Segmentation non supervisée d'image hyperspectrale par mélange de gausienne spatialisé

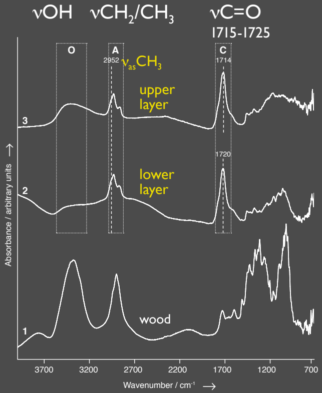
E. Le Pennec
(SELECT - INRIA Saclay / Université Paris Sud)
et
S. Cohen (IPANEMA - CNRS / Soleil)

GRETSI 11
07 septembre 2011

A. Stradivari (1644 - 1737)

Provigny (1716)

A. Giordan © Cité de la Musique

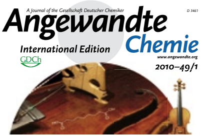


SOLEIL SYNCHROTRON

4 / 8 cm^{-1} resolution
64 / 128 scans
typ. 1 min/sp, 400sp

very simple process
no protein (amide I, amide II)
no gums, nor waxes
@SOLEIL: SMIS

cité de la musique · SOLEIL SYNCHROTRON · CRCC · CENTRE DE RECHERCHE ET DE RESTAURATION DES MUSÉES DE FRANCE · Laboratoire de Dynamique, Interactions et Réactivité



J.-P. Echard, L. Bertrand, A. von Bohlen, A.-S. Le Hô, C. Paris, L. Bellot-Gurlet, B. Soulier, A. Lattuati-Derieux, S. Thao, L. Robinet, B. Lavédrine, and S. Vaiedelich. *Angew. Chem. Int. Ed.*, 49(1), 197-201, 2010.

Segmentation d'images hyperspectrales

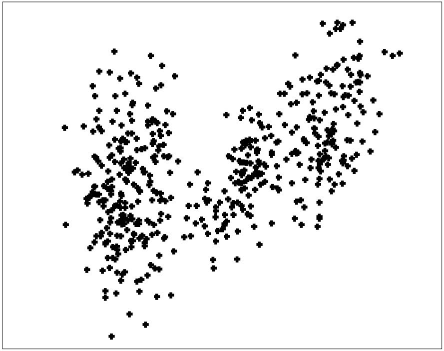
- Données :
  - image de taille N comprise entre ~ 1000 et ~ 100000 pixels,
  - spectres $\mathcal{S}$ de ~ 1024 points,
  - très bonne résolution spatiale,
  - possibilité de mesurer de très nombreux spectres par minute...
- Objectifs immédiats :
  - segmentation automatique de ces images,
  - sans intervention humaine,
  - aide à l'analyse des résultats.
- Objectifs lointains :
  - classification automatique,
  - interprétation...

Segmentation d'images hyperspectrales

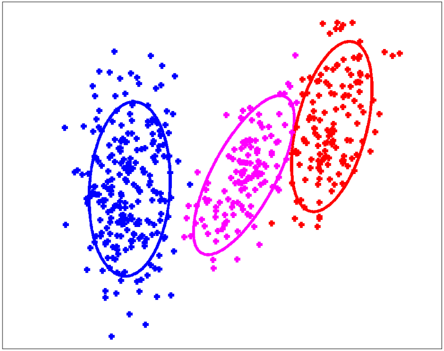
- Données :
  - image de taille N comprise entre ~ 1000 et ~ 100000 pixels,
  - spectres $\mathcal{S}$ de ~ 1024 points,
  - très bonne résolution spatiale,
  - possibilité de mesurer de très nombreux spectres par minute...
- Objectifs immédiats :
  - segmentation automatique de ces images,
  - sans intervention humaine,
  - aide à l'analyse des résultats.
- Objectifs lointains :
  - classification automatique,
  - interprétation...

Un problème “jouet”

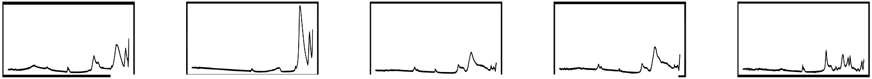
Un problème “jouet”

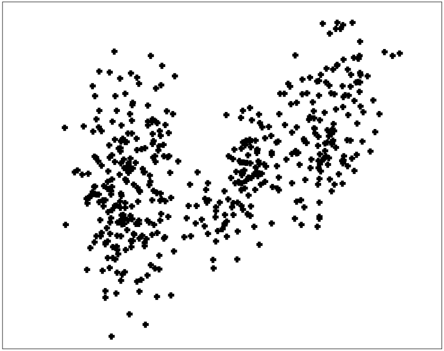


Un problème “jouet”

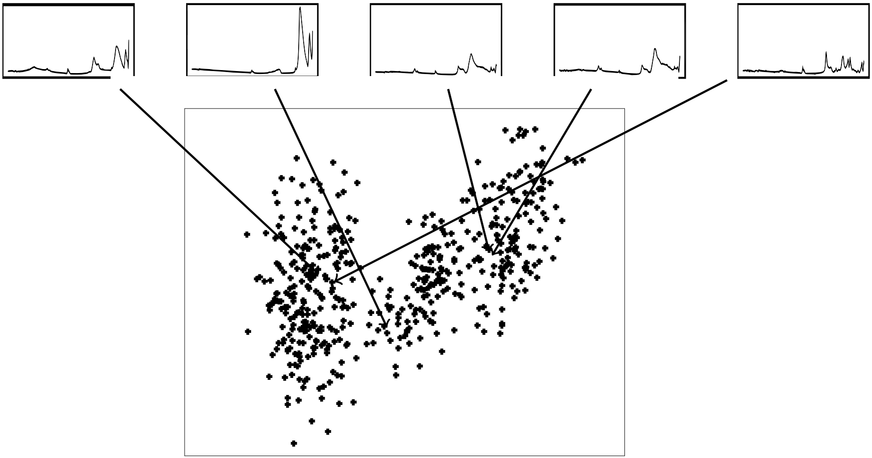


Un problème “jouet”

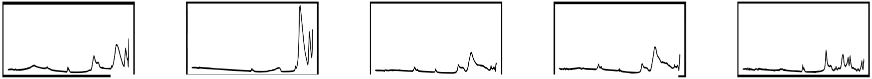


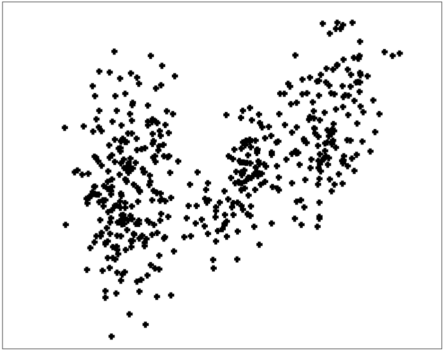


Un problème “jouet”

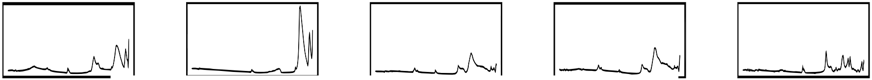


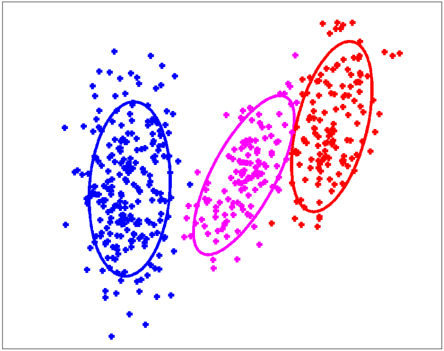
Un problème “jouet”



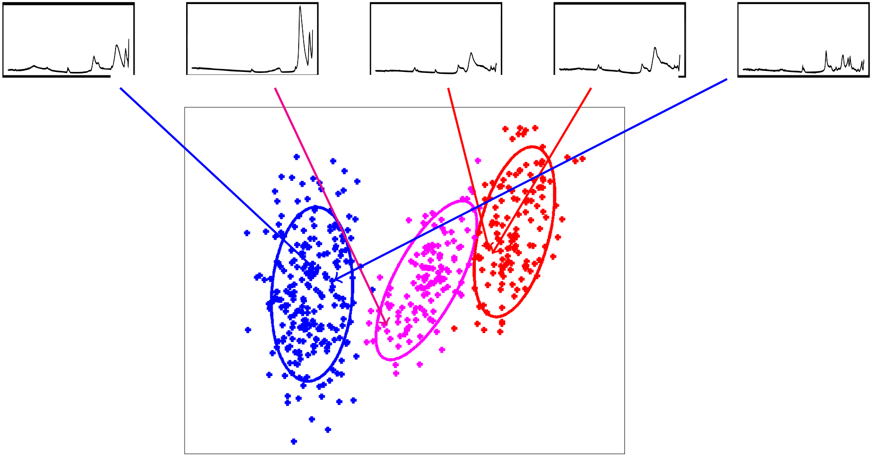


Un problème "jouet"





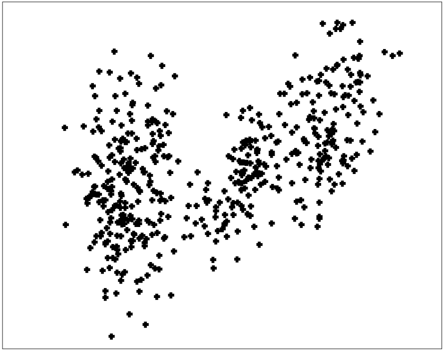
Un problème “jouet”



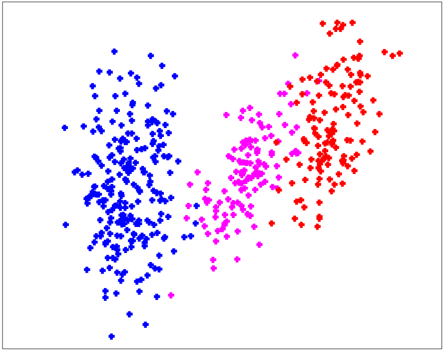
- Représentation : correspondance entre les spectres et des points dans un espace de grande dimension.
- Méthode spectrale.

Modélisation “stochastique”

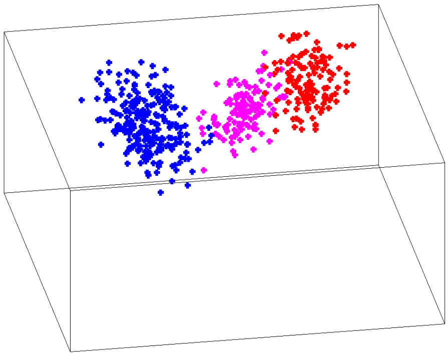
Modélisation “stochastique”



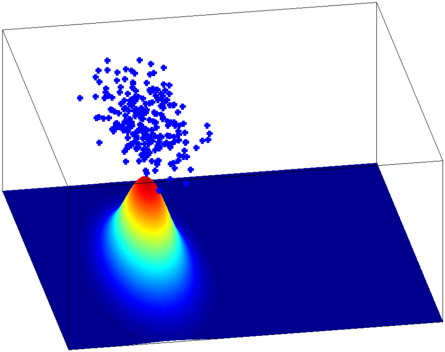
Modélisation “stochastique”



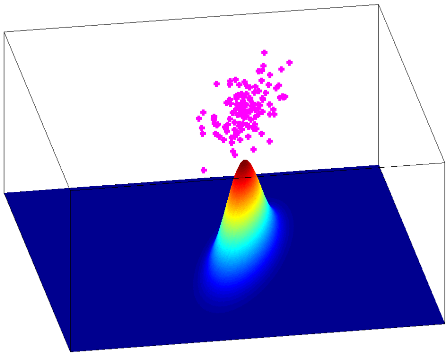
Modélisation “stochastique”



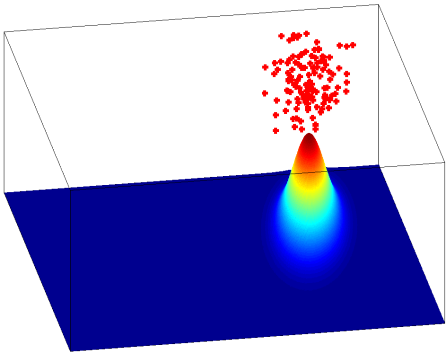
Modélisation “stochastique”



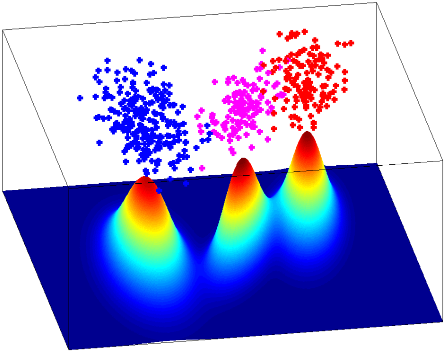
Modélisation “stochastique”



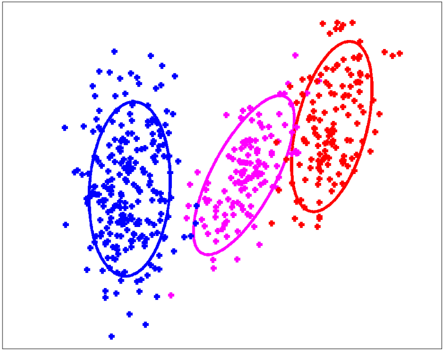
Modélisation “stochastique”



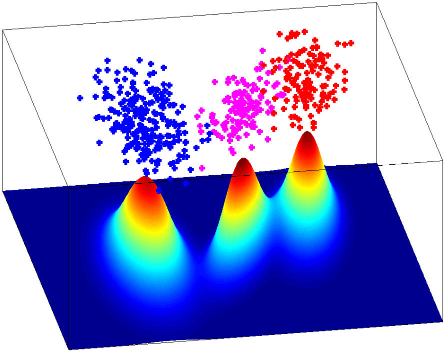
Modélisation “stochastique”



Modélisation “stochastique”



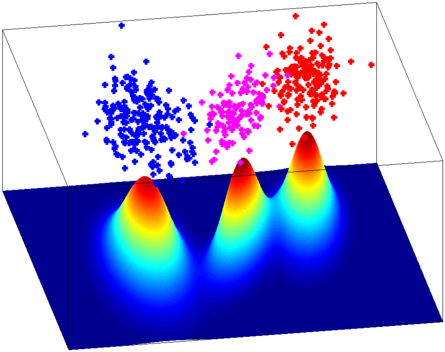
Modélisation “stochastique”



- Modèle : mélange de gaussiennes à K classes.
- Densité du mélange :

$$
\begin{aligned}
s_{K,\pi,\mu,\Sigma}(\mathcal{S}) &= \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} e^{-\frac{1}{2}(\mathcal{S}-\mu_k)^t \Sigma_k^{-1} (\mathcal{S}-\mu_k)} \\
&= \sum_{k=1}^{K} \pi_k \mathcal{N}_{\mu_k,\Sigma_k}(\mathcal{S})
\end{aligned}
$$

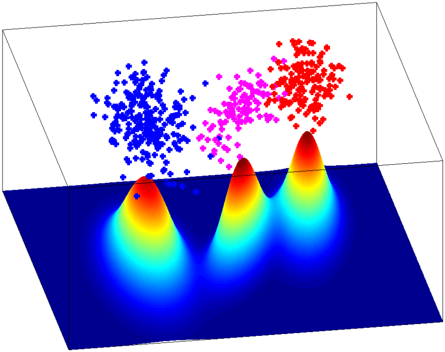
Modélisation "stochastique"



- Modèle : mélange de gaussiennes à K classes.
- Densité du mélange :

$$
\begin{aligned}
s_{K,\pi,\mu,\Sigma}(\mathcal{S}) &= \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} e^{-\frac{1}{2}(\mathcal{S}-\mu_k)^t \Sigma_k^{-1} (\mathcal{S}-\mu_k)} \\
&= \sum_{k=1}^{K} \pi_k \mathcal{N}_{\mu_k,\Sigma_k}(\mathcal{S})
\end{aligned}
$$

Modélisation "stochastique"

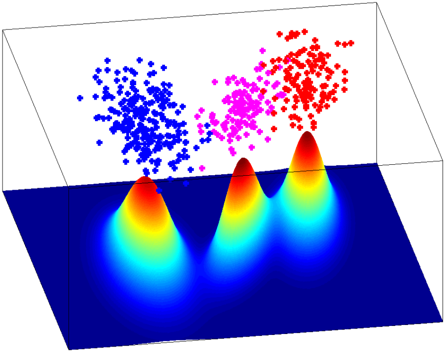


- Modèle : mélange de gaussiennes à K classes.
- Densité du mélange :

$$s_{K,\pi,\mu,\Sigma}(\mathcal{S}) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} e^{-\frac{1}{2}(\mathcal{S}-\mu_k)^t \Sigma_k^{-1}(\mathcal{S}-\mu_k)}$$

$$= \sum_{k=1}^{K} \pi_k \mathcal{N}_{\mu_k,\Sigma_k}(\mathcal{S})$$

Modélisation “stochastique”

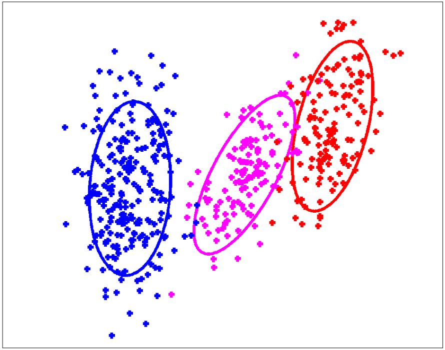


- Modèle : mélange de gaussiennes à K classes.
- Densité du mélange :

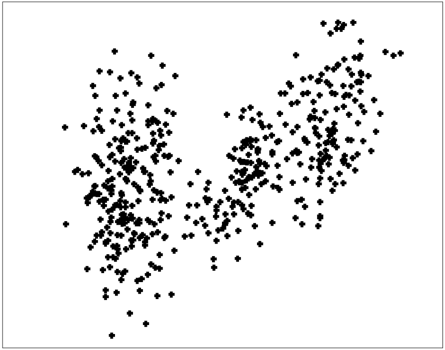
$$\begin{aligned} s_{K,\pi,\mu,\Sigma}(\mathcal{S}) &= \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} e^{-\frac{1}{2}(\mathcal{S}-\mu_k)^t \Sigma_k^{-1} (\mathcal{S}-\mu_k)} \\ &= \sum_{k=1}^{K} \pi_k \mathcal{N}_{\mu_k,\Sigma_k}(\mathcal{S}) \end{aligned}$$

Estimation “statistique”

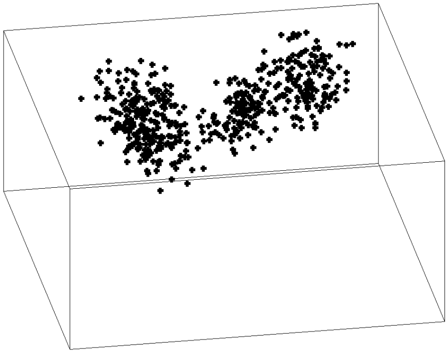
Estimation “statistique”



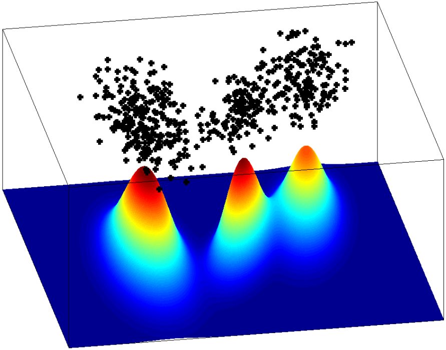
Estimation "statistique"



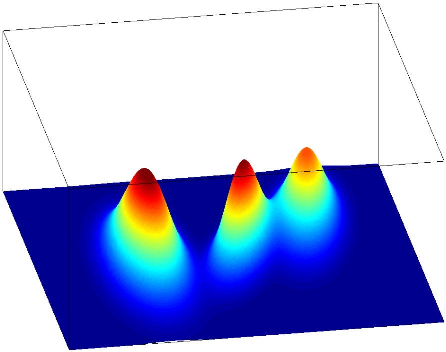
Estimation “statistique”



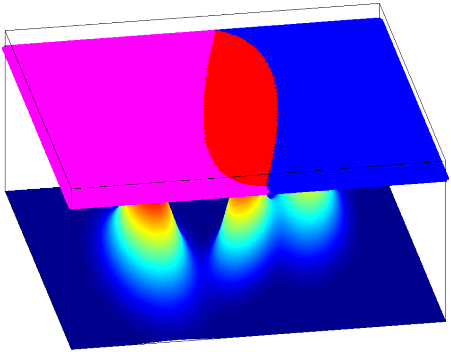
Estimation "statistique"



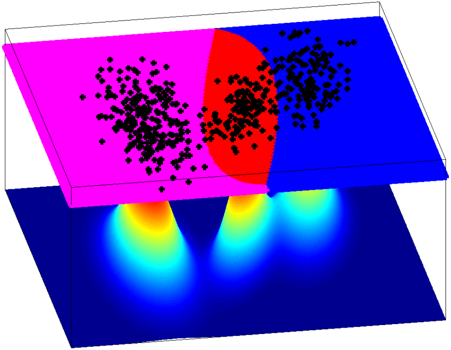
Estimation “statistique”



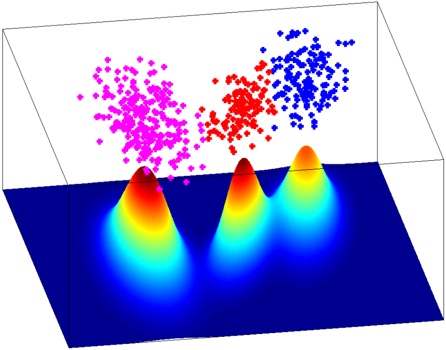
Estimation "statistique"



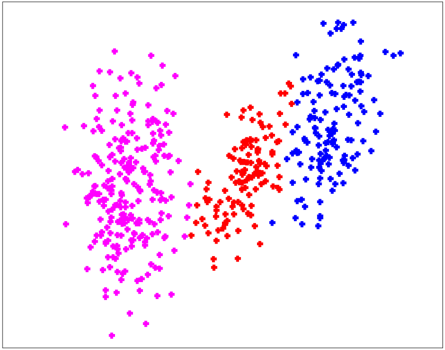
Estimation “statistique”



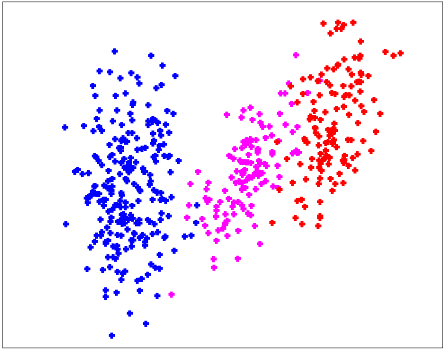
Estimation “statistique”



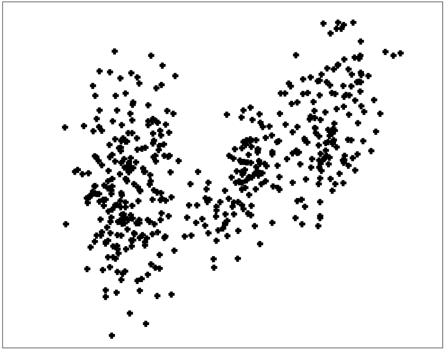
Estimation “statistique”



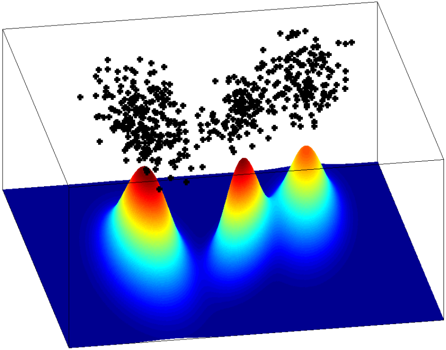
Estimation “statistique”



Estimation "statistique"



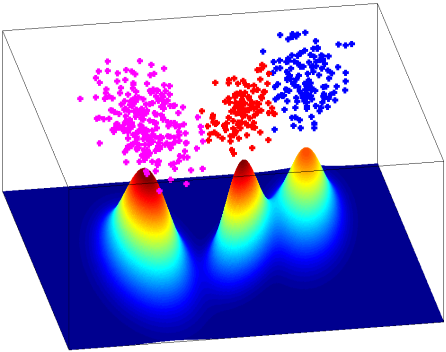
Estimation “statistique”



- Estimation des π_k, $\widehat{\mu_k}$ et $\widehat{\Sigma_k}$ par maximum de vraisemblance :

$$(\widehat{\pi_k}, \widehat{\mu_k}, \widehat{\Sigma_k}) = \operatorname{argmax} \sum_{i=1}^{N} \log s_{K,(\pi_k,\mu_k,\Sigma_k)}(\mathcal{S}_i)$$

Estimation "statistique"



- Estimation des π_k, $\widehat{\mu_k}$ et $\widehat{\Sigma_k}$ par maximum de vraisemblance :

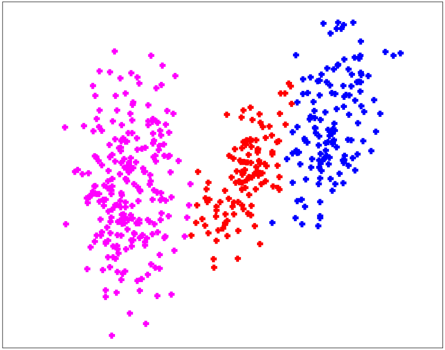
$$(\widehat{\pi_k}, \widehat{\mu_k}, \widehat{\Sigma_k}) = \operatorname{argmax} \sum_{i=1}^{N} \log s_{K,(\pi_k,\mu_k,\Sigma_k)}(\mathcal{S}_i)$$

- Estimation de $\widehat{k}(\mathcal{S})$ par maximum à posteriori :

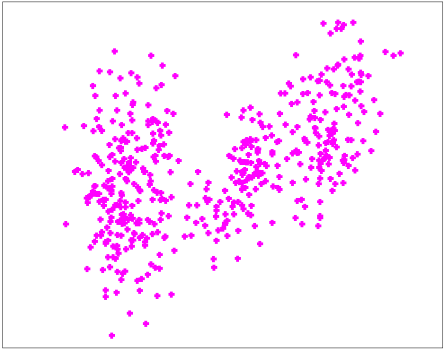
$$\widehat{k}(\mathcal{S}) = \operatorname{argmax} \widehat{\pi_k} \, \mathcal{N}_{\mu_k,\Sigma_k}(\mathcal{S})$$

Combien de classes ?

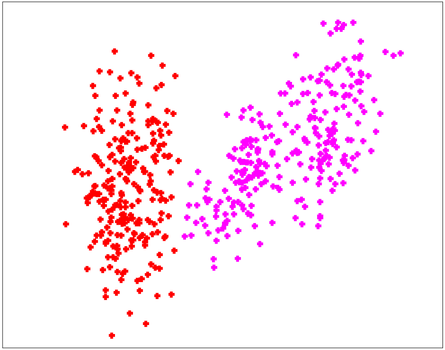
Combien de classes ?



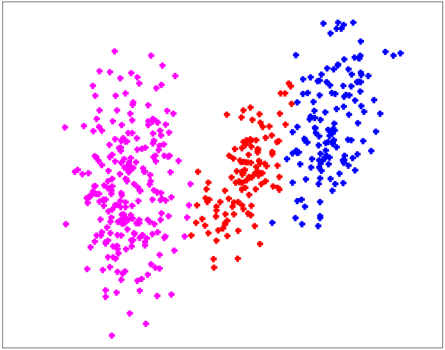
Combien de classes ?



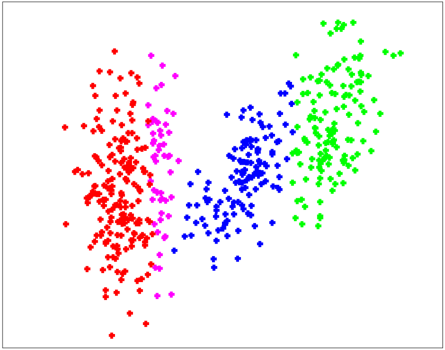
Combien de classes ?



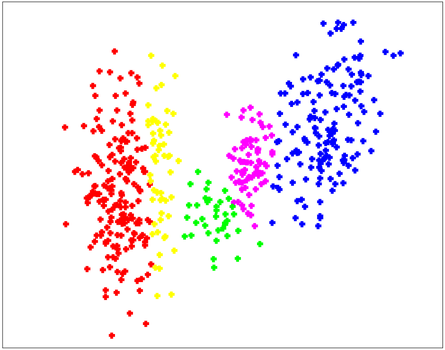
Combien de classes ?



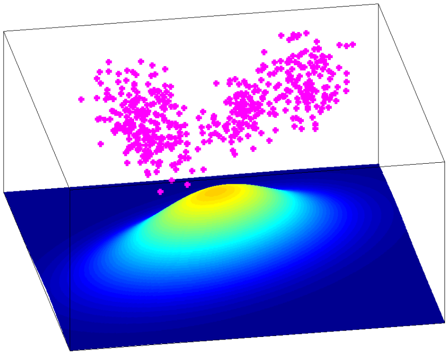
Combien de classes ?



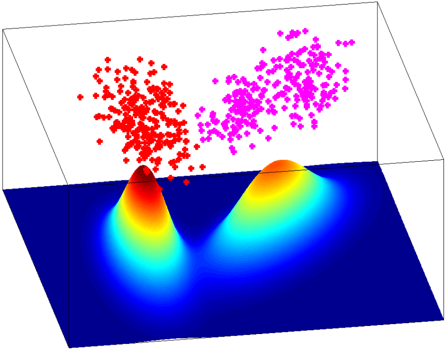
Combien de classes ?



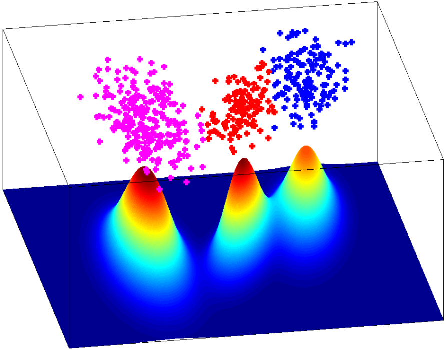
Combien de classes ?



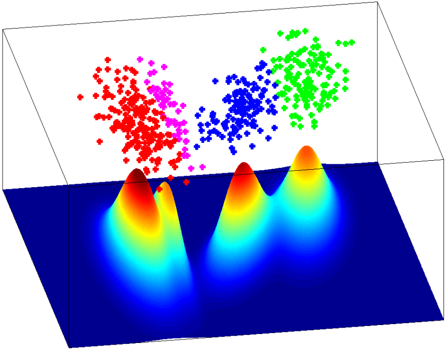
Combien de classes ?



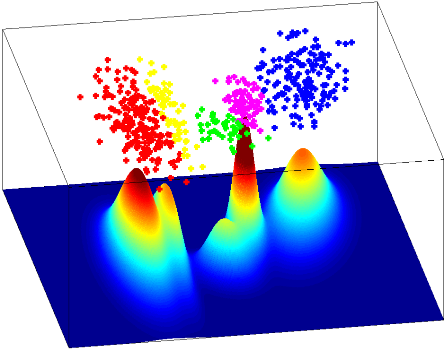
Combien de classes ?



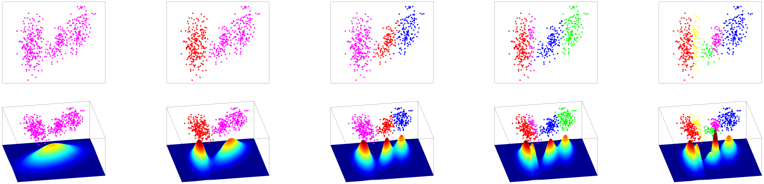
Combien de classes ?



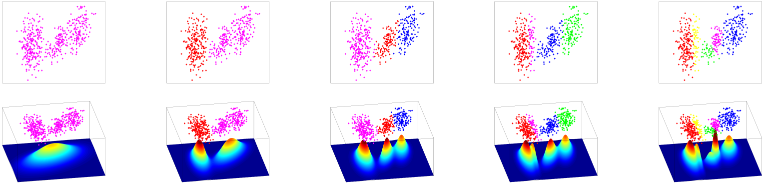
Combien de classes ?



Combien de classes ?

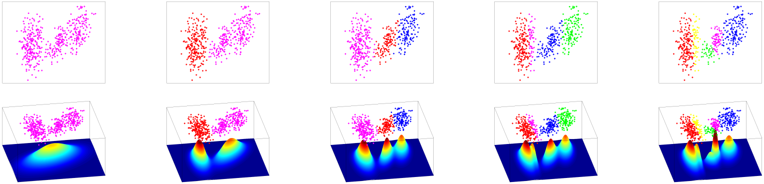


Combien de classes ?



Fidélité	$--$	$+$	$+++$	$+++\cdot$	$++++$

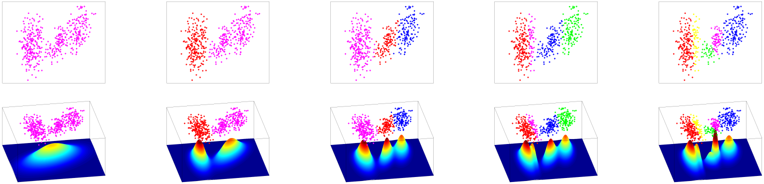
Combien de classes ?



Fidélité	$--$	$+$	$+++$	$+++\cdot$	$++++$

- Question difficile où la vraisemblance (la fidélité) ne suffit pas !

Combien de classes ?



Fidélité	$--$	$+$	$+++$	$+++\cdot$	$++++$
Simplicité	$++++$	$+++$	$++$	$+$	$-$

- Question difficile où la vraisemblance (la fidélité) ne suffit pas !
- Prise en compte de la complexité du modèle ?

Le rasoir d'Ockham

Le rasoir d'Ockham

Les multiples ne doivent pas être utilisés sans nécessité.
Guillaume d'Ockham ($\sim$ 1285 - 1347)

Le rasoir d'Ockham

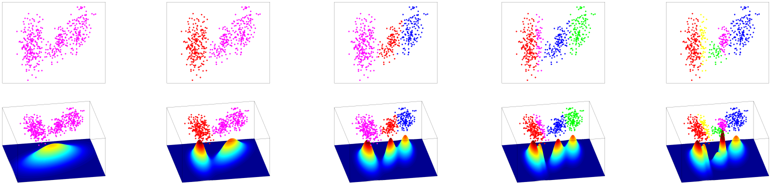
Les multiples ne doivent pas être utilisés sans nécessité.

Guillaume d'Ockham ($\sim$ 1285 - 1347)

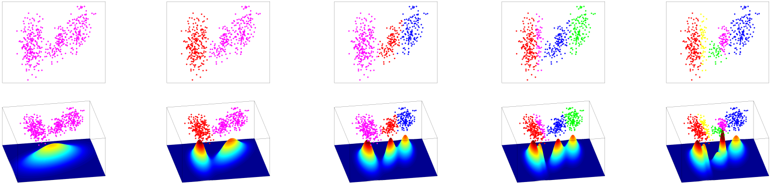
- Rasoir d'Ockham (principe de simplicité) : il ne faut pas ajouter des hypothèses, si celles utilisées suffisent déjà !
- Compromis entre pouvoir d'explication et simplicité.

Sélection par pénalisation

Sélection par pénalisation

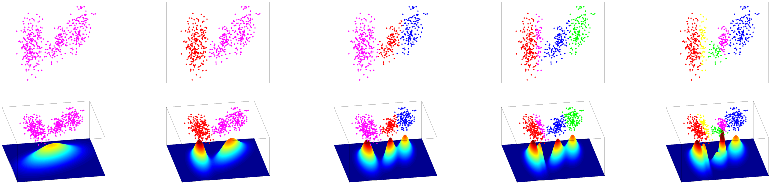


Sélection par pénalisation



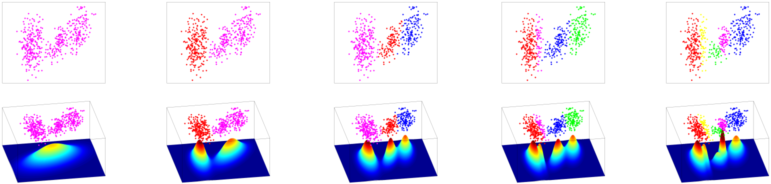
Vraisemblance	$--$	$+$	$+++$	$+++\cdot$	$++++$
Simplicité	$++++$	$+++$	$++$	$+$	$-$

Sélection par pénalisation



Vraisemblance	$--$	$+$	$+++$	$+++\cdot$	$++++$
$+$ Simplicité	$++++$	$+++$	$++$	$+$	$-$
$=$ Compromis	$++$	$++++$	$+++++$	$++++\cdot$	$+++$

Sélection par pénalisation

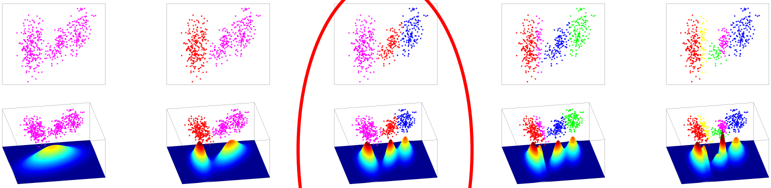


Vraisemblance	$--$	$+$	$+++$	$+++\cdot$	$++++$
$+$ Simplicité	$++++$	$+++$	$++$	$+$	$-$
$=$ Compromis	$++$	$++++$	$+++++$	$++++\cdot$	$+++$

- Vraisemblance : $\sum_{i=1}^{N} \log \hat{s}_K(X_i)$.
- Simplicité : $-\lambda \mathrm{Dim}(S_K)$ (beaucoup de théorie derrière).
- Estimateur pénalisé :

$$\mathrm{argmin} \underbrace{-\sum_{i=1}^{N} \log \hat{s}_K(X_i)}_{\text{Vraisemblance}} + \underbrace{\lambda \mathrm{Dim}(S_K)}_{\text{Pénalité}}$$

Sélection par pénalisation

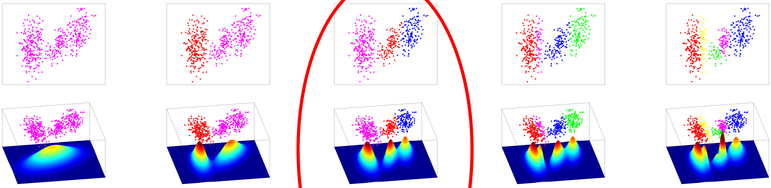


Vraisemblance	−−	+	+++	+++·	++++
+ Simplicité	++++	+++	++	+	−
= Compromis	++	++++	+++++	++++·	+++

- Vraisemblance : $\sum_{i=1}^{N} \log \hat{s}_K(X_i)$.
- Simplicité : $-\lambda \mathrm{Dim}(S_K)$ (beaucoup de théorie derrière).
- Estimateur pénalisé :

$$\operatorname{argmin} \underbrace{- \sum_{i=1}^{N} \log \hat{s}_K(X_i)}_{\text{Vraisemblance}} + \underbrace{\lambda \mathrm{Dim}(S_K)}_{\text{Pénalité}}$$

Sélection par pénalisation



Vraisemblance	$--$	$+$	$+++$	$+++\cdot$	$++++$
$+$ Simplicité	$++++$	$+++$	$++$	$+$	$-$
$=$ Compromis	$++$	$++++$	$+++++$	$++++\cdot$	$+++$

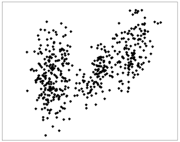
- Vraisemblance : $\sum_{i=1}^{N} \log \hat{s}_K(X_i)$.
- Simplicité : $-\lambda \mathrm{Dim}(S_K)$ (beaucoup de théorie derrière).
- Estimateur pénalisé :

$$\operatorname{argmin} \underbrace{-\sum_{i=1}^{N} \log \hat{s}_K(X_i)}_{\text{Vraisemblance}} + \underbrace{\lambda \mathrm{Dim}(S_K)}_{\text{Pénalité}}$$

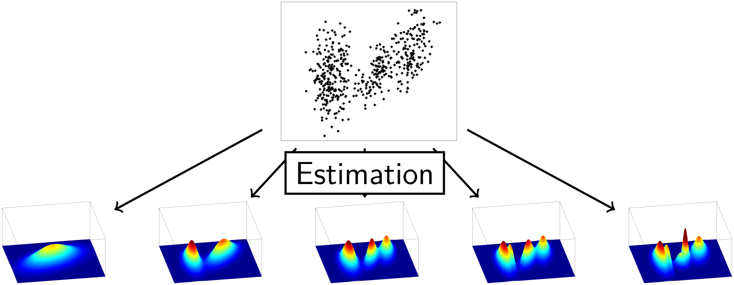
- Optimisation en K par exploration exhaustive !

Méthodologie

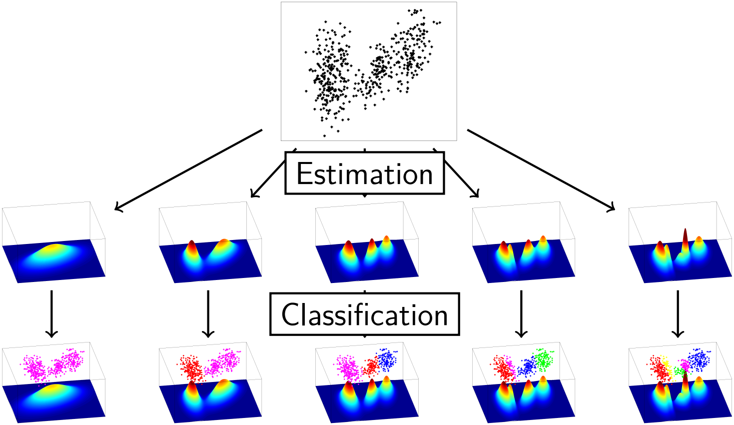
Méthodologie



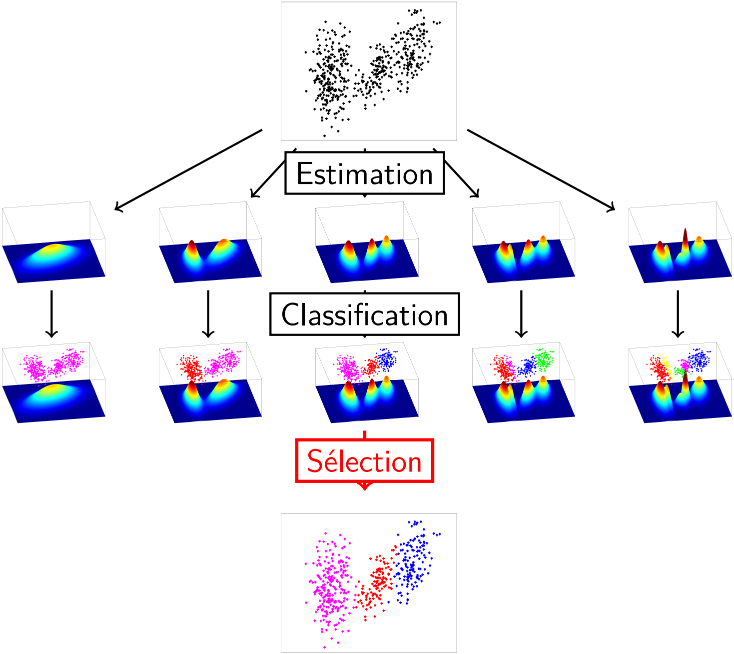
Méthodologie



Méthodologie

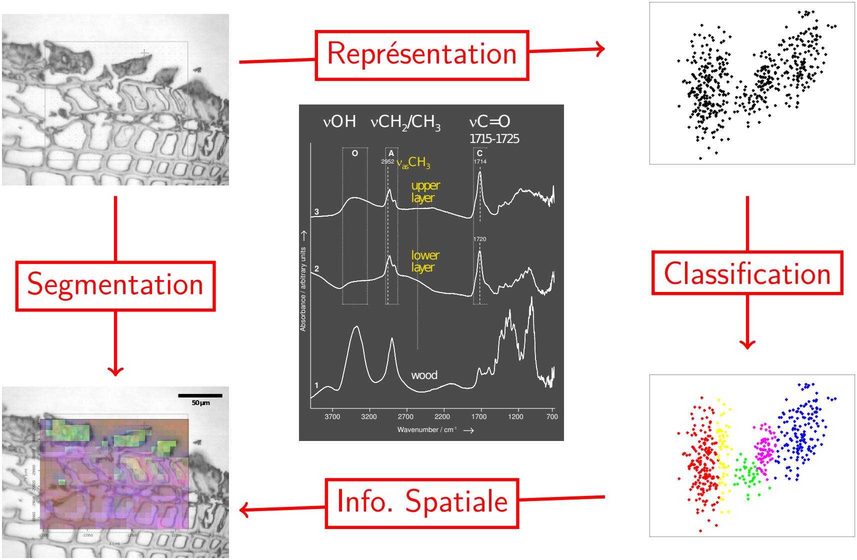


Méthodologie

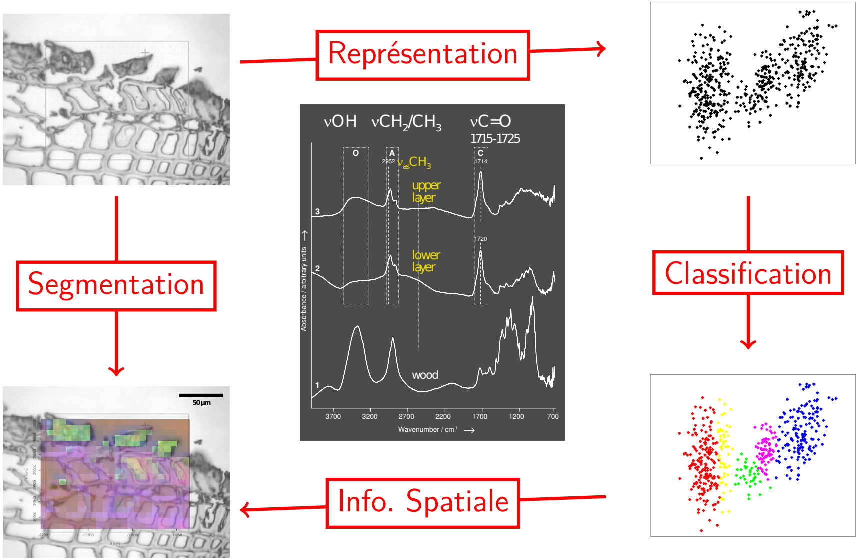


Retour à nos violons

Retour à nos violons



Retour à nos violons



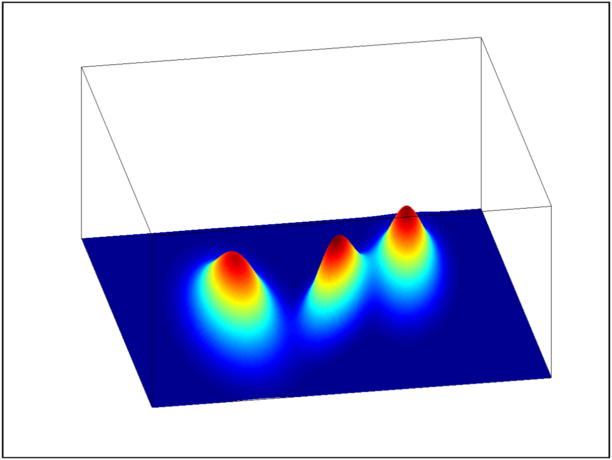
- Pb : classification $\neq$ segmentation !

Mélange de gaussien spatialisé

Mélange de gaussien spatialisé

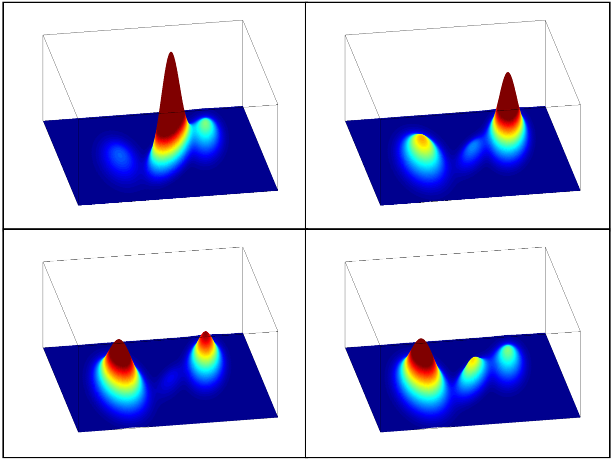
- Prise en compte du caractère spatial à travers un partitionnement récursif.

Mélange de gaussien spatialisé



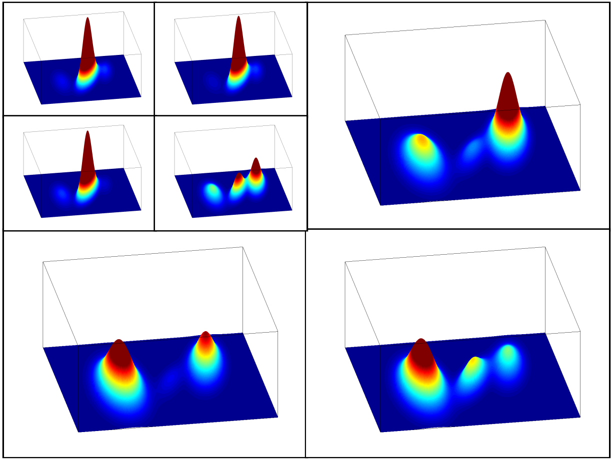
- Prise en compte du caractère spatial à travers un partitionnement récursif.

Mélange de gaussien spatialisé



- Prise en compte du caractère spatial à travers un partitionnement récursif.

Mélange de gaussien spatialisé



- Prise en compte du caractère spatial à travers un partitionnement récursif.
- Sélection de la partition (et du nombre de classes) par pénalisation.

Mélange de gaussiennes et partition hiérarchique

- Prise en compte de la position spatiale x du spectre à travers les proportions du mélange (Kolaczyk et al) : modèle de densités conditionnelles

$$\sum_{k=1}^{K} \left(\sum_{\mathcal{R} \in \mathcal{P}} \pi_k[\mathcal{R}] \chi_{\{x \in \mathcal{R}\}} \right) \mathcal{N}_{\mu_k, \Sigma_k}(\mathcal{S}).$$

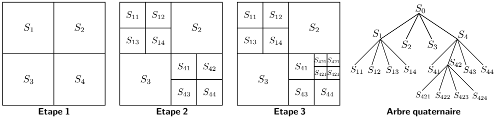
  - constant par morceau sur une partition "hiérarchique",
  - optimisation efficace possible,
  - performance d'approximation raisonnable.
- $\dim(S_{K,\mathcal{P}}) = |\mathcal{P}|(K-1) + \dim(S_K)$.
- Pénalité $\kappa \ln(n) \dim(S_{K,\mathcal{P}})$ suffisante pour
  - le contrôle théorique en terme d'estimation de densité,
  - l'optimisation numérique (EM + programmation dynamique).

Mélange de gaussiennes et partition hiérarchique

- Prise en compte de la position spatiale x du spectre à travers les proportions du mélange (Kolaczyk et al) : modèle de densités conditionnelles

$$\sum_{k=1}^{K} \left(\sum_{\mathcal{R}\in\mathcal{P}} \pi_k[\mathcal{R}] \chi_{\{x\in\mathcal{R}\}} \right) \mathcal{N}_{\mu_k,\Sigma_k}(\mathcal{S}).$$

  - constant par morceau sur une partition "hiérarchique",
  - optimisation efficace possible,
  - performance d'approximation raisonnable.



- $\dim(S_{K,\mathcal{P}}) = |\mathcal{P}|(K-1) + \dim(S_K)$.
- Pénalité $\kappa \ln(n) \dim(S_{K,\mathcal{P}})$ suffisante pour
  - le contrôle théorique en terme d'estimation de densité,
  - l'optimisation numérique (EM + programmation dynamique).

Segmentation automatique

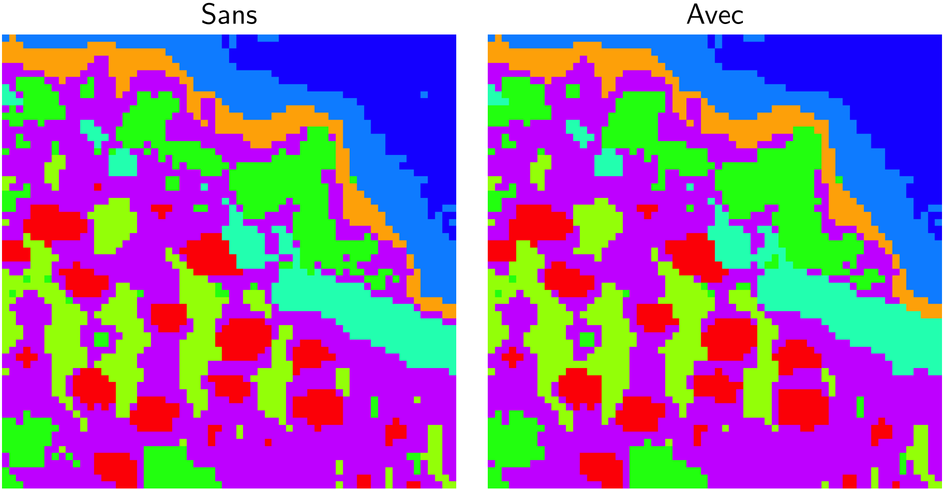
- Résultat numérique selon la prise en compte du caractère spatial :

Sans | Avec

- $K = 8$, $[L_k\, D\, B]^K$ et partition optimale.
- Calibration de la pénalité par heuristique de pente.
- Réduction de dimension par (simple) ACP...

Segmentation automatique

- Résultat numérique selon la prise en compte du caractère spatial :

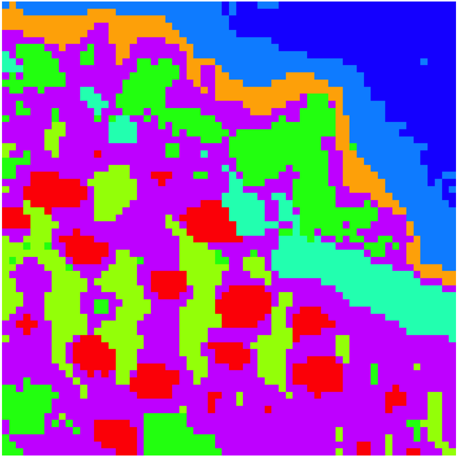


- $K = 8$, $[L_k\, D\, B]^K$ et partition optimale.
- Calibration de la pénalité par heuristique de pente.
- Réduction de dimension par (simple) ACP...

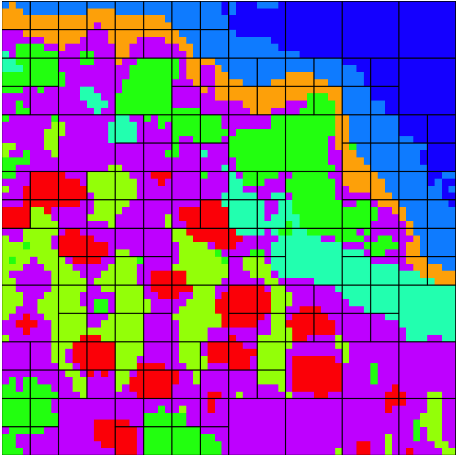
Segmentation automatique

- Résultat numérique selon la prise en compte du caractère spatial :

Sans

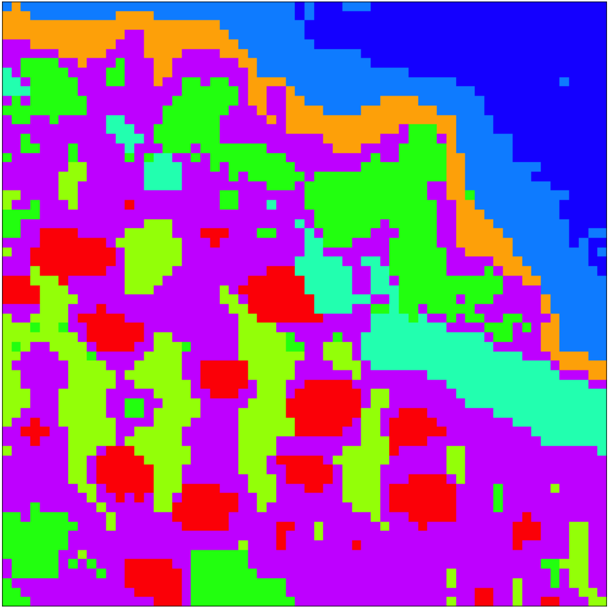


Avec



- $K = 8$, $[L_k\, D\, B]^K$ et partition optimale.
- Calibration de la pénalité par heuristique de pente.
- Réduction de dimension par (simple) ACP...

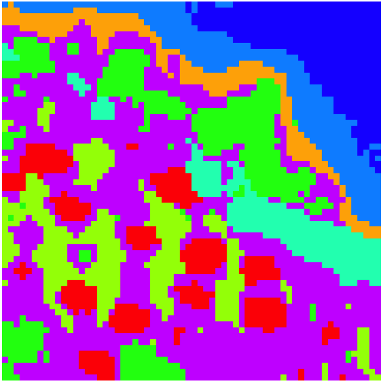
Segmentation et classification



Le secret de Stradivarius

- Deux couches fines de vernis :
  - une première couche d'huile simple, similaire à celle des peintres, pénétrant légèrement le bois,
  - une seconde d'un mélange huile, résine de pin, pigments donnant cette couleur rouge caractéristique.
- Technique classique pour l'époque.
- Le secret de Stradivarius n'est pas dans le vernis !

Le secret de Stradivarius



- Deux couches fines de vernis :
  - une première couche d'huile simple, similaire à celle des peintres, pénétrant légèrement le bois,
  - une seconde d'un mélange huile, résine de pin, pigments donnant cette couleur rouge caractéristique.
- Technique classique pour l'époque.
- Le secret de Stradivarius n'est pas dans le vernis !

Conclusion

- Cadre :
  - Problème de segmentation non supervisée.
  - Utilisation d'un modèle de mélange de gaussiennes localisé.
  - Localisation d'une méthode spectrale.
- Résultats :
  - Algorithme automatisé de segmentation non supervisée.
  - Algorithme efficace pour la minimisation (EM + Programmation dynamique).
  - Justification théorique à travers un problème d'estimation de densités conditionnelles.
- Perspectives :
  - Comparaison fine avec les méthodes *spectrales*, les méthodes *spatiales* et leurs combinaisons !
  - Lien précis entre l'estimation de densités conditionnelles et les performances de segmentation.
  - Calibration par heuristique de pente des deux problèmes.
  - Robustesse au bruit d'acquisition améliorée ?
  - Réduction de dimension adaptée à la classification non supervisée...

Conclusion

- Cadre :
  - Problème de segmentation non supervisée.
  - Utilisation d'un modèle de mélange de gaussiennes localisé.
  - Localisation d'une méthode spectrale.
- Résultats :
  - Algorithme automatisé de segmentation non supervisée.
  - Algorithme efficace pour la minimisation (EM + Programmation dynamique).
  - Justification théorique à travers un problème d'estimation de densités conditionnelles.
- Perspectives :
  - Comparaison fine avec les méthodes *spectrales*, les méthodes *spatiales* et leurs combinaisons !
  - Lien précis entre l'estimation de densités conditionnelles et les performances de segmentation.
  - Calibration par heuristique de pente des deux problèmes.
  - Robustesse au bruit d'acquisition améliorée ?
  - Réduction de dimension adaptée à la classification non supervisée...