Workshop

APEX

Plans d'expériences

10 -12 Septembre 2013

Hostellerie La Magnaneraie

VILLENEUVE LÈS AVIGNON

Visualisation de données par l'Analyse en Composantes Curvilignes

L'analyse de données

- L'accroissement du nombre de dimensions dans la collecte des données crée un réel besoin de **techniques de réduction de dimensionnalité**
- Ex : acquisition d'images, astronomie, génétique…



- **Nécessité de visualiser** ces données en grande dimension pour qu'elles soient **plus facilement interprétables**

Utilisation de techniques de réduction de dimensionnalité

Quelques méthodes de réduction de dimensionnalité

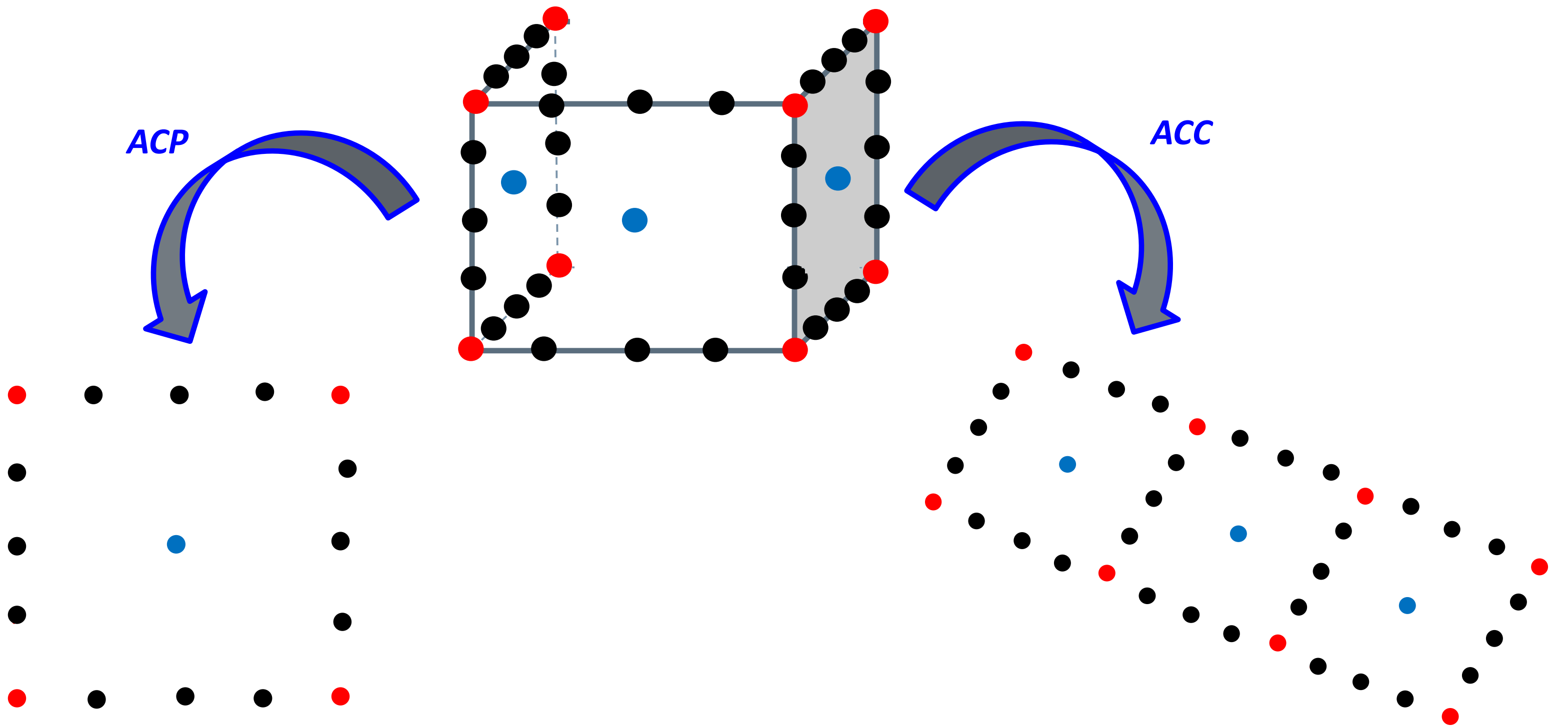
Visualisation de données en grande dimension

- **ACP** : Analyse en Composantes Principales
- **MDS** : MultiDimensional Scaling
- **Isomap**
- **Cartes de Sammon**
- **SOM** : cartes auto-organisatrices de Kohonen
- **ACC** : Analyse en Composantes Curvilignes
- **ADC** : Analyse en Distances Curvilignes

Comparaison ACP / ACC

- **ACP** = méthode la plus ancienne et la plus souvent utilisée
  - Limitation **ACP** : méthode exclusivement **linéaire**

- **ACC :**
  - Méthode **non linéaire**, proposée par Demartines en 1992*
  - Efficace dans le cadre du **dépliement** de données

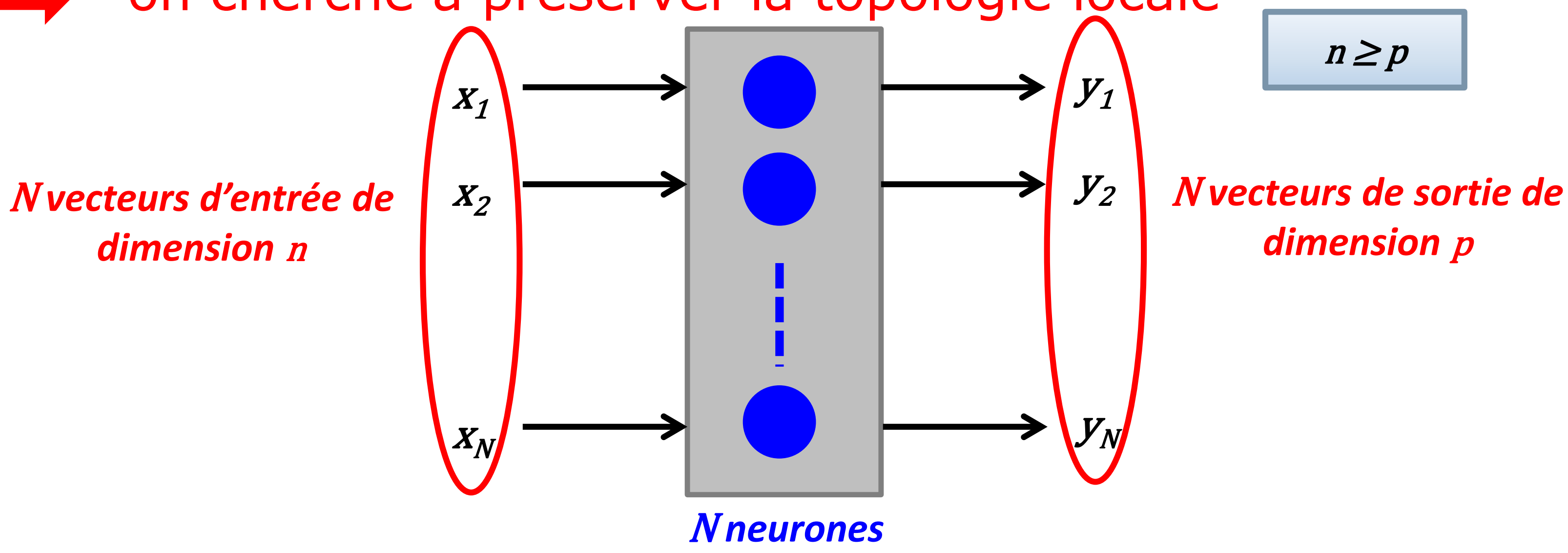


* P. Demartines, Analyse de données par réseaux de neurones auto-organisés, PhD Thesis, Institut National Polytechnique de Grenoble,1992

ACC : Analyse en Composantes Curvilignes

- Reproduire la topologie de l'espace d'entrée $\mathbf{E} = \mathbf{R}^n$ dans l'espace de sortie $\mathbf{F} = \mathbf{R}^p$ dans lequel on souhaite projeter les points ($n \geq p$)

➡ on cherche à préserver la topologie locale



- Principe général : minimiser un critère caractérisant les différences de topologie entre l'espace initial et l'espace de projection

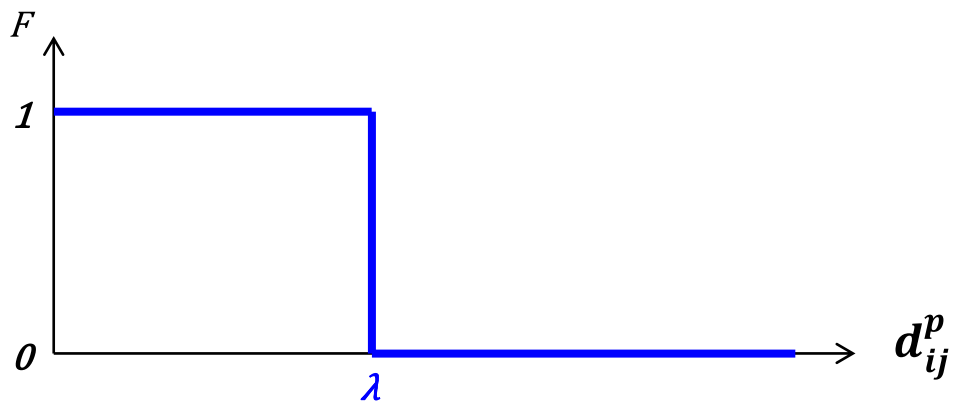
$$E_{ACC} = \frac{1}{2} \sum_{i} \sum_{j \neq i} \left(d_{ij}^{n} - d_{ij}^{p} \right)^{2} F_{\lambda}(d_{ij}^{p})$$

ACC : analyse en composantes curvilignes

- Critère à minimiser :

$$E_{ACC} = \frac{1}{2}\sum_{i}\sum_{j\neq i}\left(d_{ij}^{n} - d_{ij}^{p}\right)^{2} F_{\lambda}(d_{ij}^{p})$$

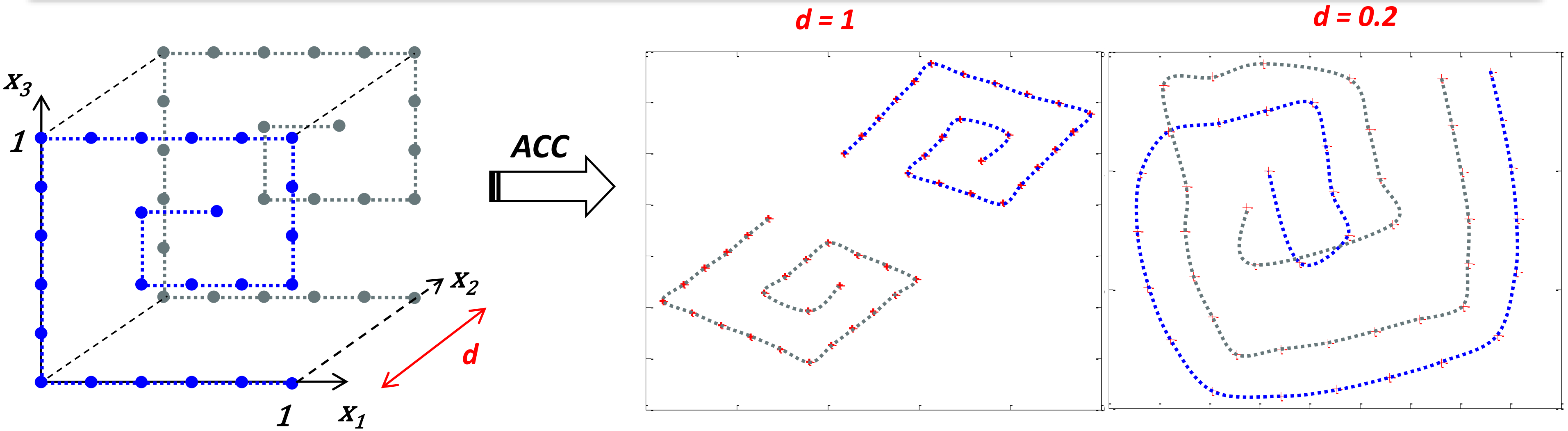
- $F_{\lambda}(d_{ij}^{p})$ est une fonction positive monotone et décroissante, qui favorise le respect local de la topologie

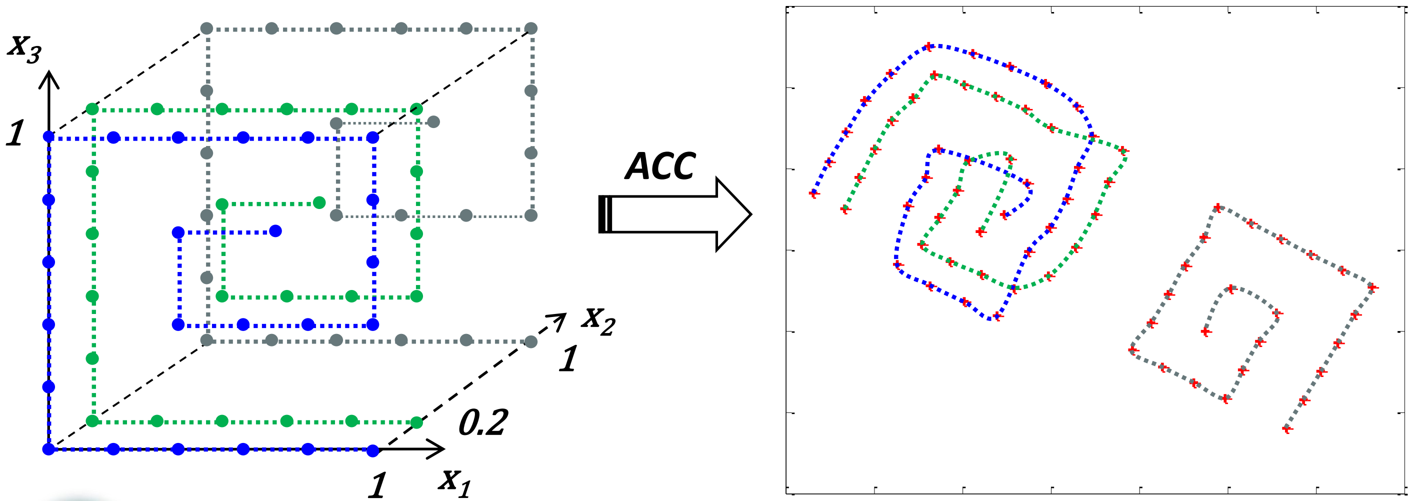


Exemples d'application de l'ACC :

- Cas ***simulé***
  - étude de spirales en trois dimensions
- Cas ***réel***
  - étude de fromages par IR

Application de l'ACC sur un cas *simulé* en faible dimension

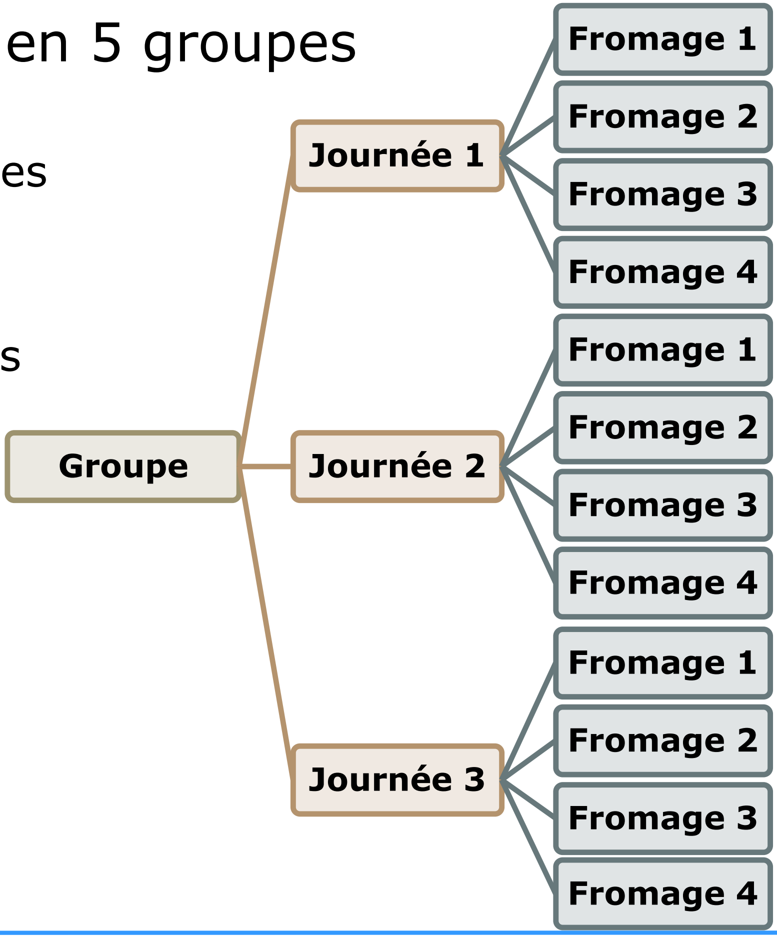




Application de l’ACC sur un cas *réel*

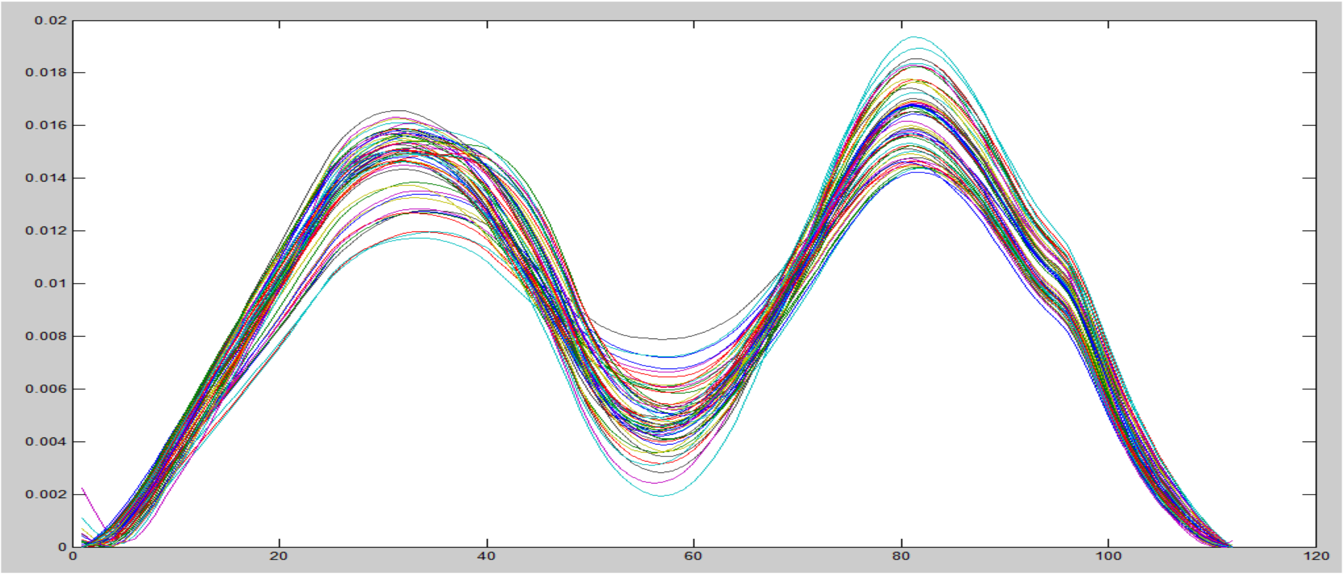
- Présentation des données :
  - Spectres IR de fromages mesurant 112 absorbances aux nombres d’onde correspondants de 1700.9 cm^{-1} à 1486.8 cm^{-1}

- Etude de 60 fromages répartis en 5 groupes
  - Groupe 1 = Pâtes pressées cuites
  - Groupe 2 = Pâtes pressées mi-cuites
  - Groupe 3 = Pâtes pressées
  - Groupe 4 = Pâtes molles
  - Groupe 5 = Pâtes molles stabilisées

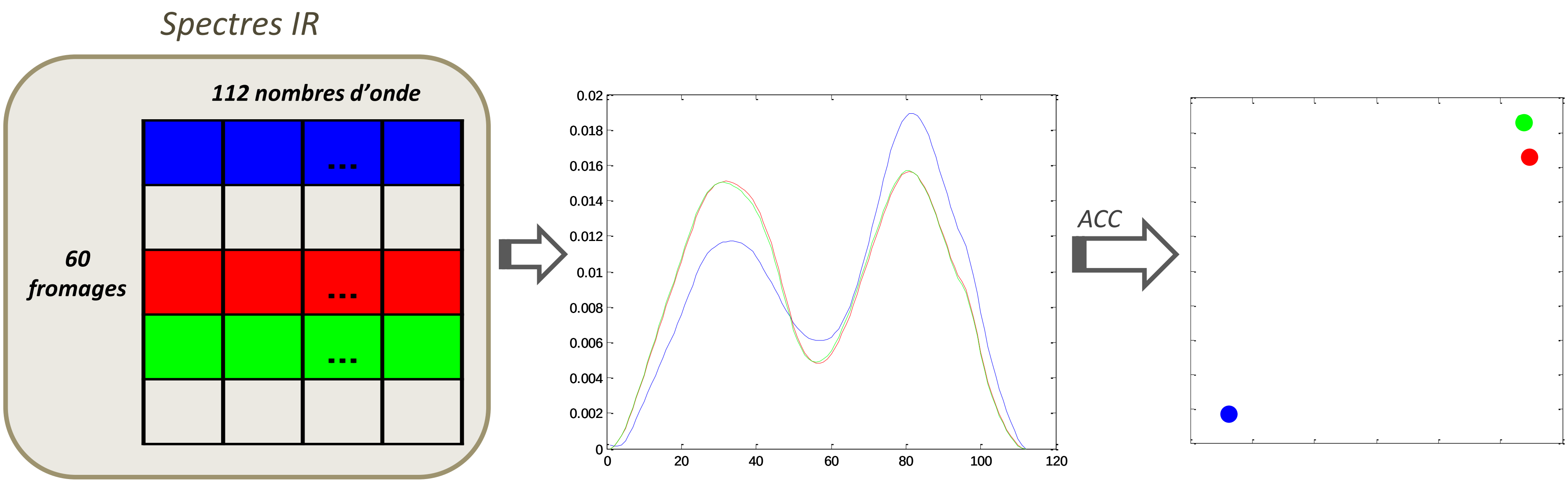


Application de l'ACC sur un cas réel

- Spectres des 60 fromages obtenus à partir des données initiales



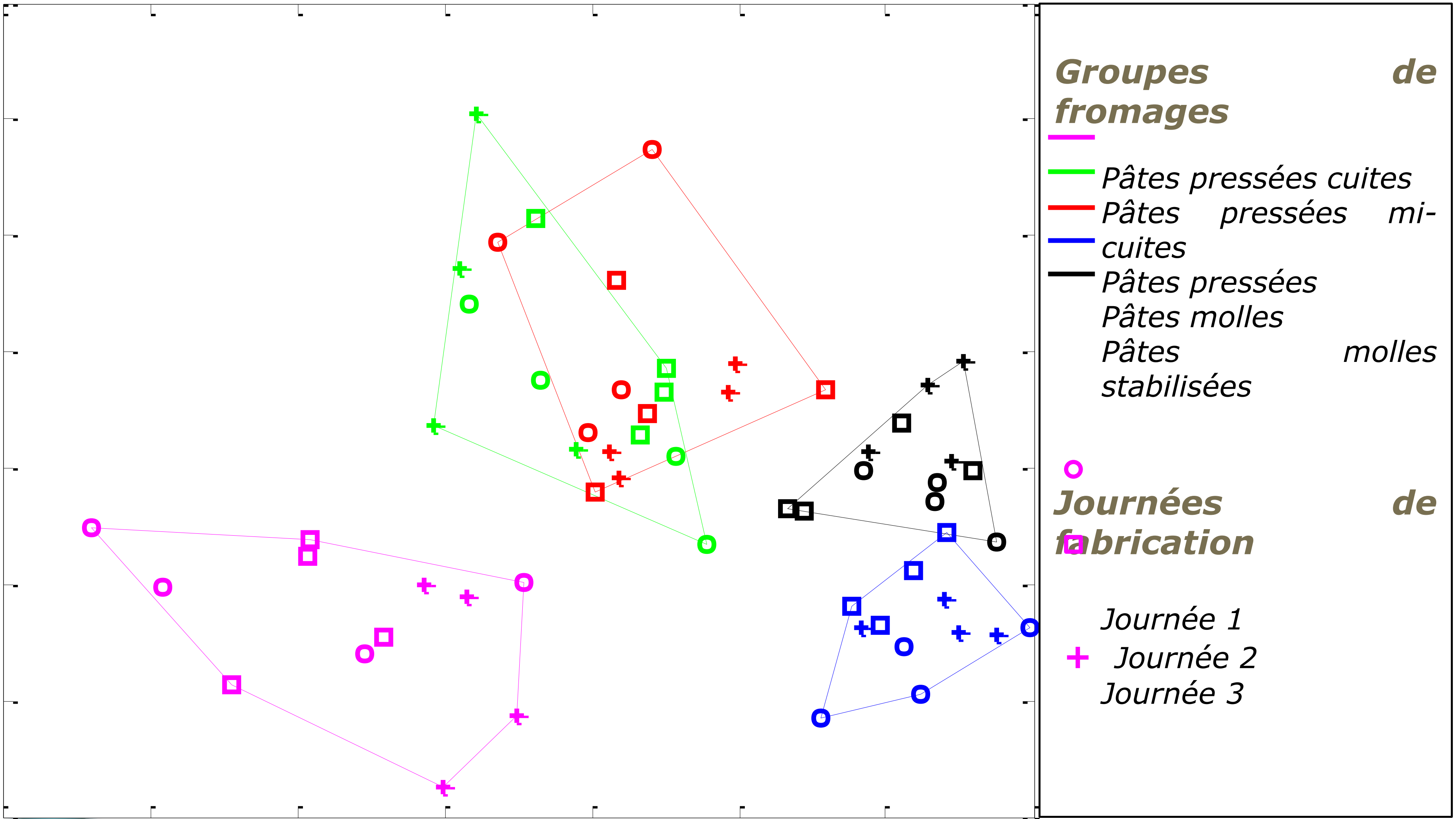
Application de l'ACC sur un cas réel



❑ *L'ACC permet de visualiser dans un espace à deux dimensions, les spectres ayant des caractéristiques proches dans l'espace initial*

Application de l'ACC sur un cas *réel*

- ACC sur les 60 spectres



Conclusion

- ACC est une méthode de réduction de dimensionnalité par projection non linéaire
- ACC permet de conserver les points proches lors de la réduction de dimensionnalité d’un espace en grande dimension vers un espace en plus faible dimension, sans grande perte d’information locale