BAYESIAN SEGMENTATION OF HYPERSPECTRAL IMAGES

Adel Mohammadpour, Nadia Bali and Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes

CNRS-ESE-UPS

Supélec, Plateau de Moulon

91192 Gif-sur-Yvette, FRANCE.

Adel.Mohammadpour@lss.supelec.fr

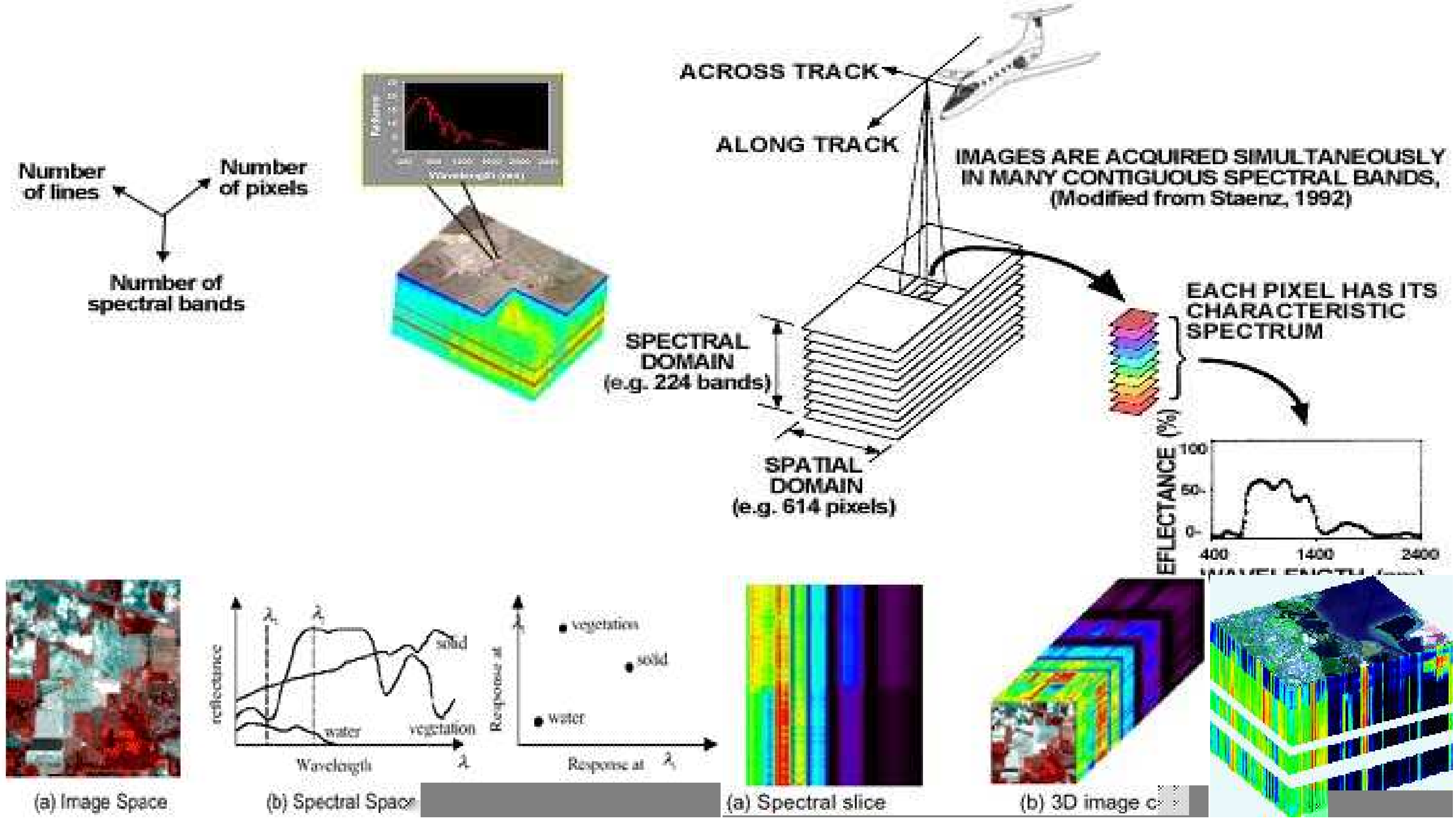
Olivier.Feron@lss.supelec.fr

djafari@lss.supelec.fr

Contents

- Introduction to hyperspectral images
- Spectral classification methods
  (Spatial distribution of spectra is neglected)
- Spatial classification methods
  (Spectral shapes of the voxels are neglected)
- Modeling for segmentation methods which account for both spatial and spectral structures of the data
- Bayesian approach and general MCMC Gibbs sampling
- Comparison of the proposed approach with classical methods
- Proposed algorithm
- Simulation results
- Conclusions

Introduction to hyperspectral images

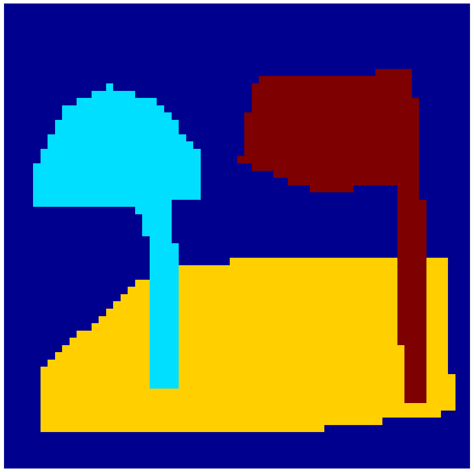


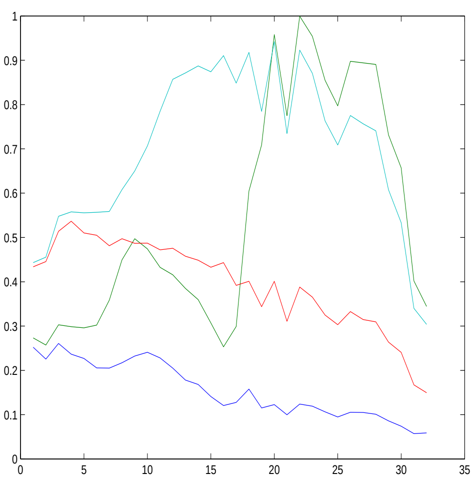
Classification, segmentation and data reduction

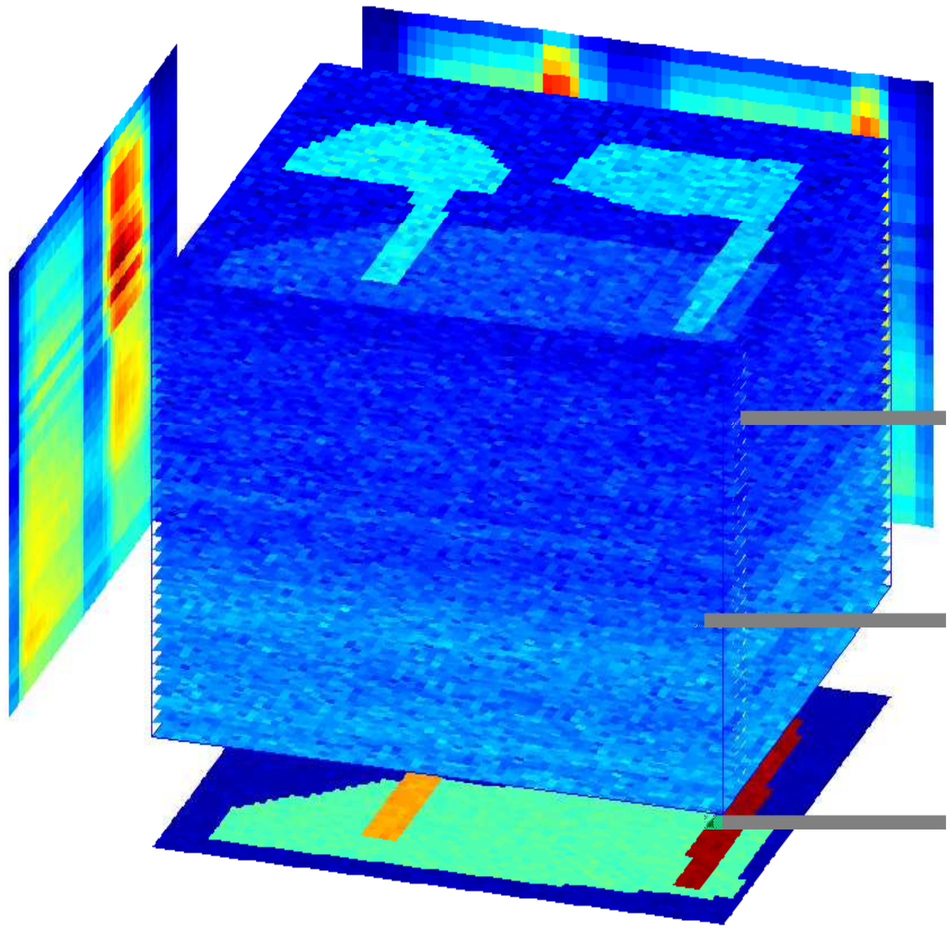
- Hyperspectral data : $g(\omega, \boldsymbol{r})$
  - A set of spectra : $g_{\boldsymbol{r}}(\omega)$
  - A set of images : $g_{\omega}(\boldsymbol{r})$
- Redundancy due to spectral and spatial structure
- Main objectif 1 : Find the type of materials in a given position (labeling)
  - Classification
  - Segmentation
- Main objectif 2 : Data reduction and compression
  - ACP, ACI and using classification and segmentation for better compression.
  - Proposing a method which does data reduction and segmentation at the same time.

Generating simulated data

Synthetic data 1:



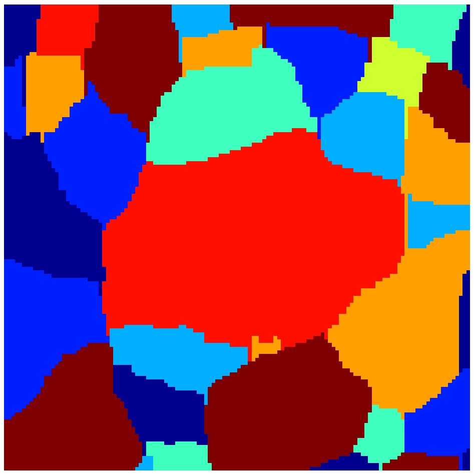


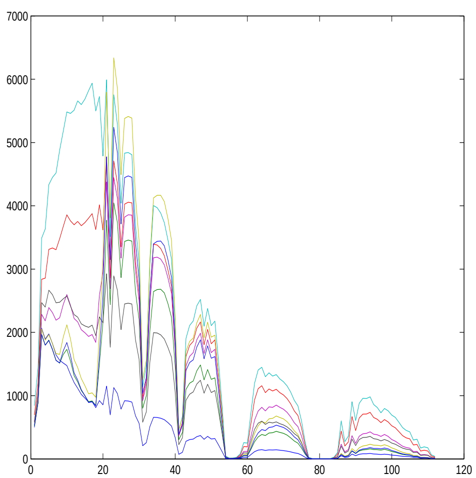


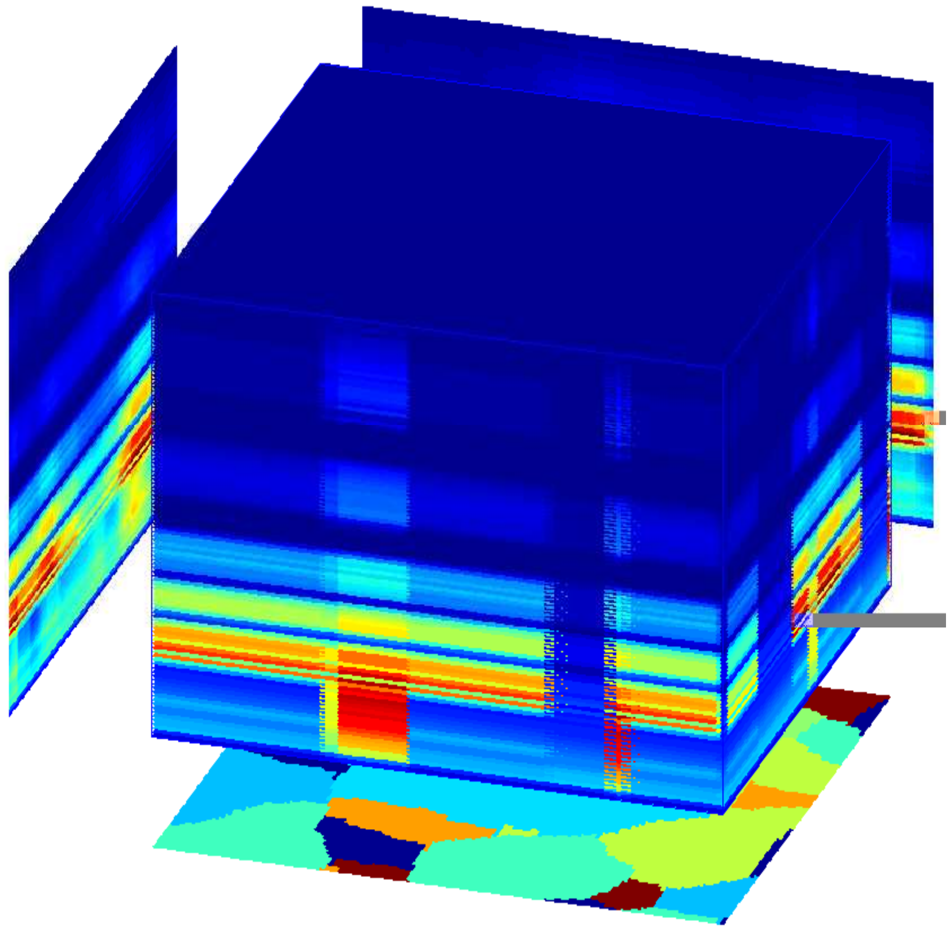
4 classes, 32 images (64x64)

Generating simulated data

Synthetic data 2:

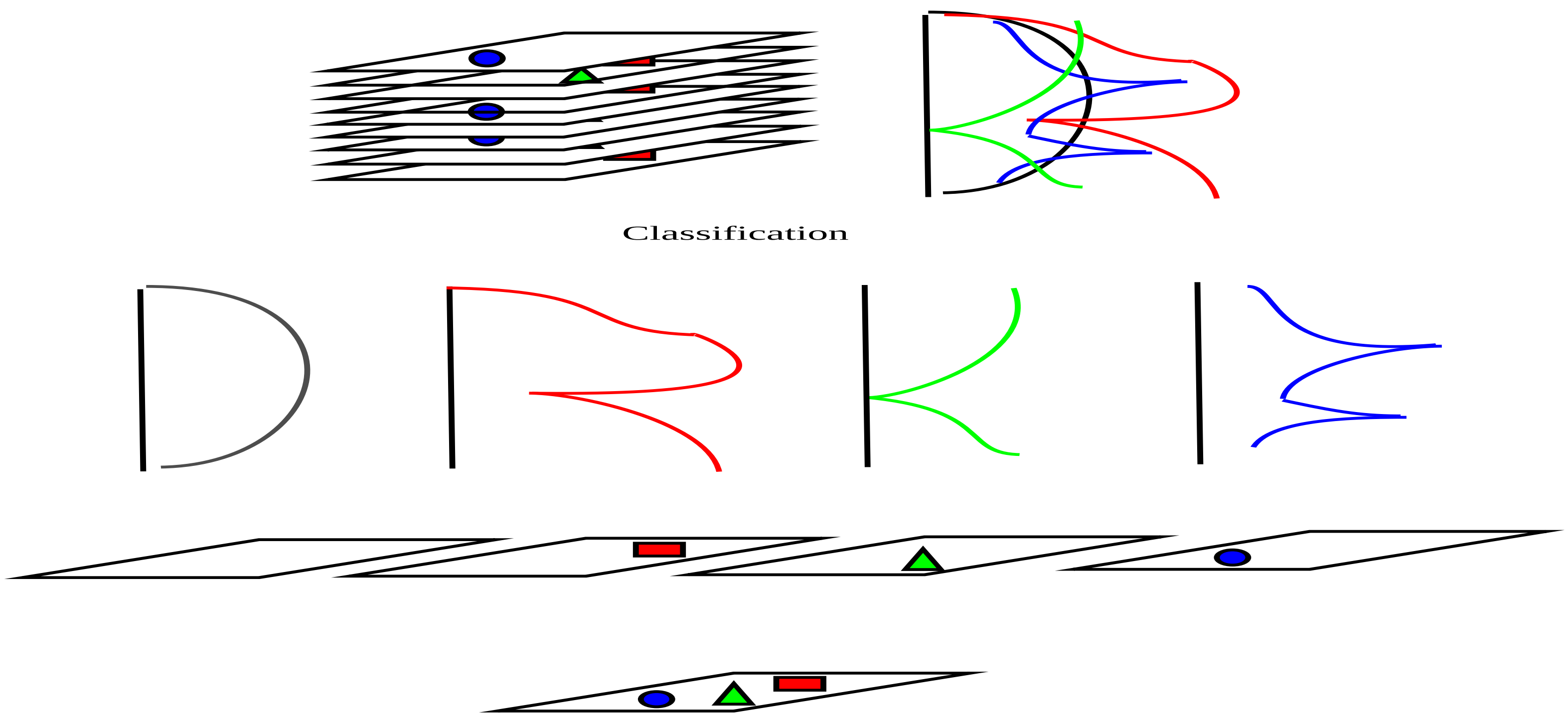




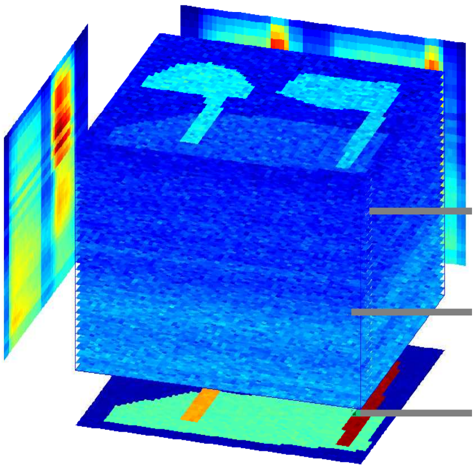


8 classes, 112 images (128x128)

Spectral classification methods

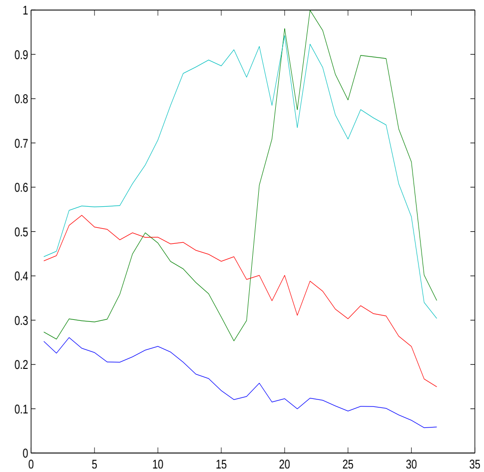


- Each spectral line is considered as a point in a vectorial space
- Different classification methods are used: K-means, mixture of gaussians, ...
- Spatial distribution of the spectra is neglected

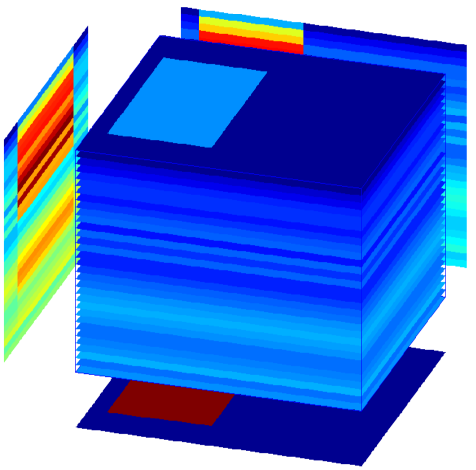


32 images

Original

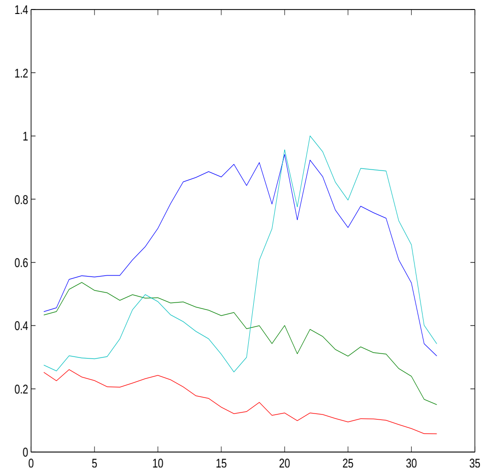


4 classes

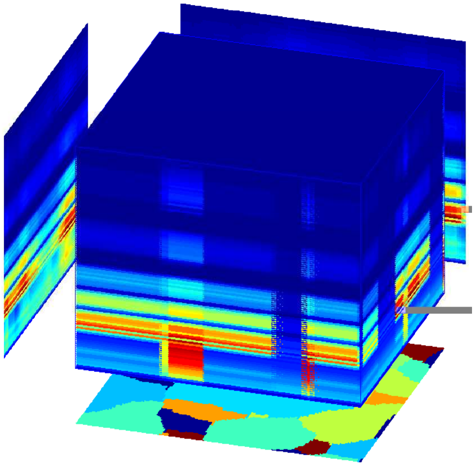


32 images

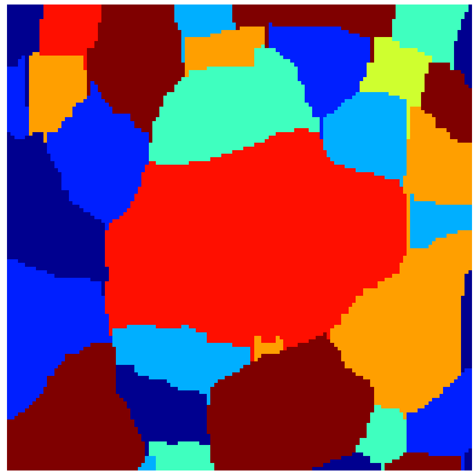
Estimated



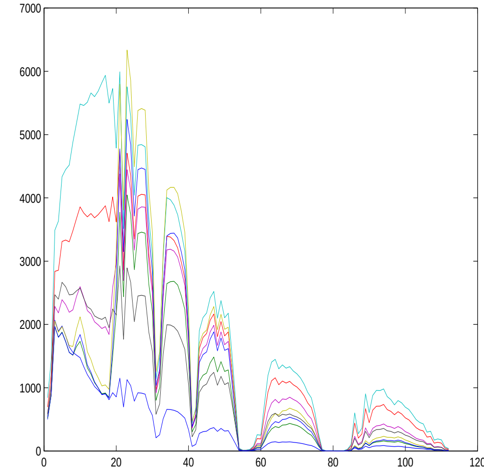
4 classes



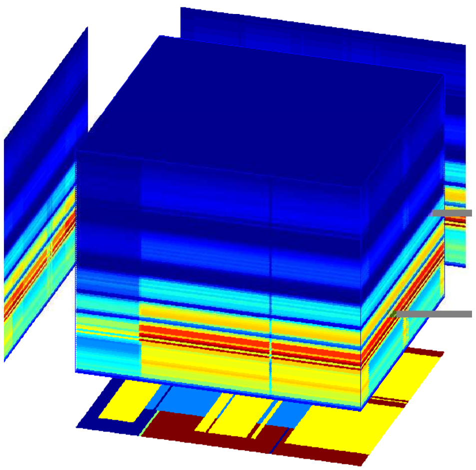
112 images



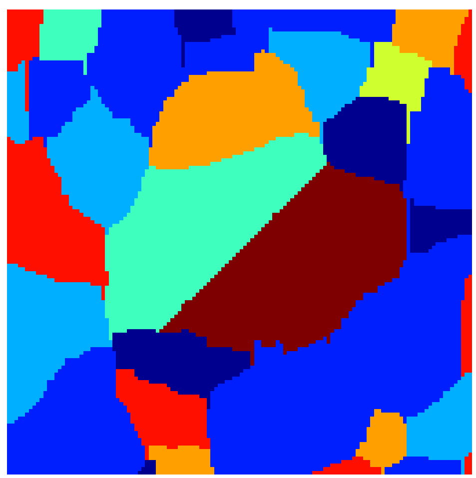
Original



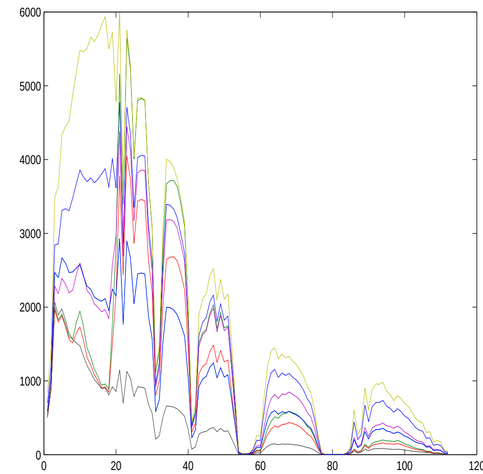
8 classes



112 images

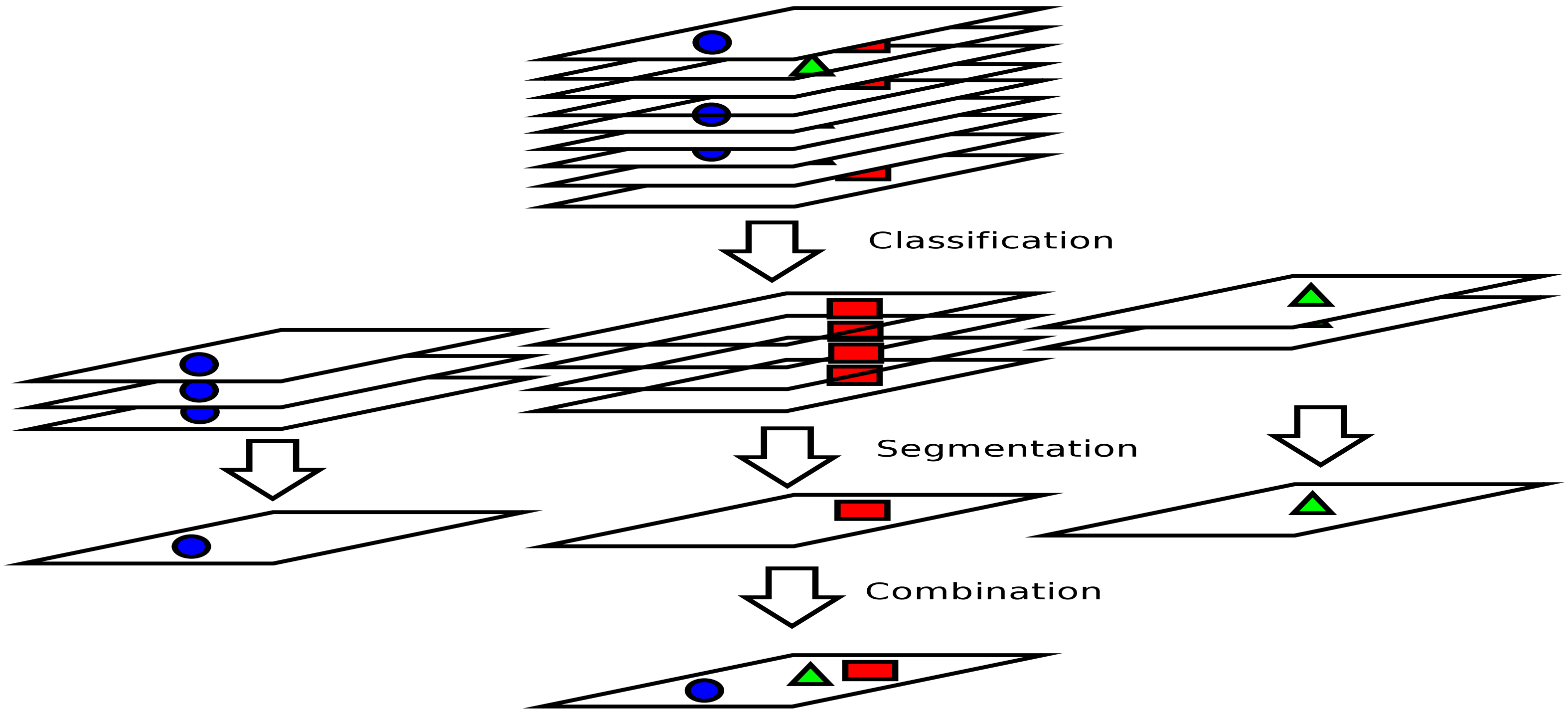


Estimated

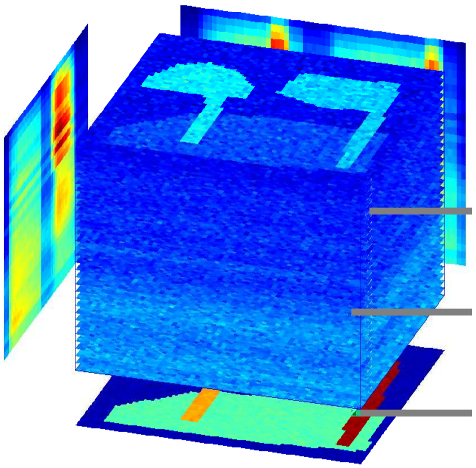


8 classes

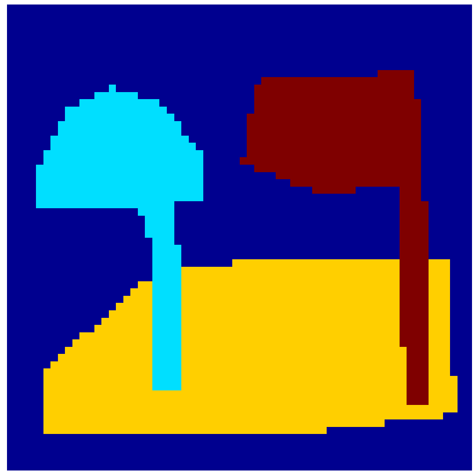
Spatial classification methods



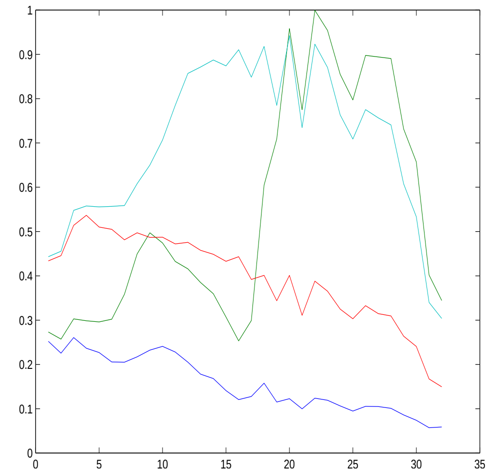
- Each image is considered as a point in a vectorial space
- Different classification methods are used: K-means, mixture of gaussians, ...
- Spectral structures of the image pixels are neglected



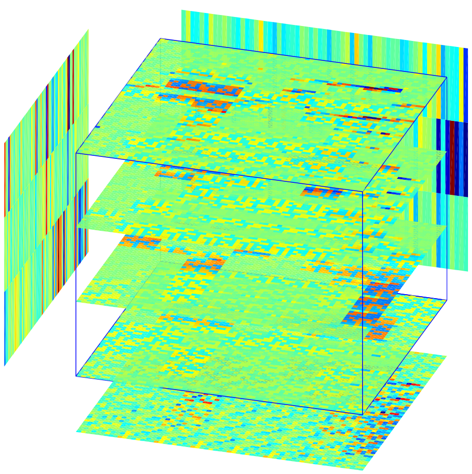
32 images



Original

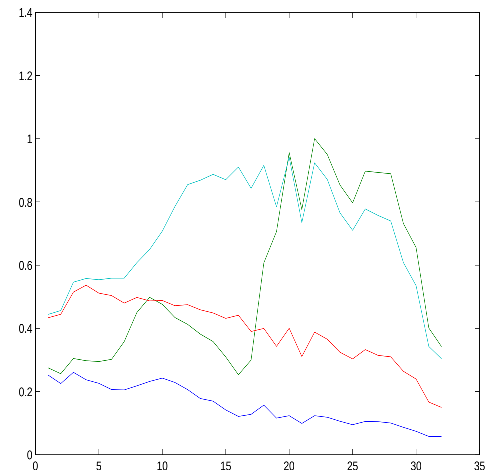


4 classes

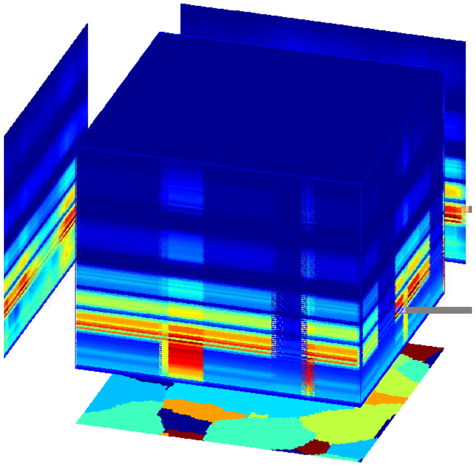


8 images

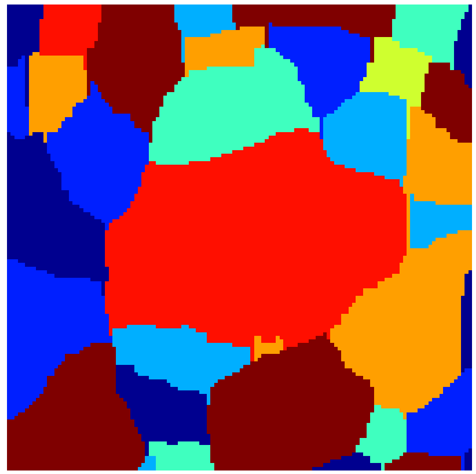
Estimated



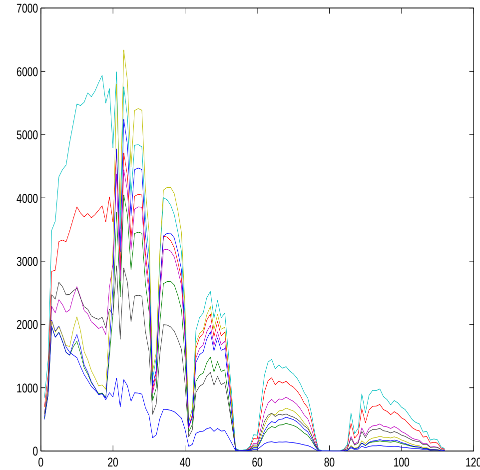
4 classes



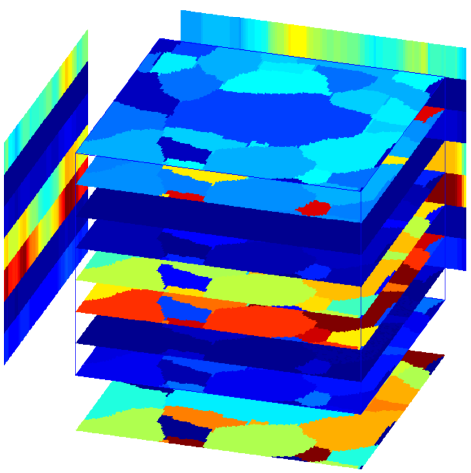
112 images



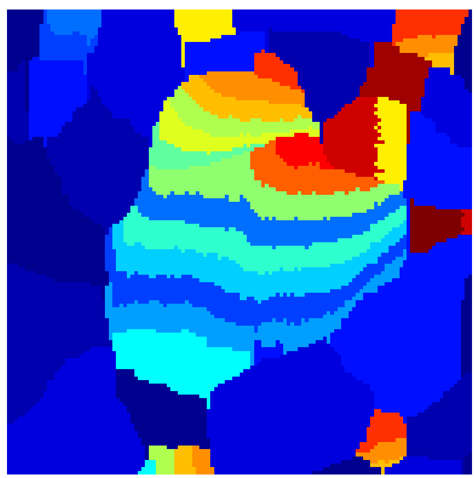
Original



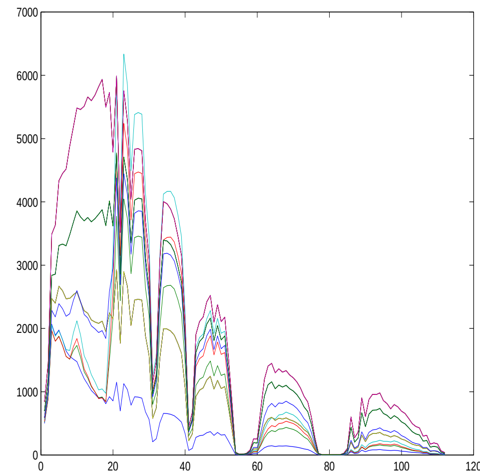
8 classes



8 images

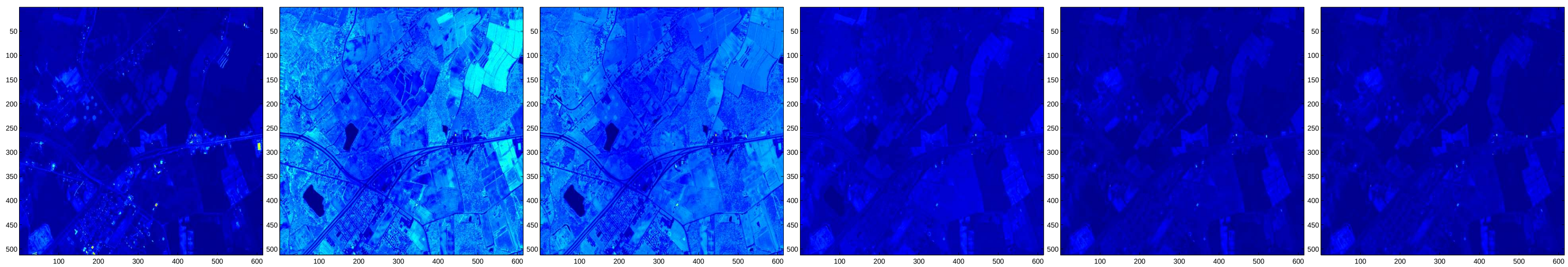


Estimated



8 classes

Modelling for accounting for both spatial and spectral structures



$g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}), \quad i = 1, \cdots, M$

$\underline{g}(\boldsymbol{r}) = \{g_i(\boldsymbol{r}),\ i = 1, M\}$

$\underline{g}(\boldsymbol{r}) = \underline{f}(\boldsymbol{r}) + \underline{\epsilon}(\boldsymbol{r})$

$\boldsymbol{g}_i = \{g_i(\boldsymbol{r}),\ \boldsymbol{r} \in \mathcal{R}\}$

$\underline{\boldsymbol{g}} = \{\boldsymbol{g}_i(\boldsymbol{r}),\ i = 1, M\}$

$\underline{\boldsymbol{g}} = \underline{\boldsymbol{f}} + \underline{\boldsymbol{\epsilon}}$

Segmentation:

– Hidden variables rep. regions

$z(\boldsymbol{r}) = k,\ k = 1, \cdots, K$

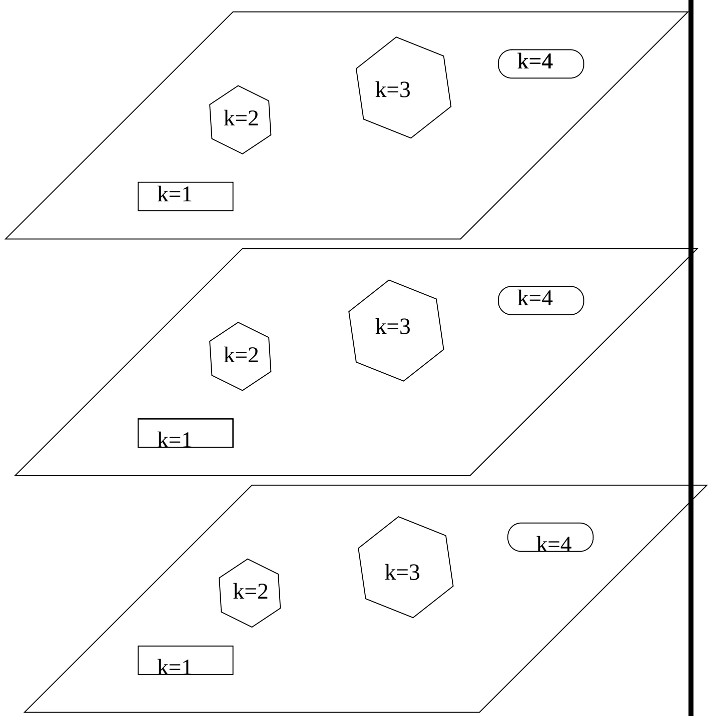
$\mathcal{R}_k = \{\boldsymbol{r} \ :\ z(\boldsymbol{r}) = k\},\ \mathcal{R} = \cup_k \mathcal{R}_k$

– Homogeneity in regions:

$p(f_i(\boldsymbol{r}) | z(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2)$

Modelling for accounting for both spatial and spectral structures

$$\begin{cases} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}), \quad i = 1, \cdots, M \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma^2_{i\,k}) \end{cases}$$



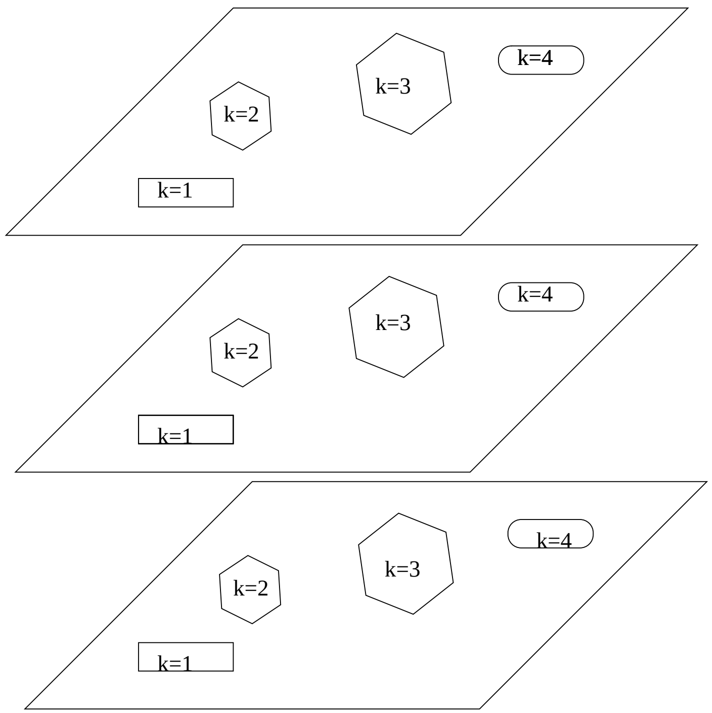
Prior hypothesis about $f_i(\boldsymbol{r})$:

- Pixels values of $f_i(\boldsymbol{r})$ in different regions of an image are independent. They may share however the same parameters $\theta_{ik} = (m_{i\,k}, \sigma^2_{i\,k})$
- For pixels values in a given region of an image, two possibilities:
  - i.i.d.: $p(f_j(\boldsymbol{r})|z_j(\boldsymbol{r}) = k) = \mathcal{N}(m_{j\,k}, \sigma^2_{j\,k})$
    $p(f_j(\boldsymbol{r}), \boldsymbol{r} \in R_{j\,k}) = \mathcal{N}(m_{j\,k}\mathbf{1}, \sigma^2_{j\,k}\boldsymbol{I})$
  - Markovien: $p(f_j(\boldsymbol{r}), \boldsymbol{r} \in R_{j\,k}) = \mathcal{N}(m_{j\,k}\mathbf{1}, \boldsymbol{\Sigma}_{j\,k})$

Modelling for accounting for both spatial and spectral structures

$$\left\{ \begin{array}{l} g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}), \quad i = 1, \cdots, M \\ p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2) \end{array} \right.$$

Pixels along the channels represent spectra



- A Markovien model for $f_i(\boldsymbol{r})$:

$$p(f_i(\boldsymbol{r})|z(\boldsymbol{r}) = k, f_{i-1}(\boldsymbol{r})) = \mathcal{N}(\psi_{ik} f_{i-1}(\boldsymbol{r}), \sigma_{i\,k}^2)$$

$$f_{ik}(\boldsymbol{r}) = \psi_{ik} f_{i-1,k}(\boldsymbol{r}) + \eta_{ik} \quad \sim AR(1)$$

- A Markovien model for the means:

$$p(f_i(\boldsymbol{r})|z_i(\boldsymbol{r}) = k) = \mathcal{N}(m_{i\,k}, \sigma_{i\,k}^2)$$

$$m_{i\,k} = \phi_k m_{i-1,k} + \eta_k \quad \sim AR(1)$$

Modeling the labels

$$p(f_j(\boldsymbol{r})|z_j(\boldsymbol{r})=k)=\mathcal{N}(m_{j\,k},\sigma^2_{j\,k}) \longrightarrow p(f_j(\boldsymbol{r}))=\sum_k P(z_j(\boldsymbol{r})=k)\,\mathcal{N}(m_{j\,k},\sigma^2_{j\,k})$$

- Independent Gaussian Mixture model (IGM), where $\boldsymbol{z}_j=\{z_j(\boldsymbol{r}),\boldsymbol{r}\in\mathcal{R}\}$ are assumed to be independent and

$$P(z_j(\boldsymbol{r})=k)=p_k, \quad \text{with} \quad \sum_k p_k=1 \;\text{ and }\; p(\boldsymbol{z}_j)=\prod_k p_k$$

- Contextual Gaussian Mixture model (CGM): $\boldsymbol{z}_j$ Markovien

$$p(\boldsymbol{z}_j)\propto \exp\left[\alpha \sum_{\boldsymbol{r}\in\mathcal{R}}\sum_{\boldsymbol{s}\in\mathcal{V}(\boldsymbol{r})}\delta(z_j(\boldsymbol{r})-z_j(\boldsymbol{s}))\right]$$

which is the Potts Markov random feild (PMRF).
The parameter α controls the mean value of the regions' sizes.

Expressions of likelihood, prior and posterior laws

$$g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r})$$

$$\boldsymbol{g}_i = \boldsymbol{f}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \cdots, M \quad \longrightarrow \quad \underline{\boldsymbol{g}} = \underline{\boldsymbol{f}} + \underline{\boldsymbol{\epsilon}}$$

- Likelihood: $\boldsymbol{\theta}_1 = \{\sigma_{\epsilon_i}^2, i = 1, \cdots, M\}, \quad \longrightarrow \quad \boldsymbol{\Sigma}_{\epsilon i} = \sigma_{\epsilon_i}^2 \boldsymbol{I}$

$$p(\underline{\boldsymbol{g}}|\underline{\boldsymbol{f}}, \boldsymbol{\theta}_1) = \prod_{i=1}^{M} p(\underline{\boldsymbol{g}}|\underline{\boldsymbol{f}}, \boldsymbol{\Sigma}_{\epsilon i}) = \prod_{i=1}^{M} \mathcal{N}(\underline{\boldsymbol{f}}, \boldsymbol{\Sigma}_{\epsilon i})$$

- HMM for the images: $\boldsymbol{\theta}_2 = \{(m_{i\,k}, \sigma_{i\,k}^2), j = 1, \cdots, M\}$
  - Markovian model for $\boldsymbol{f}_i|\boldsymbol{z}$:

$$p(\underline{\boldsymbol{f}}|\boldsymbol{z}, \boldsymbol{\theta}_2) = p(\boldsymbol{f}_1|\boldsymbol{z}, m_{i\,k}, \sigma_{i\,k}^2) \prod_{i=2}^{M} p(\boldsymbol{f}_i|\boldsymbol{f}_{i-1}, \boldsymbol{z}, m_{i\,k}, \sigma_{i\,k}^2)$$

  - Markovian model for $m_{i\,k}$:

$$p(\underline{\boldsymbol{f}}|\boldsymbol{z}, \boldsymbol{\theta}_2) = \prod_{i=1}^{M} p(\boldsymbol{f}_i|\boldsymbol{z}, m_{i\,k}, \sigma_{i\,k}^2)$$

but $\quad m_{i\,k} = \phi_k m_{i-1,k} + \eta_{ik} \sim AR(1)$

- PMRF for the labels:

$$p(\boldsymbol{z}) \propto \exp\left[\alpha \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{s}\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{s}))\right]$$

- Conjugate priors for the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$:
$\boldsymbol{\theta} = \{\{\sigma_{\epsilon_i}^{\,2}, i = 1, \cdots, M\}, \{(m_{i\,k}, \sigma_{i\,k}^2), i = 1, \cdots, M, k = 1, \cdots, K\}\}$

$$\begin{aligned} p(\sigma_{\epsilon\,i}) &= \mathcal{IG}(\alpha_{i0}, \beta_{i0}) \\ p(m_{i\,k}) &= \mathcal{N}(\phi_k m_{i-1\,k}, \sigma_{i\,k0}^2) \\ p(\sigma_{i\,k}^2) &= \mathcal{IG}(\alpha_{i0}, \beta_{i0}) \\ p(\boldsymbol{\Sigma}_{i\,k}) &= \mathcal{IW}(\alpha_{i0}, \Lambda_{i0}) \end{aligned}$$

- Joint posterior law of $\underline{\boldsymbol{f}}$, $\boldsymbol{z}$ and $\underline{\boldsymbol{\theta}}$

$$p(\underline{\boldsymbol{f}}, \boldsymbol{z}, \underline{\boldsymbol{\theta}}|\underline{\boldsymbol{g}}) \propto p(\underline{\boldsymbol{g}}|\underline{\boldsymbol{f}}, \boldsymbol{\theta}_1)\; p(\underline{\boldsymbol{f}}|\boldsymbol{z}, \boldsymbol{\theta}_2)\; p(\boldsymbol{z}|\alpha)\; p(\underline{\boldsymbol{\theta}})$$

General MCMC sampling scheme

$$\begin{aligned} p(\underline{\boldsymbol{f}}, \boldsymbol{z}, \underline{\boldsymbol{\theta}} | \underline{\boldsymbol{g}}) & \propto p(\underline{\boldsymbol{g}} | \underline{\boldsymbol{f}}, \boldsymbol{\theta}_1)\, p(\underline{\boldsymbol{f}} | \boldsymbol{z}, \boldsymbol{\theta}_2)\, p(\boldsymbol{z} | \boldsymbol{\theta}_2) \\ & \propto \left(\prod_i p(\boldsymbol{g}_i | \boldsymbol{f}_i, \boldsymbol{\theta}_1)\right)\; p(\underline{\boldsymbol{f}} | \boldsymbol{z}, \boldsymbol{\theta}_2)\, p(\boldsymbol{z} | \alpha)\, p(\underline{\boldsymbol{\theta}}) \end{aligned}$$

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \begin{cases} \boldsymbol{\theta}_1 = \{\sigma_{\epsilon_i}^2, i = 1, \cdots, M\} \\ \boldsymbol{\theta}_2 = \{(m_{i\,k}, \sigma_{i\,k}^2), i = 1, \cdots, M, k = 1, \cdots, K\} \end{cases}$$

Gibbs sampling:

- Generate samples $(\underline{\boldsymbol{f}}, \underline{\boldsymbol{z}}, \underline{\boldsymbol{\theta}})^{(1)}, \cdots, (\underline{\boldsymbol{f}}, \underline{\boldsymbol{z}}, \underline{\boldsymbol{\theta}})^{(N)}$ using
  - $\underline{\boldsymbol{f}} \sim p(\underline{\boldsymbol{f}} | \underline{\boldsymbol{g}}, \boldsymbol{z}, \underline{\boldsymbol{\theta}}) \propto p(\underline{\boldsymbol{g}} | \underline{\boldsymbol{f}}, \underline{\boldsymbol{z}}, \underline{\boldsymbol{\theta}})\, p(\underline{\boldsymbol{f}} | \boldsymbol{z})$
  - $\boldsymbol{z} \sim p(\boldsymbol{z} | \underline{\boldsymbol{g}}, \underline{\boldsymbol{f}}, \underline{\boldsymbol{\theta}}) \propto p(\underline{\boldsymbol{g}} | \underline{\boldsymbol{f}}, \boldsymbol{z}, \underline{\boldsymbol{\theta}})\, p(\underline{\boldsymbol{f}} | \boldsymbol{z})\, p(\boldsymbol{z} | \alpha)$
  - $\underline{\boldsymbol{\theta}} \sim p(\underline{\boldsymbol{\theta}} | \underline{\boldsymbol{g}}, \underline{\boldsymbol{f}}, \boldsymbol{z})$
- Compute any statistics such as mean, median, variance, ...

Comparison with classical methods

Classical

Observation model: no noise

$g_i(\boldsymbol{r}) = f_i(\boldsymbol{r})$

Ind. Gaussian Mixture model:

$p(f_i(\boldsymbol{r})|z(\boldsymbol{r})=k) = \mathcal{N}(m_{i\,k}, \sigma^2_{i\,k})$

$z(\boldsymbol{r}) \perp z(\boldsymbol{s}), \quad \boldsymbol{s} \neq \boldsymbol{r},$

$p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z})$

No correlation in diff. channels:

$m_{i\,k} \perp m_{j\,k}, \quad i \neq j$

Proposed

Observation model: accounts for noise

$g_i(\boldsymbol{r}) = f_i(\boldsymbol{r}) + \epsilon_i(\boldsymbol{r}), \quad \boldsymbol{r} \in \mathcal{R},$

Hidden Markov Model:

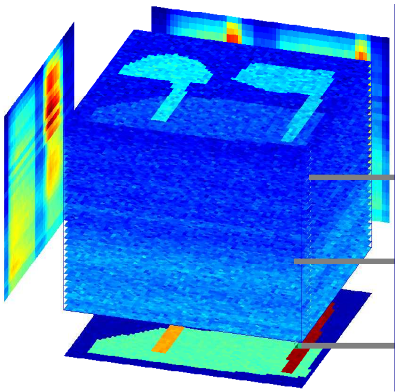
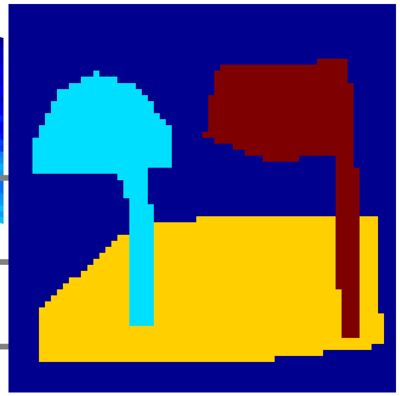
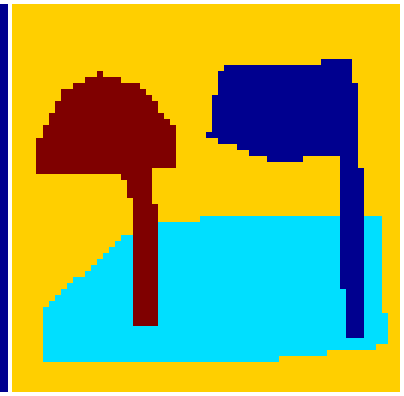
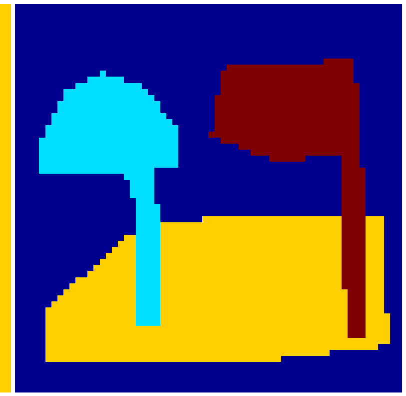
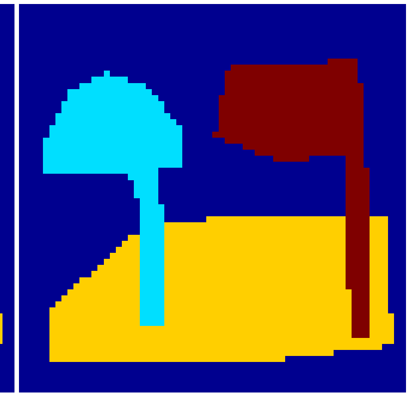
$p(f_i(\boldsymbol{r})|z(\boldsymbol{r})=k) = \mathcal{N}(m_{i\,k}, \sigma^2_{i\,k})$

$p(\boldsymbol{z}) \propto \exp\left[\alpha \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{s}\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{s}))\right]$

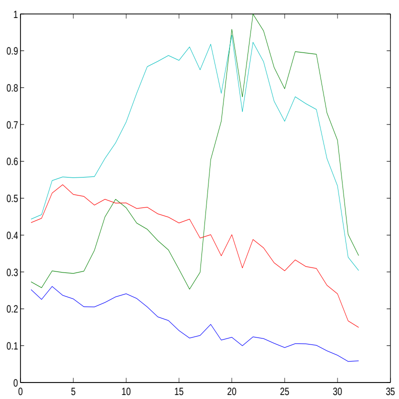
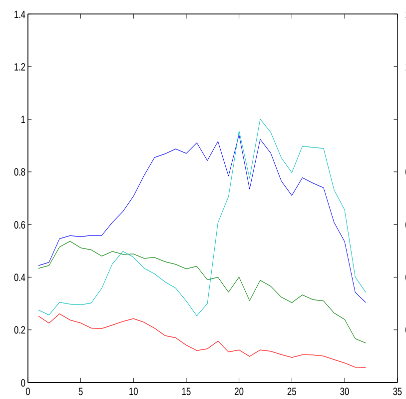
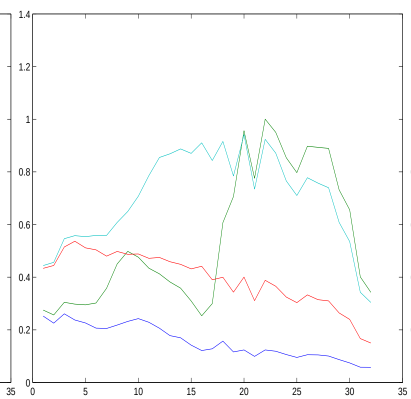
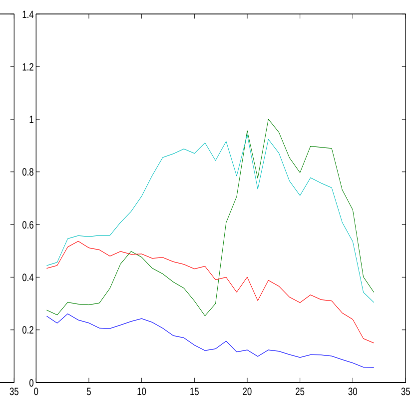
$p(\underline{\boldsymbol{f}}|\boldsymbol{z}) = \prod_i p(\boldsymbol{f}_i|\boldsymbol{z})$

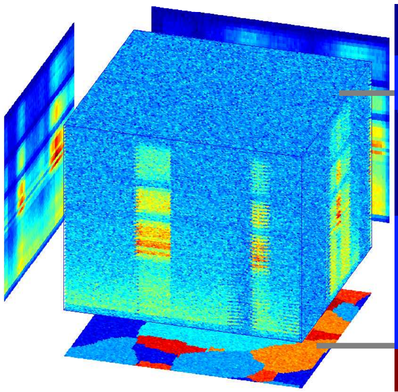
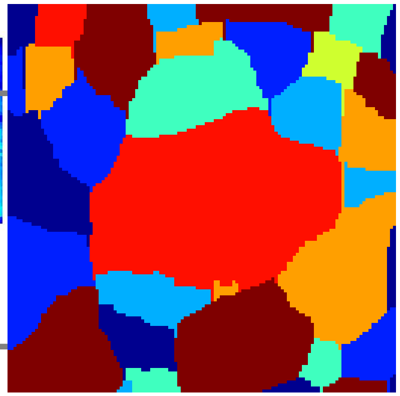
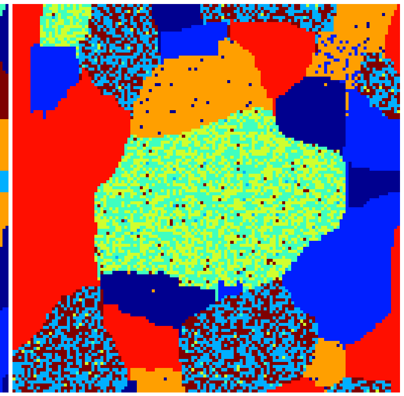
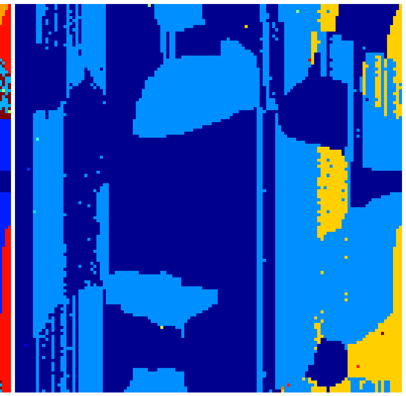
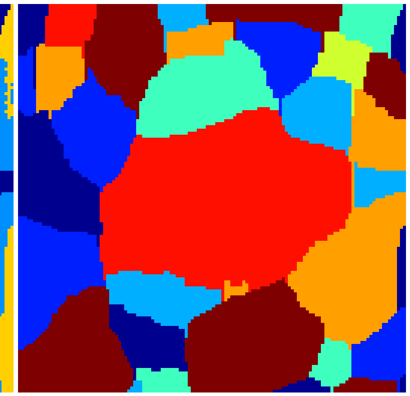
Accounts for correlation in diff. channels:

$m_{i\,k} = \phi_k m_{i-1\,k} + \eta_{ik}$

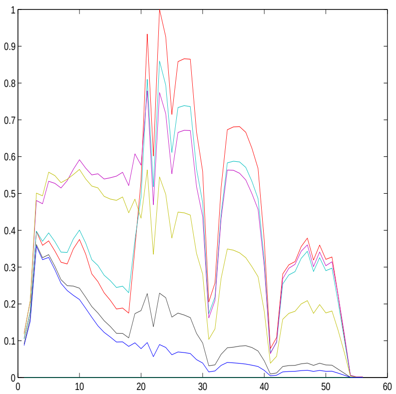
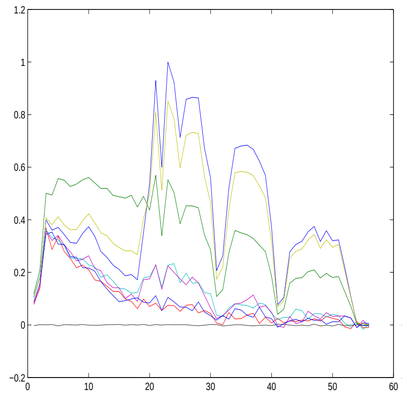
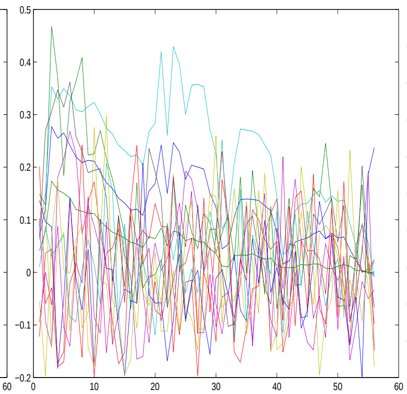
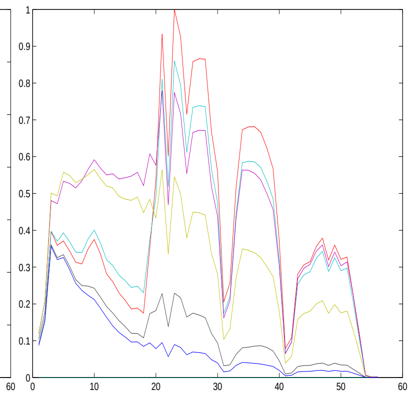






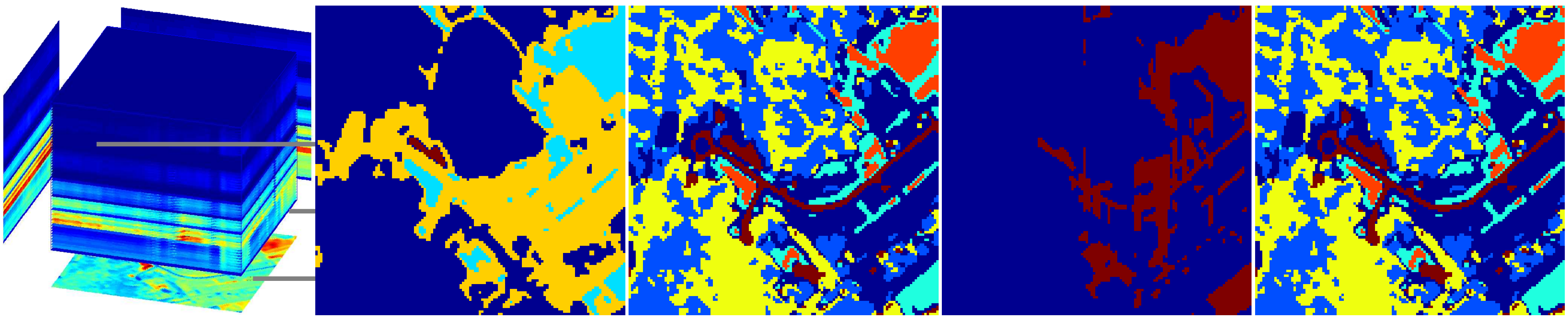
Data Original Kmeans1 Kmeans2 BFJsegment

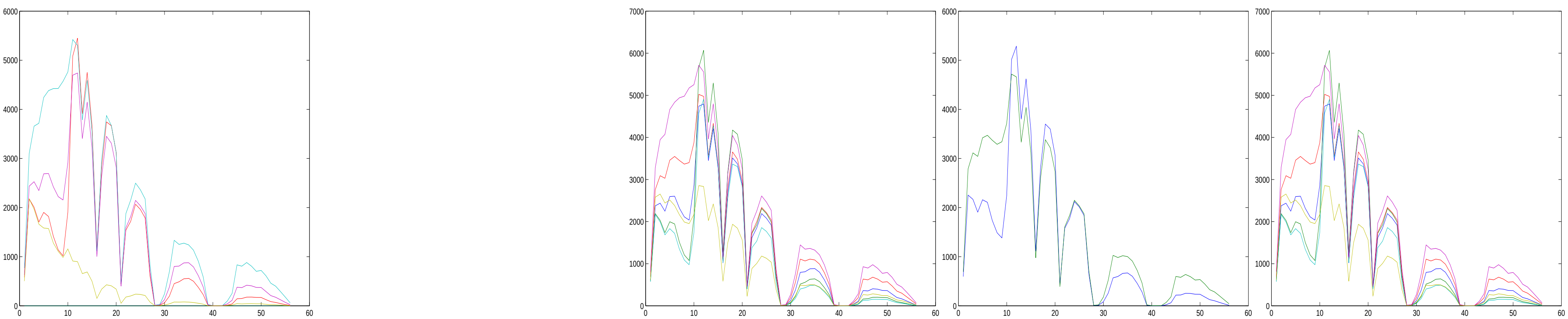
    

Data Original Kmeans1 Kmeans2 BFJsegment



Data Original Kmeans1 Kmeans2 BFJsegment



DataOriginalKmeans1Kmeans2BFJsegment

DataOriginalKmeans1Kmeans2BFJsegment

DataOriginalKmeans1Kmeans2BFJsegment