Unmixing hyperspectral images using Markov random fields

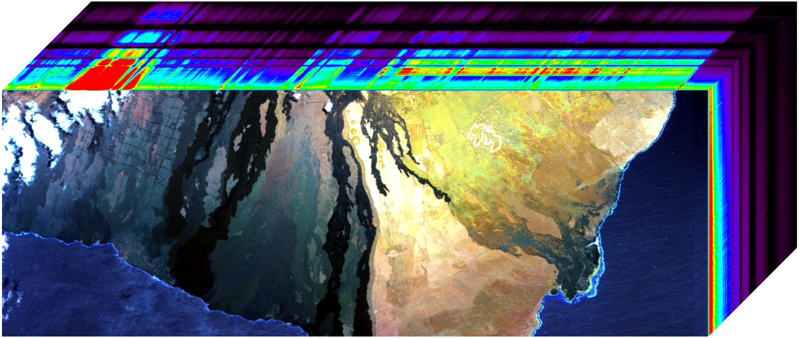
Olivier Eches, Nicolas Dobigeon and Jean-Yves Tourneret

Université de Toulouse - IRIT/INP-ENSEEIHT Toulouse, FRANCE
http://www.enseeiht.fr/~eches

MaxEnt 2010, Chamonix, France

What is an hyperspectral image?

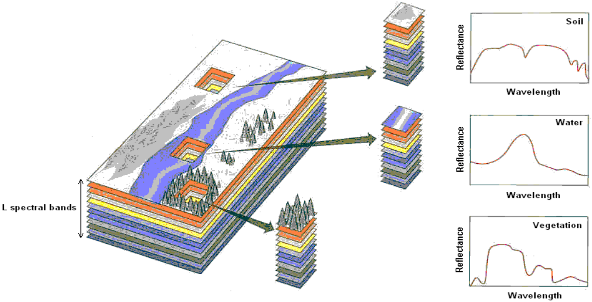
- Same scene observed in different spectral bands
- 3-Dimension image: length, width and wavelength



Example of hyperspectral cube (Big Island, Hawaii, USA)

Single pixel

Pixel represented by a vector of hundreds of measurements



Applications

- mineral exploration
- agriculture: soil quality, crop forecasting, forest monitoring
- environment: pollution detection, climatic change detection
- military: target detection (minefields, vehicles,...), cartography

Spectral mixing

Measured pixel: mixture of pure spectra (*endmembers*) characterized by their corresponding fractions (*abundances*).

Common assumption: Linear mixing model

If the pure materials are spatially disjoint in the pixel, the measured spectrum is the linear combination of the corresponding pure spectra.

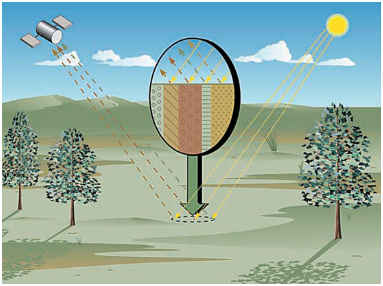


Figure: Linear observation model

Linear mixing model (LMM)

For a given pixel p:

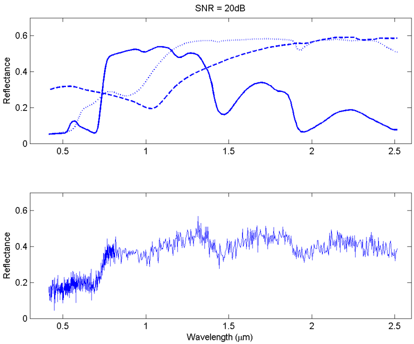
$$\boldsymbol{y}_p = \sum_{r=1}^{R} \boldsymbol{m}_r a_{r,p} + \boldsymbol{n}_p,$$

- $\boldsymbol{y}_p = [y_{1,p}, \ldots, y_{L,p}]^T$ the observed pixel p in L bands,
- R number of pure materials or *endmembers*,
- $\boldsymbol{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ the spectrum of the rth endmember,
- $a_{r,p}$ fraction or *abundance* of the rth endmember in the pth pixel,
- $\boldsymbol{n}_p = [n_{1,p}, \ldots, n_{L,p}]^T$ the additive noise in the pth observed pixel (assumed white Gaussian),

Constraints on the abundance vectors $\boldsymbol{a}_p = [a_{1,p}, \ldots, a_{R,p}]^T$

$$\begin{cases} a_{r,p} \geq 0, \ \forall r = 1, \ldots, R \\ \sum_{r=1}^{R} a_{r,p} = 1. \end{cases} \tag{1}$$

Linear Mixing Model (LMM): $\boldsymbol{y}_p = \sum_{r=1}^{R} \boldsymbol{m}_r a_{r,p} + \boldsymbol{n}_p$



- $L = 825$ (0.4μm $\to 2.5\mu$m),
- $R = 3$:
  - green grass (solid line),
  - galvanized steel metal (dashed line),
  - bare red brick (dotted line),
- $\boldsymbol{a}_p = [0.4, 0.2, 0.4]^T$,
- SNR $\approx$ 20dB.

Spectral unmixing problem

Estimation of $\{\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R\}$ and $\boldsymbol{\alpha}_p$.

Unmixing steps

1. Endmember extraction step: estimation of R and $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R$ (Vertex Component Analysis, N-FINDR, Pixel Purity Index,...)
2. Inversion step: estimation of the corresponding abundances $\boldsymbol{a}_p = [a_{1,p}, \ldots, a_{R,p}]^T$ (LS, ML and Bayesian approaches)

Most inversion strategies ignore the possible interactions between the pixels.

Unmixing steps

1. Endmember extraction step: estimation of R and $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R$ (Vertex Component Analysis, N-FINDR, Pixel Purity Index,...)
2. Inversion step: estimation of the corresponding abundances $\boldsymbol{a}_p = [a_{1,p}, \ldots, a_{R,p}]^T$ (LS, ML and Bayesian approaches)

Most inversion strategies ignore the possible interactions between the pixels.

Problem addressed in this work

Estimation of $\boldsymbol{a}_p$ under positivity and additivity constraints.

Main contributions

- ☞ Exploiting spatial correlations in a new Bayesian inversion procedure
- ☞ Using Markov random fields (MRFs) to model spatial interactions
- ☞ Spatial correlations $\Rightarrow$ image classification/segmentation

1. Introducing spatial structures
2. Hierarchical Bayesian model
   - Likelihood
   - Prior distributions
   - Joint posterior distribution
3. Simulations
   - Synthetic data
   - Real data
4. Conclusions

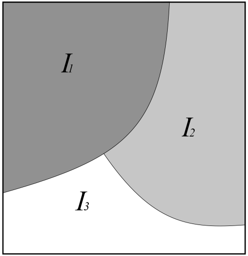
Outline

Image partitioning

Defining homogeneous regions

- Image of P pixels divided into K regions or *classes*.
- In each region, the pixels approximately share the same composition.
- Introducing hidden variable: label vector $\boldsymbol{z} = [z_1, \ldots, z_P]^T$, where

$$z_p = k \Leftrightarrow p \in \mathcal{I}_k, \ \ k = \{1, \ldots, K\}.$$



- $\forall p \in \mathcal{I}_1$: $z_p = 1$, $\mathrm{E}[\boldsymbol{a}_p] = \boldsymbol{\mu_1}$, $\mathrm{Covar}(\boldsymbol{a}_p) = \boldsymbol{\Gamma}_1$,
- $\forall p \in \mathcal{I}_2$: $z_p = 2$, $\mathrm{E}[\boldsymbol{a}_p] = \boldsymbol{\mu_2}$, $\mathrm{Covar}(\boldsymbol{a}_p) = \boldsymbol{\Gamma}_2$,
- $\forall p \in \mathcal{I}_3$: $z_p = 3$, $\mathrm{E}[\boldsymbol{a}_p] = \boldsymbol{\mu_3}$, $\mathrm{Covar}(\boldsymbol{a}_p) = \boldsymbol{\Gamma}_3$.

Image partitioning

Example

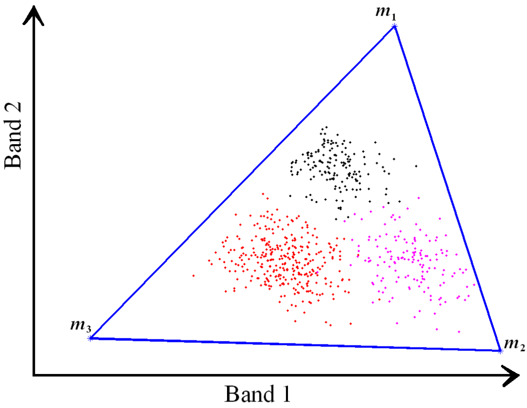


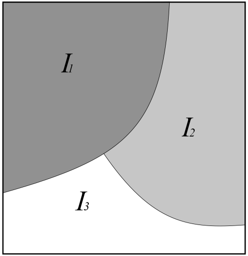
Figure: Scatter-plot of dual-band data of an image partitioned in 3 classes.

Image partitioning

Defining homogeneous regions

- Image of P pixels divided into K regions or *classes*.
- In each region, the pixels approximately share the same composition.
- Introducing hidden variable: label vector $\boldsymbol{z} = [z_1, \ldots, z_P]^T$, where

$$z_p = k \Leftrightarrow p \in \mathcal{I}_k, \;\; k = \{1, \ldots, K\}.$$



- $\forall p \in \mathcal{I}_1$: $z_p = 1$, $\mathrm{E}[\boldsymbol{a}_p] = \boldsymbol{\mu_1}$, $\mathrm{Covar}(\boldsymbol{a}_p) = \boldsymbol{\Gamma}_1$,
- $\forall p \in \mathcal{I}_2$: $z_p = 2$, $\mathrm{E}[\boldsymbol{a}_p] = \boldsymbol{\mu_2}$, $\mathrm{Covar}(\boldsymbol{a}_p) = \boldsymbol{\Gamma}_2$,
- $\forall p \in \mathcal{I}_3$: $z_p = 3$, $\mathrm{E}[\boldsymbol{a}_p] = \boldsymbol{\mu_3}$, $\mathrm{Covar}(\boldsymbol{a}_p) = \boldsymbol{\Gamma}_3$.

Image partitioning

Abundance reparametrization

Introducing *logistic* coefficients[a] $\boldsymbol{t}_p = [t_{1,p} \ldots, t_{R,p}]^T$ where:

$$a_{r,p} = \frac{\exp(t_{r,p})}{\sum_{r=1}^{R} \exp(t_{r,p})}. \tag{2}$$

$\Rightarrow$ Ensure positivity and sum-to-one constraints.

$\Rightarrow$ Each class k fully characterized by:

$$\mathrm{E}\left[\boldsymbol{t}_p | z_p = k\right] = \boldsymbol{\psi}_k,$$
$$\mathrm{Covar}\left[\boldsymbol{t}_p | z_p = k\right] = \boldsymbol{\Sigma}_k.$$

[a][A. Gelman *et al.*, *J. Amer. Math. Soc.*, 1996]

Outline

Unknown parameter vector

Unknown parameter vector

$$\boldsymbol{\Theta} = \{\boldsymbol{T}, \boldsymbol{z}, s\}$$

with

- $\boldsymbol{T} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_P]$ logistic coefficient matrix
- $\boldsymbol{z} = [z_1, \ldots, z_P]^T$ label vector
- $\boldsymbol{s} = \left[s_1^2, \ldots, s_P^2\right]^T$ noise variance vector

Likelihood

Likelihood

The LMM model and the Gaussian property of the noise vector yield

$$f\left(\boldsymbol{y}_p \,|\boldsymbol{t}_p, s_p^2\right) = \left(\frac{1}{2\pi s_p^2}\right)^{\frac{L}{2}} \exp\left[-\frac{\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_p(\boldsymbol{t}_p)\|^2}{2s_p^2}\right], \tag{3}$$

where

- $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}$ is the standard ℓ^2 norm,
- $\boldsymbol{a}_p(\boldsymbol{t}_p)$ explicitly mention dependance of the abundance vector $\boldsymbol{a}_p$ over the logistic coefficient vector $\boldsymbol{t}_p$.

Assuming independence between the different noise vector $\boldsymbol{n}_p$

$$\Rightarrow \quad f\left(\boldsymbol{Y}|\boldsymbol{T}, \boldsymbol{s}\right) = \prod_{p=1}^{P} f\left(\boldsymbol{y}_p|\boldsymbol{t}_p, s_p^2\right). \tag{4}$$

Label prior

Which prior for the labels?

- Independent prior distributions:

$$f(\boldsymbol{z}) = \prod_{p=1}^{P} f(z_p)$$

$\Rightarrow$ no correlations between the pixels!

Label prior

Which prior for the labels?

- Independent prior distributions:

$$f(\boldsymbol{z}) = \prod_{p=1}^{P} f(z_p)$$

$\Rightarrow$ no correlations between the pixels!

- Markov Random fields...

Potts-Markov random field

$$f(\boldsymbol{z}) = \frac{1}{G(\beta)} \exp\left[\beta \sum_{p=1}^{P} \sum_{p' \in \mathcal{V}(p)} \delta(z_p - z_{p'})\right]$$

- $\mathcal{V}(p)$ the 1-order neighborhood (4 pixels),
- $G(\beta)$ normalizing constant (or *partition function*),
- β: granularity coefficient (known in this study),
- $\delta(x) = \begin{cases} 1, & \text{if } x = 0, \\ 0, & \text{otherwise.} \end{cases}$

☞ Naturally introduces spatial correlations between labels of neighboring pixels

Influence of β on the label prior

Generation of synthetic images with $K = 3$ using Potts-Markov random fields with different values of β and 1-order neighborhood structure.

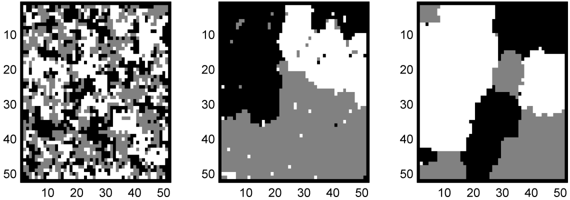


Figure: From left to right: $\beta = 0.8,\ 1.4,\ 2.$

- small values of β: many small regions,
- large values of β: few large regions,

Logistic coefficient prior distribution

Logistic coefficient prior distribution

For each class k ($k = 1, \ldots, K$), $p \in \mathcal{I}_k$

$$\boldsymbol{t}_p | z_p = k, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k \sim \mathcal{N}\left(\boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k\right),$$

$\Rightarrow$ Two hyperparameters (fixed or estimated) that characterize the kth class: $\boldsymbol{\psi}_k$ and $\boldsymbol{\Sigma}_k$.

For the whole set of pixels, the distribution for $\boldsymbol{T} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_P]$ is:

$$f\left(\boldsymbol{T} | \boldsymbol{\Psi}, \boldsymbol{\Sigma}\right) = \prod_{k=1}^{K} \prod_{p \in \mathcal{I}_k} f\left(\boldsymbol{t}_p | z_p = k, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k\right)$$

with $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K]$ and $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K\}$.

Noise variance and hyperparameter prior distributions

Noise variance prior

$$s_p^2 | \delta \sim \mathcal{E}(\delta),$$

where δ is an adjustable hyperparameter.

$$f(\boldsymbol{s}|\delta) = \prod_{p=1}^{P} f\left(s_p^2|\delta\right) \tag{5}$$

Hyperparameter priors (hierarchical model)

Let $\Omega = \{\delta, \boldsymbol{\Psi}, \boldsymbol{\Sigma}\}$ the hyperparameter vector:

$$f(\Omega) = f(\delta) \prod_{k=1}^{K} f(\boldsymbol{\psi}_k) f(\boldsymbol{\Sigma}_k)$$

with

- $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_K]$ (means of the logistic coefficient vectors)
- $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K\}$ (covariance matrices of the logistic coefficient vectors).

Joint posterior distribution of $\mathbf{\Theta}$ and Ω

Joint posterior of parameter $\mathbf{\Theta}$ and hyperparameter Ω vectors:

$$f(\mathbf{\Theta}, \Omega | \boldsymbol{Y}) \propto f(\boldsymbol{Y} | \mathbf{\Theta}) f(\mathbf{\Theta} | \Omega) f(\Omega). \tag{6}$$

Straightforward computations lead to:

$$\begin{aligned} f(\mathbf{\Theta}, \Omega | \boldsymbol{Y}) \propto & \prod_{p=1}^{P} \left(\frac{1}{s_p^2}\right)^{\frac{L}{2}} \exp\left[-\frac{\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_p(\boldsymbol{t}_p)\|^2}{2s_p^2}\right] \\ & \times \prod_{r,k} \frac{1}{\sigma_{r,k}^{n_k+1}} \exp\left[-\left(\frac{\psi_{r,k}^2}{2\upsilon^2} + \frac{2\gamma + \sum_{p\in\mathcal{I}_k}(t_{r,p} - \psi_{r,k})^2}{2\sigma_{r,k}^2}\right)\right] \\ & \times \delta^{P-1} \prod_{p=1}^{P} \left(\frac{1}{w_p^2}\right)^2 \exp\left(-\frac{\delta}{w_p^2}\right) \left(\frac{1}{\upsilon^2}\right)^{\frac{RK}{2}+1} \\ & \times \exp\left[\sum_{p=1}^{P} \sum_{p'\in\mathcal{V}(p)} \beta\delta(z_p - z_{p'})\right] \end{aligned} \tag{7}$$

where $n_k = \mathrm{card}(\mathcal{I}_k)$ (combinatorial problem).

Computing estimates of $\boldsymbol{\Theta}$

Joint posterior distribution too complex to obtain Bayesian estimates of $\boldsymbol{\Theta}$

- Simulation of samples $\boldsymbol{\Theta}^{(t)}$ asymptotically distributed according to $f(\boldsymbol{\Theta}, \Omega|\boldsymbol{Y})$ using MCMC methods
- MMSE and MAP estimators

$$\hat{\boldsymbol{\Theta}}_{\text{MMSE}} \approx \frac{1}{N_{\text{MC}} - N_{\text{bi}}} \sum_{t=1}^{N_{\text{MC}} - N_{\text{bi}}} \boldsymbol{\Theta}^{(t)}, \tag{8}$$

$$\hat{\boldsymbol{\Theta}}_{\text{MAP}} \approx \arg \max_{\boldsymbol{\Theta}^{(t)}} f(\boldsymbol{\Theta}^{(t)}|\boldsymbol{Y}). \tag{9}$$

Outline

Synthetic data

Simulation parameters

- 25×25 synthetic image composed of $K = 3$ classes

Synthetic data

Simulation parameters

- 25×25 synthetic image composed of $K = 3$ classes
- label map generated using a Potts-Markov random field with $\beta = 1.1$

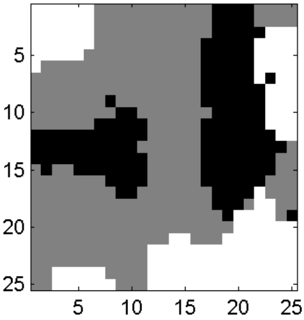


Figure: Synthetic label map

Synthetic data

Simulation parameters

- 25×25 synthetic image composed of $K = 3$ classes
- label map generated using a Potts-Markov random field with $\beta = 1.1$
- Each pixel mixed with $R = 3$ endmembers whose spectra are: construction concrete, green grass, micaceous loam ($L = 413$ bands)

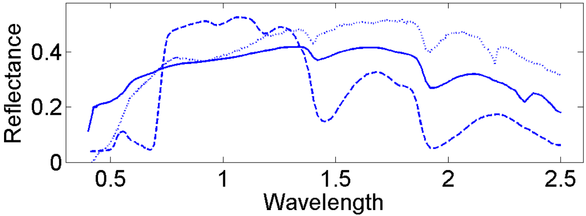


Figure: Endmember spectra

Synthetic data

Simulation parameters

- 25×25 synthetic image composed of $K = 3$ classes
- label map generated using a Potts-Markov random field with $\beta = 1.1$
- Each pixel mixed with $R = 3$ endmembers whose spectra are: construction concrete, green grass, micaceous loam ($L = 413$ bands)
- Noise level SNR $\simeq$ 15dB

Synthetic data

Simulation parameters

- 25×25 synthetic image composed of $K = 3$ classes
- label map generated using a Potts-Markov random field with $\beta = 1.1$
- Each pixel mixed with $R = 3$ endmembers whose spectra are: construction concrete, green grass, micaceous loam ($L = 413$ bands)
- Noise level SNR $\simeq$ 15dB
- Abundances maps generated from truncated Gaussian distributions (with different means and covariance matrices in each class)

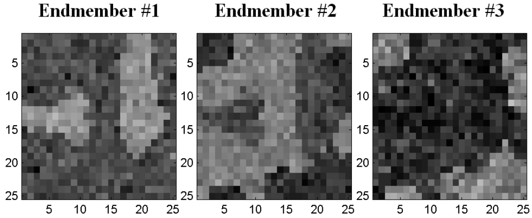


Figure: Synthetic abundances map

Synthetic data

Simulation results: label map

- Marginal MAP label estimates

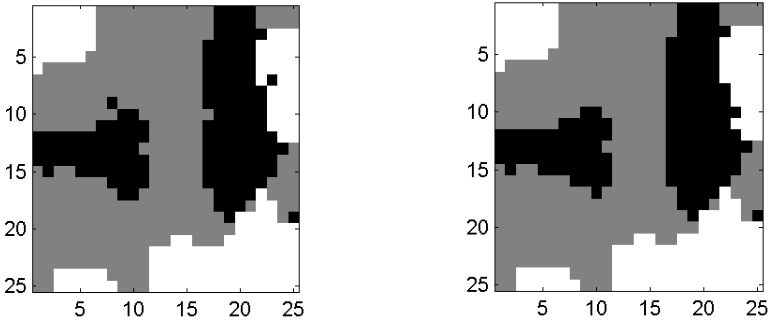


Figure: Left: original labels. Right: estimated labels.

Synthetic data

Simulation results: abundance maps

- MMSE logistic coefficient estimates (conditionally to the MAP label estimates)

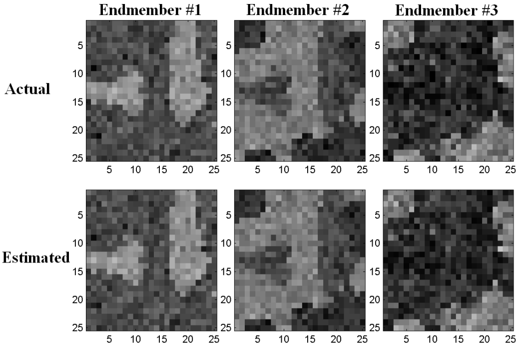


Figure: Top: Actual abundances. Bottom: estimated abundances.

Synthetic data

Simulation results: abundance means

Table: Actual and estimated abundance mean and variance for each class.

		Real values	Estimated values
Class 1	$\boldsymbol{\mu}_1 = \mathrm{E}[\boldsymbol{a}_p \vert \mathcal{I}_1]$	$[0.6, 0.3, 0.1]^T$	$[0.58, 0.29, 0.13]^T$
	$\mathrm{Var}[a_p \vert \mathcal{I}_1]$ $(\times 10^{-3})$	$[5, 5, 5]^T$	$[4.5, 4.3, 5.5]^T$
Class 2	$\boldsymbol{\mu}_2 = \mathrm{E}[\boldsymbol{a}_p \vert \mathcal{I}_2]$	$[0.3, 0.5, 0.2]^T$	$[0.29, 0.49, 0.2]^T$
	$\mathrm{Var}[a_p \vert \mathcal{I}_2]$ $(\times 10^{-3})$	$[5, 5, 5]^T$	$[4.5, 4.7, 5.3]^T$
Class 3	$\boldsymbol{\mu}_3 = \mathrm{E}[\boldsymbol{a}_p \vert \mathcal{I}_3]$	$[0.3, 0.2, 0.5]^T$	$[0.31, 0.19, 0.49]^T$
	$\mathrm{Var}[a_p \vert \mathcal{I}_3]$ $(\times 10^{-3})$	$[5, 5, 5]^T$	$[7, 4.2, 11.7]^T$

Example of results with the 2nd class

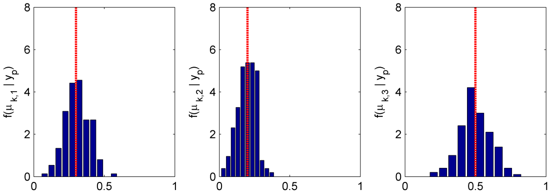
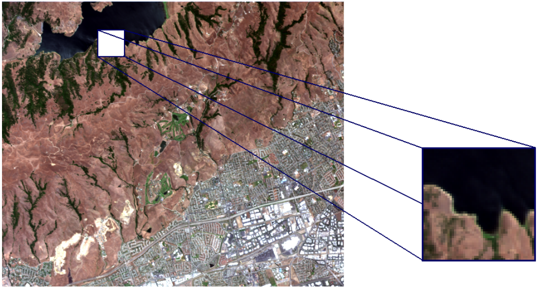


Figure: Histograms of the abundance means $\boldsymbol{\mu}_2$

Real data

Simulation parameters

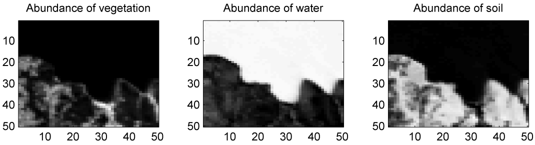
- Real hyperspectral image of 50×50 pixels extracted from a larger image acquired in 1997 by AVIRIS (Moffett Field, CA, USA) data set reduced from 224 to 189 bands (water absorption bands removed)
- N-FINDR used to extract the $R = 3$ endmember spectra associated to water, soil and vegetation.
- $K = 4$ classes have been considered



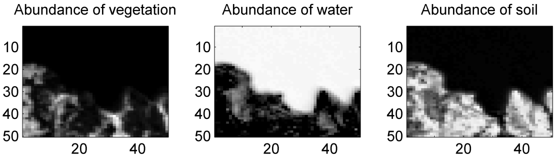
Real hyperspectral data: Moffett field acquired by AVIRIS in 1997 (left) and the region of interest shown in true colors (right).

Real data

Abundance maps



Comparison with other results



Results from FCLS algorithm[a].

[a] [D. C. Heinz, C.-I Chang, *IEEE Trans. on Geosci. and Remote Sensing*, 2001]

Real data

Label map

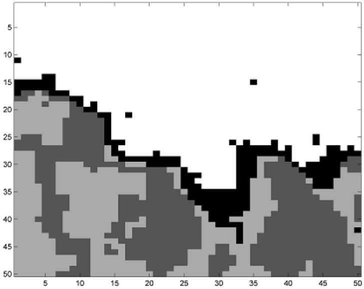


Figure: $\beta = 1.1$.

Outline

Conclusions

“Spatial” unmixing algorithm

- Exploitation of spatial correlations using MRFs within a Bayesian framework
- Hidden labels introduced to identify several classes defined by homogeneous composition of macroscopic materials
- Appropriate reparametrization for the abundance vector

Results

- Good estimation of abundance coefficients
- Classification map obtained thanks to the estimates of the underlying labels
- Price to pay: high computational cost

Perspectives

- Estimation of the granularity coefficient β,
- Other models for spatial correlations (discriminative random fields...)
- Replace MCMC methods by other computational methods: variational methods, (constrained ?) Hamiltonian MCMC,...

Unmixing hyperspectral images using Markov random fields

Olivier Eches, Nicolas Dobigeon and Jean-Yves Tourneret

Université de Toulouse - IRIT/INP-ENSEEIHT Toulouse, FRANCE
http://www.enseeiht.fr/~eches

MaxEnt 2010, Chamonix, France

Simulation results: comparisons

Synthetic data: comparisons with other algorithms

$$\mathrm{MSE}_r^2 = \frac{1}{P}\sum_{p=1}^{P}(\hat{a}_{r,p} - a_{r,p})^2$$

Table: MSEs for abundance coefficients.

	FCLS[a]	Bayesian independent[b]	Spatial
MSE_1^2	0.0019	0.0016	0.001
MSE_2^2	4.3×10^{-4}	4.1×10^{-4}	3.1×10^{-4}
MSE_3^2	0.0014	0.0013	8.6×10^{-4}

[a] [D. C. Heinz, C.-I Chang, *IEEE Trans. on Geosci. and Remote Sensing*, 2001]
[b] [N. Dobigeon *et al.*, *IEEE Trans. on Signal Proc.*, 2008]

Real data

Label map

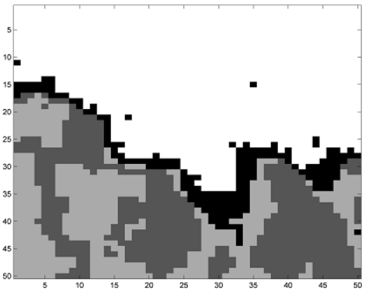


Figure: $\beta = 1.1$.

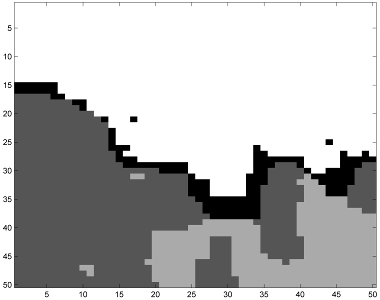


Figure: $\beta = 2$.