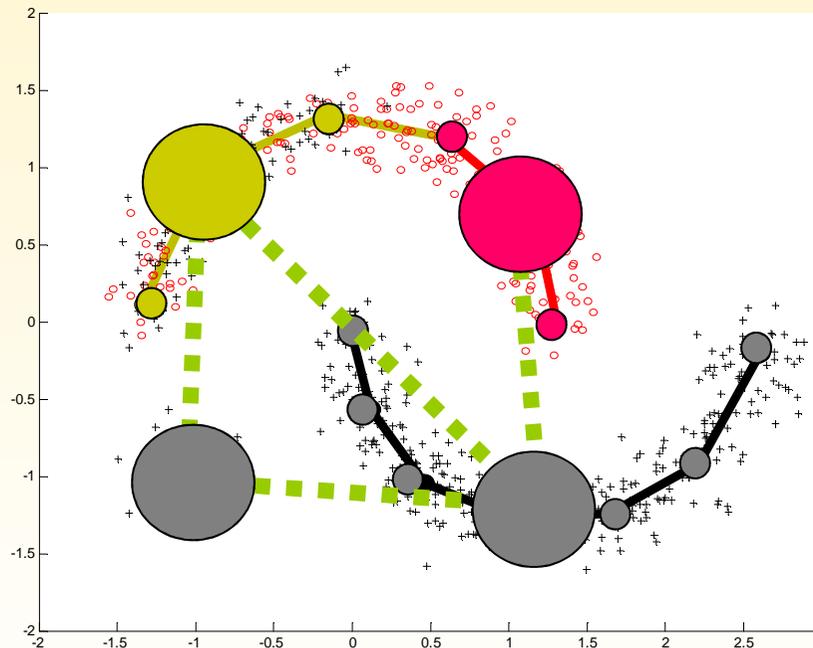




## Analyse *in situ* > Le graphe des classes



- On construit un graphe synthétisant l'information topologique



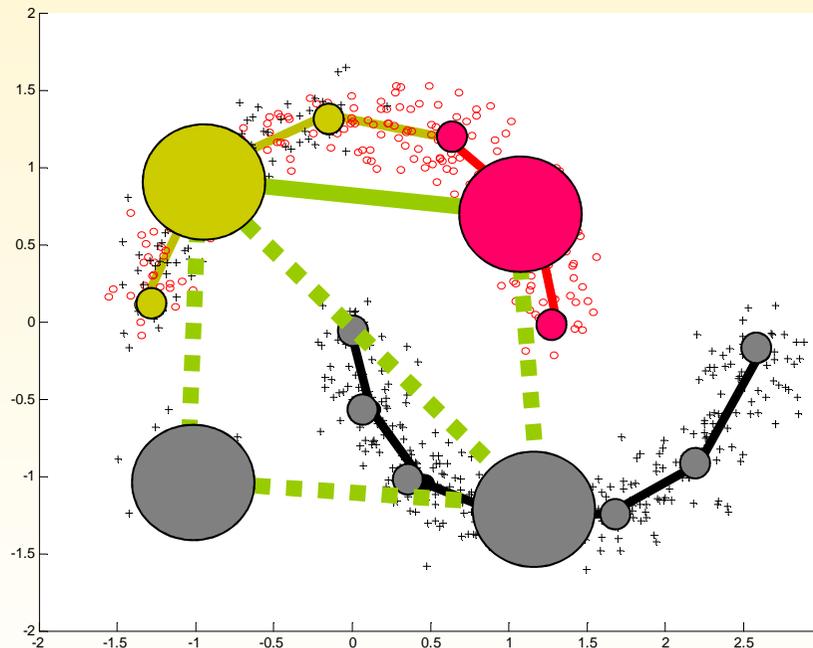
**Delaunay-adjacence :**  
un lien pointillé entre composantes initialement connectées



## Analyse *in situ* > Le graphe des classes



- On construit un graphe synthétisant l'information topologique



### Densité-adjacence :

Un lien en trait continu pour les classes densité-adjacentes



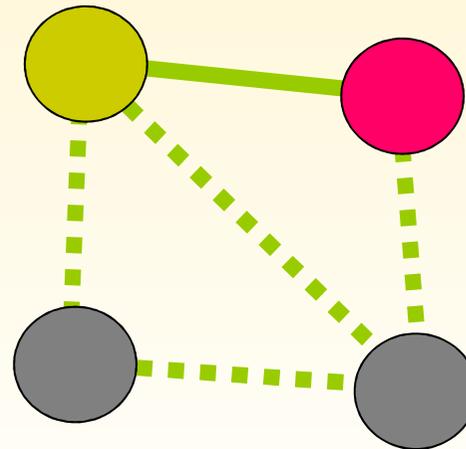


## Analyse *in situ* > Le graphe des classes



- On construit un graphe synthétisant l'information topologique

### Le graphe des classes



Permet une **visualisation *in situ*** car sans distorsion de la connexité **authentique** telle qu'elle est (au sens du modèle) dans l'espace d'origine multidimensionnel

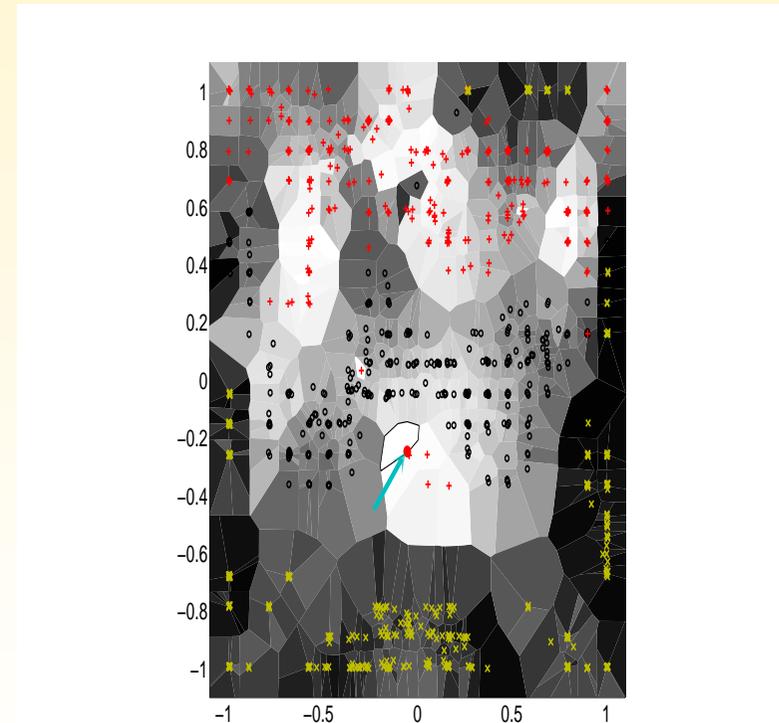
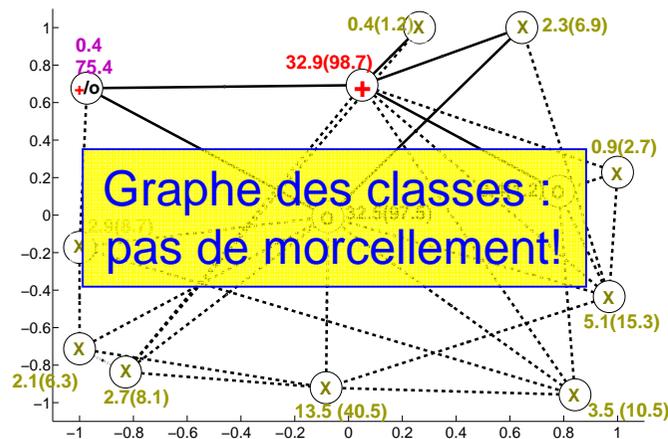
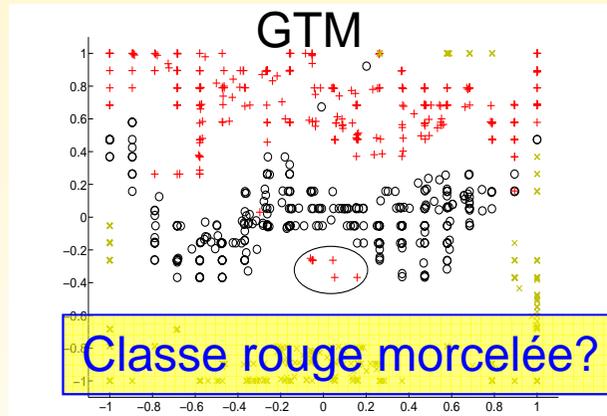


## Analyse *in situ* > Application données Oil-Flow



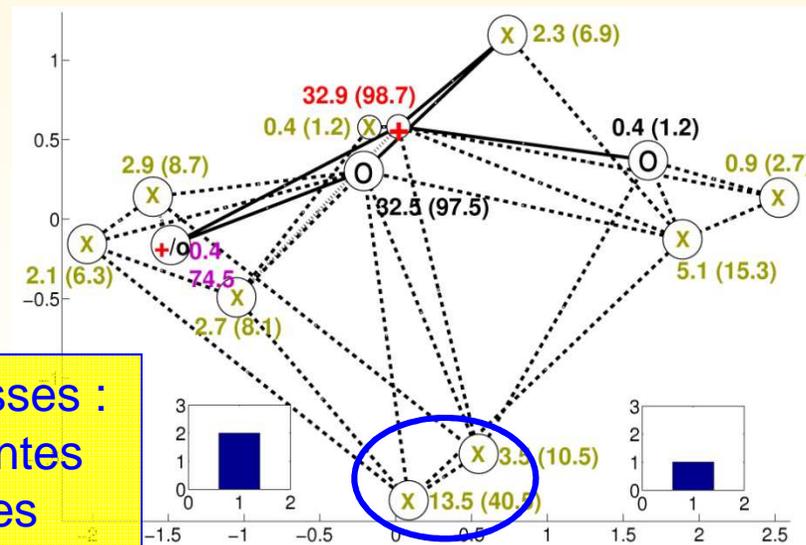
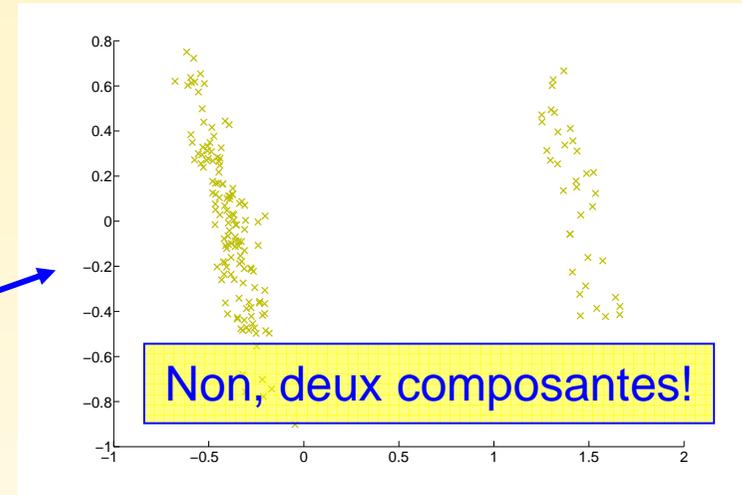
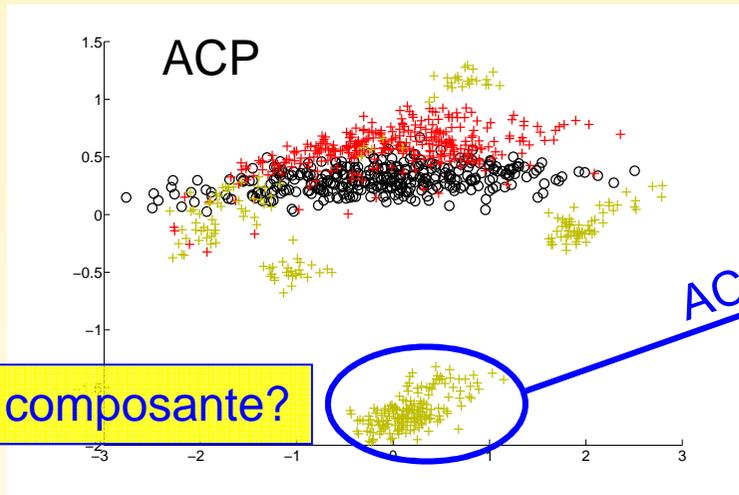
### Données Oil-Flow (Bishop)

- 1000 points de  $IR^{12}$  issues de la mesure par 12 capteurs de la configuration d'un mélange huile-eau-gaz pouvant prendre 3 configurations différentes dans un pipeline.
- Projections par GTM, ACP et ACC





# Analyse *in situ* > Application données Oil-Flow

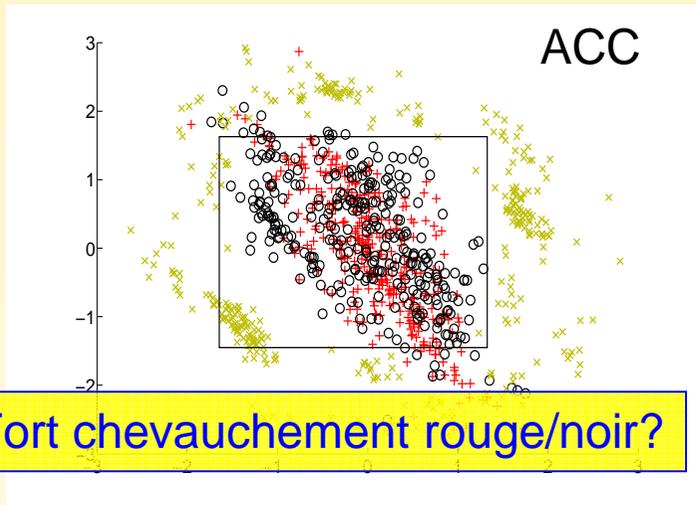


Graphe des classes :  
deux composantes  
non connexes

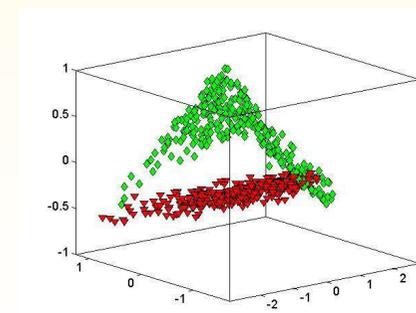
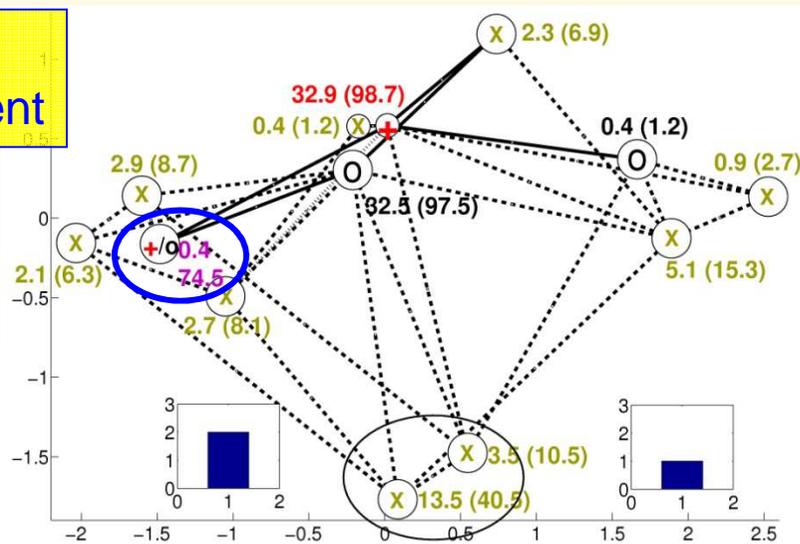




# Analyse *in situ* > Application données Oil-Flow



Graphe des classes :  
Très faible chevauchement



Vérité (ACP)





## Analyse exploratoire de données non-étiquetées

---



- **L'ensemble de données** : série de 365 images d'un objet en rotation  
⇒ 1 image = 1 donnée en 192 dimensions





## Analyse exploratoire de données non-étiquetées

---



- **L'ensemble de données** : série de 365 images d'un objet en rotation  
⇒ 1 image = 1 donnée en 192 dimensions





## Analyse exploratoire de données non-étiquetées

---



- **L'ensemble de données** : série de 365 images d'un objet en rotation  
⇒ 1 image = 1 donnée en 192 dimensions

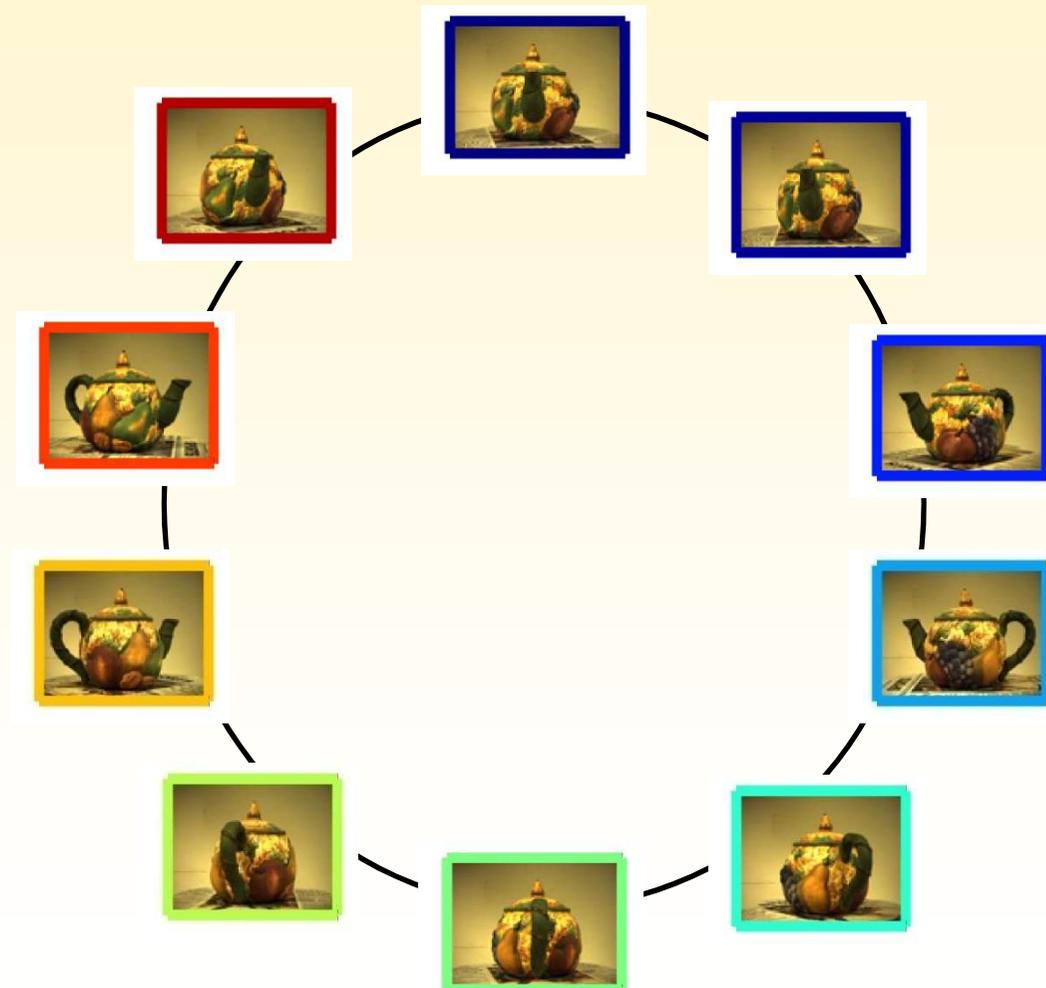




## Analyse exploratoire de données non-étiquetées



- **L'ensemble de données** : série de 365 images d'un objet en rotation  
⇒ 1 degré de liberté : la variété principale est homéomorphe à un cercle

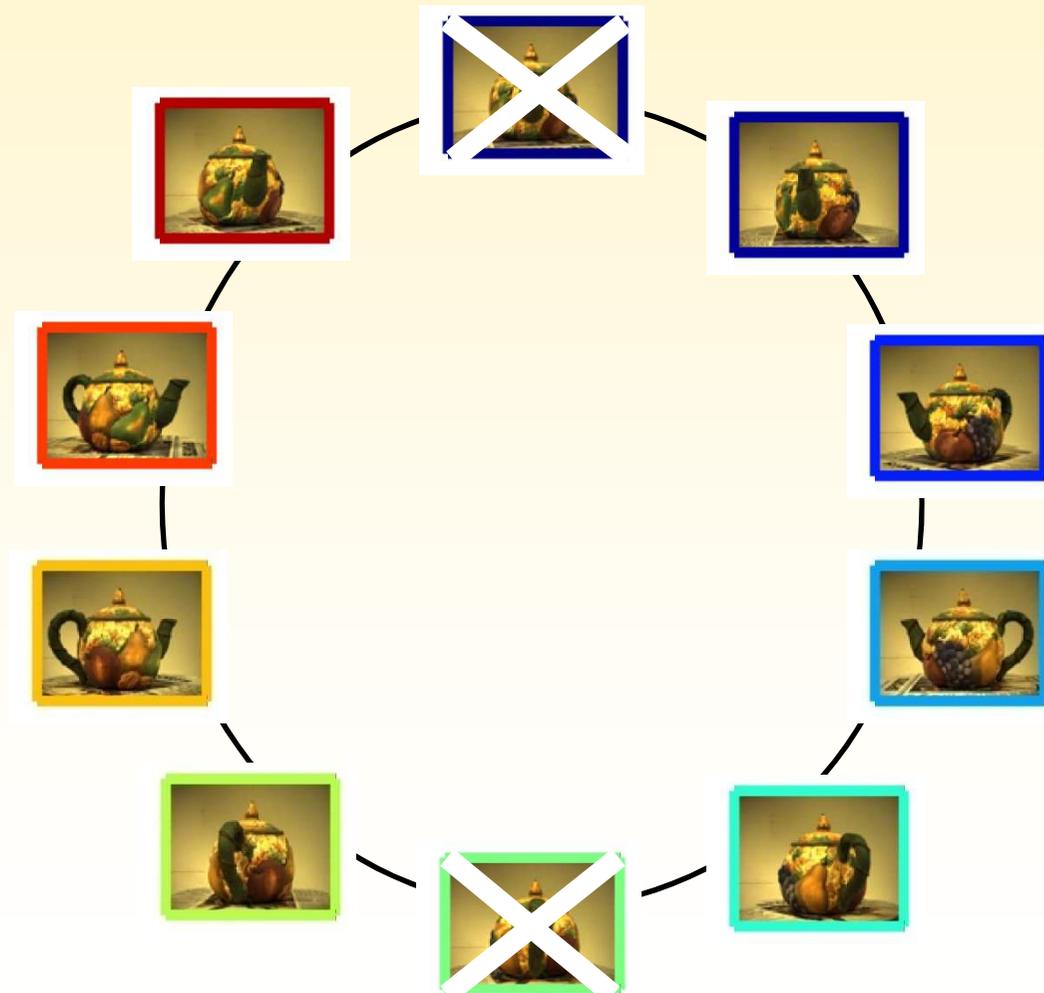




## Analyse exploratoire de données non-étiquetées



- **L'ensemble de données** : série de 365 images d'un objet en rotation  
⇒ 1 degré de liberté : la variété principale est homéomorphe à un cercle

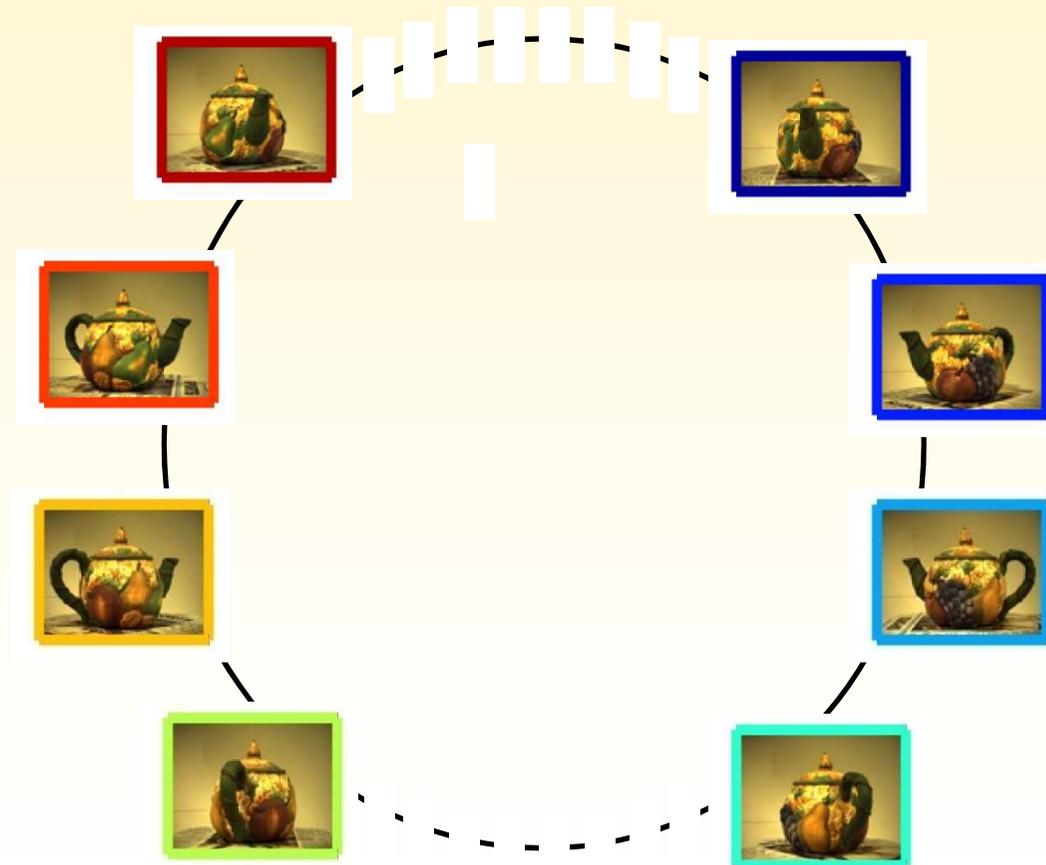




## Analyse exploratoire de données non-étiquetées



- **L'ensemble de données** : série de 365 images d'un objet en rotation  
⇒ 2 variétés déconnectées homéomorphes à un segment

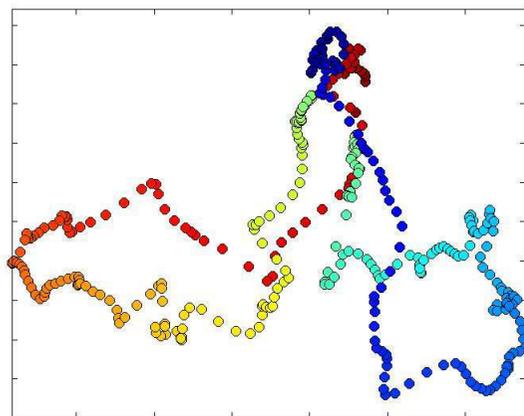




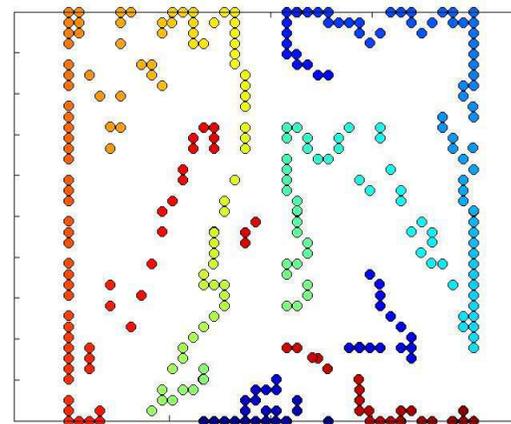
## Analyse exploratoire de données non-étiquetées



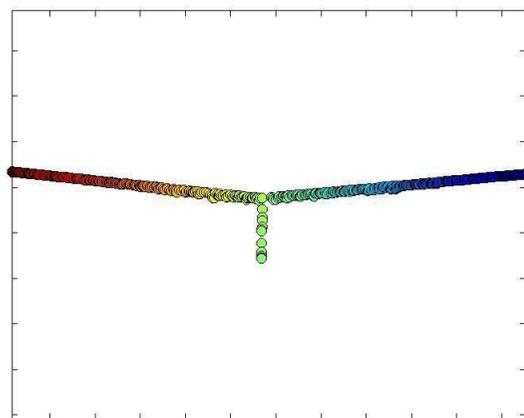
- Projection de l'ensemble de données



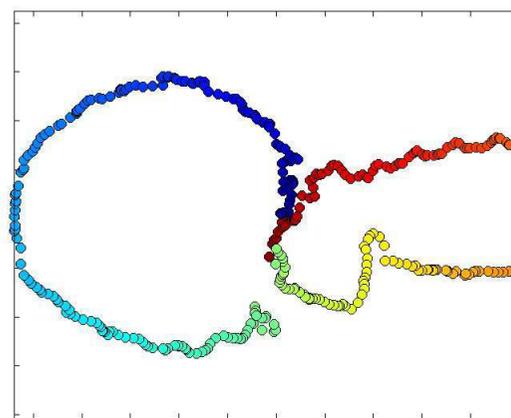
ACP



GTM



ISOMAP (k=10)



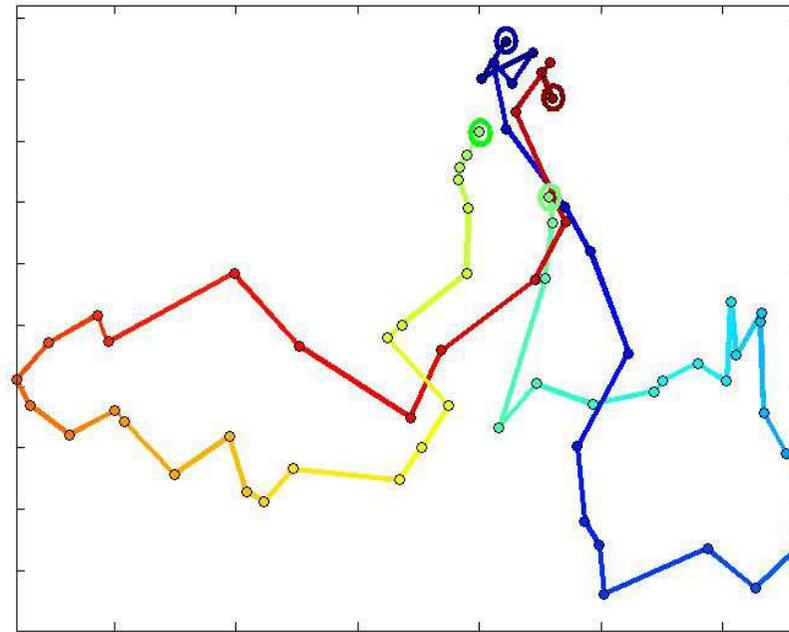
ISOMAP (k=20)





- **Caractérisation de la connexité avec le GGG**

$N^*_0 = 67$  - Projection du graphe par ACP



⇒ 2 composantes connexes



## Analyse exploratoire de données non-étiquetées

- Naviguer sur la variété

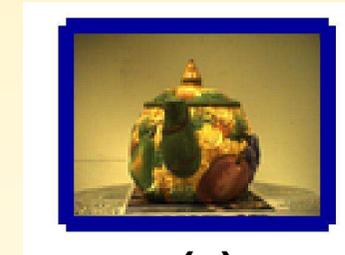
list



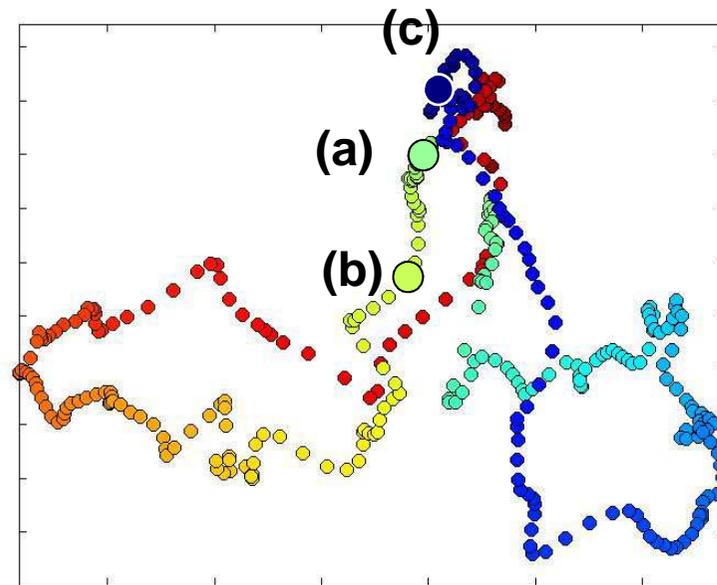
(a)



(b)



(c)





## Analyse exploratoire de données non-étiquetées

- Naviguer sur la variété

list



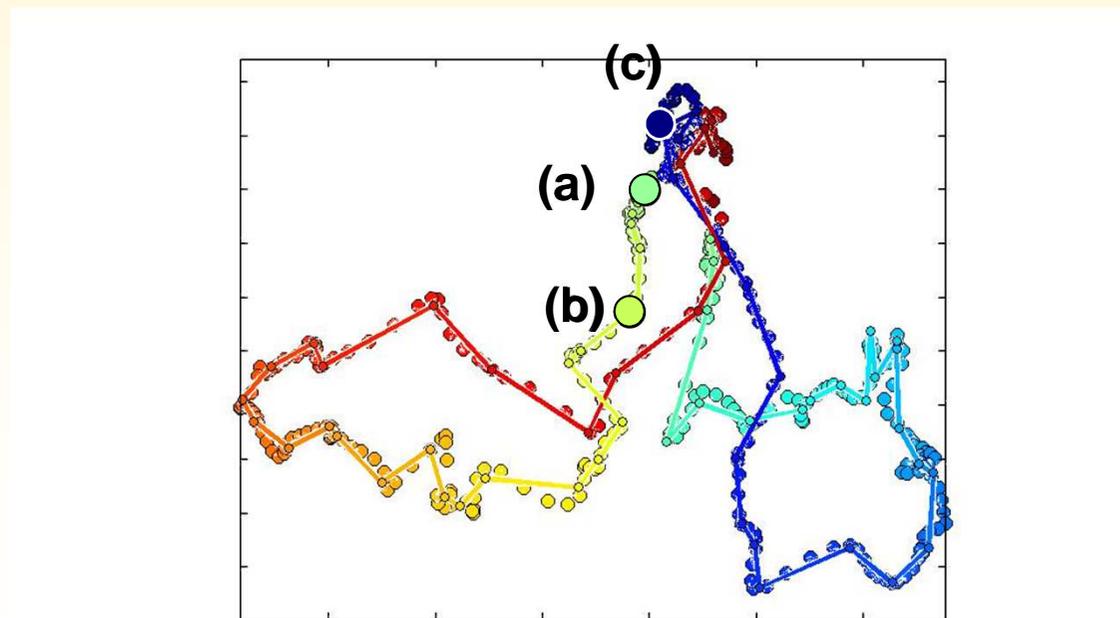
(a)



(b)



(c)





## Analyse exploratoire de données non-étiquetées

- Naviguer sur la variété

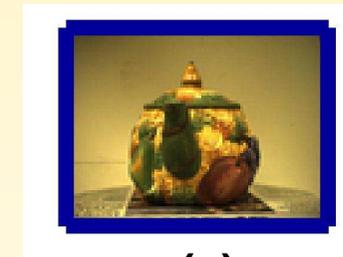
list



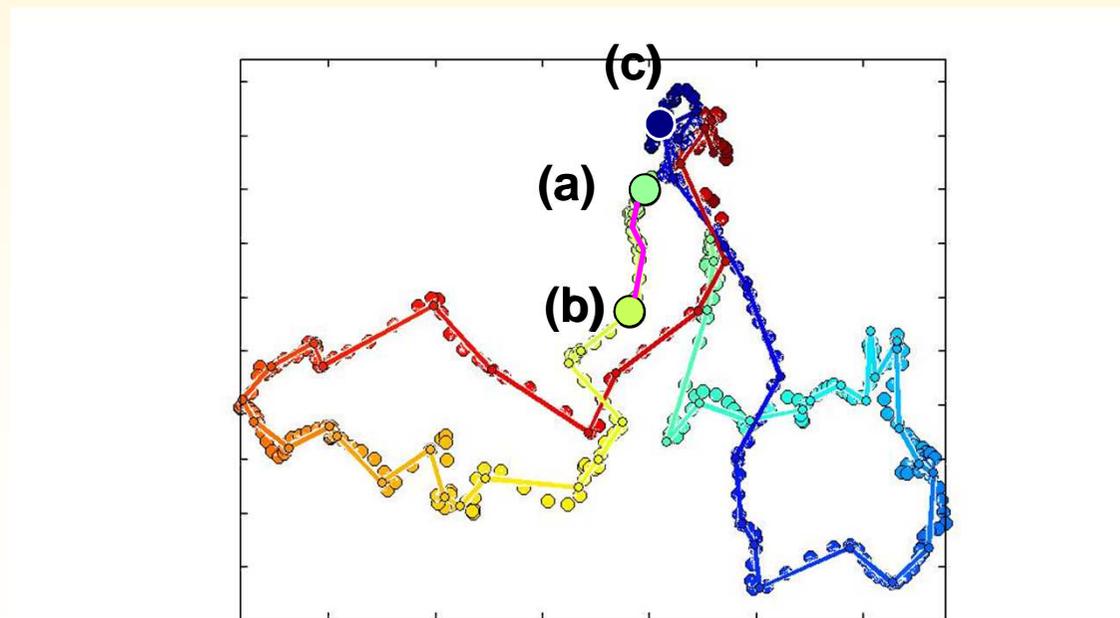
(a)



(b)



(c)





## Plan

---



- Contexte
- Notion de topologie
- Analyse *in situ*
- **Conclusion**



## Analyse *in situ* > Conclusions

---



- Un modèle génératif permettant l'extraction de la topologie
  - Construit directement dans l'espace d'origine
  - Cadre statistique
- Hypothèses
  - Graphe de Delaunay suffisamment riche pour contenir comme sous-graphe un bon modèle des variétés génératrices
  - Nécessite un modèle de bruit
- Perspectives
  - Extension aux complexes simpliciaux (points, segments, triangles, tétraèdres...)
  - Applications en classification non-supervisée, semi-supervisée et supervisée, données EEG, ECG,
  - Quel lien entre critère BIC et bonne topologie?
  - Extension aux données de type graphe? Autres données? Autres métriques?
  - Calcul sur GPU?



## Références

---



- NIPS 2005
  - M.A. – **Learning Topology with the Generative Gaussian Graph and the EM algorithm** – Advances in Neural Information Processing and Systems 18, Proceedings of the 2005 Conference, Yair Weiss, Bernhard Schölkopf and John C. Platt, eds., pp.83-90, 2006.
- Workshop « Topology Learning » NIPS 2007
  - <http://topolearnnips2007.insa-rouen.fr/>
- Neurocomputing 2008
  - Pierre Gaillard, M. A., Gérard Govaert – **Learning topology of a labeled data set with the supervised generative Gaussian graph** – Neurocomputing, 71(7-9) March 2008, 1283-1299, Elsevier
- Thèse Pierre Gaillard – UTC 2008,
  - Apprentissage de la connexité d'un nuage de points par modèle génératif** – Dir. G. Govaert, Encadrant M. A.
- Thèse Maxime Maillot en cours (2010-2012) Dir. G. Govaert, Co-Dir M. A.
  - Extension aux complexes simpliciaux
  - Application EEG, ECG... autres attracteurs

**Merci de votre attention**

