

Enhancing hyperspectral image unmixing with spatial correlations

Olivier Eches, Nicolas Dobigeon and Jean-Yves Tourneret

Abstract

This paper describes a new algorithm for hyperspectral image unmixing. Most of the unmixing algorithms proposed in the literature do not take into account the possible spatial correlations between the pixels. In this work, a Bayesian model is introduced to exploit these correlations. The image to be unmixed is assumed to be partitioned into regions (or *classes*) where the statistical properties of the abundance coefficients are homogeneous. A Markov random field is then proposed to model the spatial dependency of the pixels within any class. Conditionally upon a given class, each pixel is modeled by using the classical linear mixing model with additive white Gaussian noise. This strategy is investigated the well known linear mixing model. For this model, the posterior distributions of the unknown parameters and hyperparameters allow ones to infer the parameters of interest. These parameters include the abundances for each pixel, the means and variances of the abundances for each class, as well as a classification map indicating the classes of all pixels in the image. To overcome the complexity of the posterior distribution of interest, we consider Markov chain Monte Carlo methods that generate samples distributed according to the posterior of interest. The generated samples are then used for parameter and hyperparameter estimation. The accuracy of the proposed algorithms is illustrated on synthetic and real data.

Index Terms

Bayesian inference, Monte Carlo methods, spectral unmixing, hyperspectral images, Markov random fields, Potts-Markov model.

I. INTRODUCTION

Since the early 90's, hyperspectral imagery has been receiving growing interests in various fields of applications. For example, hyperspectral images have been recently used successfully for mapping the timber species in tropical forestry [1]. Hyperspectral image analysis involves many technical issues such as image classification, image segmentation, target detection and the

The authors are with University of Toulouse, IRIT/INP-ENSEEIHT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France (e-mail: {olivier.eches, nicolas.dobigeon, jean-yves.tourneret}@enseeiht.fr).

crucial step of spectral unmixing. The problem of spectral unmixing has been investigated for several decades in both the signal processing and geoscience communities where many solutions have been proposed (see for instance [2] and [3] and references therein). Hyperspectral unmixing consists of decomposing the measured pixel reflectances into mixtures of pure spectra whose fractions are referred to as abundances. Assuming the image pixels are linear combinations of pure materials is very common in the unmixing framework. More precisely, the linear mixing model (LMM) considers the spectrum of a mixed pixel as a linear combination of endmembers [2]. Another statistical model known as normal compositional model (NCM) proposed by Stein in [4] assumes that the measured pixel reflectances are combinations of random endmembers (with known means) instead of deterministic ones. As illustrated in [5], the NCM can be preferred to the LMM when the image does not contain enough pure pixels. However, this article will only focus on the LMM.

The LMM requires to have known endmember signatures. These signatures can be obtained from a spectral library or by using an endmember extraction algorithm (EEA). Some standard EEAAs are reviewed in [6]. Once the endmembers that appear in a given image have been identified, the corresponding abundances have to be estimated in a so-called *inversion* step. Due to obvious physical considerations, the abundances have to satisfy positivity and sum-to-one constraints. A lot of inversion algorithms respecting these constraints have been proposed in the literature. The fully constrained least squares (FCLS) [7] and scaled gradient (SGA) [8] algorithms are two optimization techniques that ensure the positivity and sum-to-one constraints inherent to the unmixing problem. Another interesting approach introduced in [9] consists of assigning appropriate prior distributions to the abundances and to solve the unmixing problem within a Bayesian framework. However, all these inversion strategies have been developed in a pixel-by-pixel context and, consequently, do not exploit the possible spatial correlations between the different pixels of the hyperspectral image. In this paper, we show that taking these spatial correlations into account allows one to improve the unmixing procedure. More precisely, the Bayesian algorithm initially developed in [9] is modified to introduce spatial constraints between the abundance coefficients to be estimated.

Within a Bayesian estimation framework, a very popular strategy for modeling spatial information in an image is based on Markov random fields (MRFs). MRFs have been widely used in the image processing literature to properly describe neighborhood dependance between image

pixels. MRFs and their pseudo-likelihood approximations have been introduced by Besag in [10]. MRFs have then been popularized by Geman in [11] by exploiting the Gibbs distribution inherent to MRFs. There are mainly two approaches that can be investigated to model spatial correlations between the abundances of an hyperspectral image with MRFs. The first idea is to define appropriate prior distributions for the abundances highlighting spatial correlation. This approach has been for instance adopted by Kent and Mardia in [12] where several techniques have been introduced for mixed-pixel classification of remote sensing data. These techniques rely on a fuzzy membership process, which implicitly casts the achieved classification task as a standard unmixing problem¹. Modeling the abundance dependencies with MRFs makes this approach particularly well adapted to unmix images with smooth abundance transition throughout the scene. Conversely, this paper proposes to exploit the pixel correlations in an underlying membership model. This standard alternative strategy allows more flexibility and appears more suited for images composed of distinct areas, as frequently encountered in remote sensing applications. Moreover, this approach has the great advantage of easily generalizing the Bayesian algorithms previously introduced in [9], [13], as detailed further in the manuscript. It consists of introducing labels that are assigned to the pixels of the image. Then MRFs are not assigned on the abundances directly but on these hidden variables, leading to a softer classification. More precisely, to take into account the possible spatial correlations between the observed pixels, a Potts-Markov field [14] is chosen as prior for the labels. This distribution enforces the neighboring pixels to belong to the same class. Potts-Markov models have been extensively used for classification/segmentation of hyperspectral data in the remote sensing and image processing literatures [15]–[20]. Note that other research works, such as [21] and [22], have proposed other strategies of modeling spatial correlations between pixels for classification of hyperspectral images. All these works have shown that taking into account the spatial correlations is of real interest when analyzing hyperspectral images.

This paper proposes to study the interest of using MRFs for unmixing hyperspectral images. More precisely, the Bayesian unmixing strategy developed in [9] is generalized to take into account spatial correlations between the pixels of an hyperspectral image. The hyperspectral

¹Note that, to our knowledge, the Kent and Mardia's paper is one of the earliest work explicitly dealing with linear unmixing of remotely sensed images.

image to be analyzed is assumed to be partitioned into homogeneous regions (or classes) in which the abundance vectors have the same first and second order statistical moments (means and covariances). This assumption implies an implicit image classification, modeled by hidden labels whose spatial dependencies follow a Potts-Markov field. Conditionally upon these labels, the abundance vectors are assigned appropriate prior distributions with unknown means and variances that depend on the pixel class. These prior distributions ensure the positivity and sum-to-one constraints of the abundance coefficients. They are based on a reparametrization of the abundance vectors and are much more flexible than the priors previously studied in [9], [5] or [13]. Of course, the accuracy of the abundance estimation procedure drastically depends on the hyperparameters associated to these priors. This paper proposes to estimate these hyperparameters in a fully unsupervised manner by introducing a second level of hierarchy in the Bayesian inference. Non-informative prior distributions are assigned to the hyperparameters. The unknown parameters (labels and abundance vectors) and hyperparameters (prior abundance mean and variance for each class) are then inferred from their joint posterior distribution. Since this posterior is too complex to derive closed-form expressions for the classical Bayesian estimators, Markov chain Monte Carlo (MCMC) techniques are studied to alleviate the numerical problems related to the LMM with spatial constraints. MCMC generate samples asymptotically distributed according to the joint posterior of interest. These samples are finally used to approximate the Bayesian estimators, such as the minimum mean square error (MMSE) or the maximum *a posteriori* estimators. Note that the underlying classification and abundance estimation problems are jointly solved within this Bayesian framework.

The paper is organized as follows. The unmixing problem associated to the LMM with spatial correlations is formulated in II. Section III introduces a hierarchical Bayesian model appropriate to this unmixing problem. The MCMC algorithm required to approximate the Bayesian LMM estimators is described in Section IV. Simulation results conducted on simulated and real data are provided in Sections V and VI. Finally, the conclusions of this work are given in Section VII.

II. TECHNICAL BACKGROUND AND PROBLEM FORMULATION

A. Unmixing statistical model

As highlighted in the previous section, the LMM has been mainly proposed in the remote sensing literature for spectral unmixing. The LMM assumes that the spectrum of a given pixel is

a linear combination of deterministic endmembers corrupted by an additive noise [2] considered here as white Gaussian. More specifically, the observed L -spectrum of a given pixel p is defined as

$$\mathbf{y}_p = \mathbf{M}\mathbf{a}_p + \mathbf{n}_p \quad (1)$$

where L is the number of spectral bands, $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_R]$ is a known $L \times R$ matrix containing the L -spectra of the endmembers, \mathbf{a}_p is the $R \times 1$ abundance vector, R is the number of endmembers that are present in the image and \mathbf{n}_p is the noise vector. The vector \mathbf{n}_p is classically assumed to be an independent and identically distributed (i.i.d.) zero-mean Gaussian sequence with unknown variance s^2

$$\mathbf{n}_p | s^2 \sim \mathcal{N}(\mathbf{0}_L, s^2 \mathbf{I}_L) \quad (2)$$

where \mathbf{I}_L is the $L \times L$ identity matrix. Note that the noise is the same for all pixels of the hyperspectral image and does not vary from one pixel to another, which has been a common assumption widely admitted in the hyperspectral literature [23]–[25].

Considering an image of P pixels, standard matrix notations can be adopted leading to $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_P]$ and $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_P]$.

B. Introducing spatial dependencies between abundances

We propose in this paper to exploit some spatial correlations between the pixels of the hyperspectral image to be analyzed. More precisely, when performing unmixing, it is interesting to consider that the abundances of a given pixel are similar to the abundances of its neighboring pixels. Formally, the hyperspectral image is assumed to be partitioned into K regions or *classes*. Let $\mathcal{I}_k \subset \{1, \dots, P\}$ denote the subset of pixel indexes belonging to the k th class. A label vector of size $P \times 1$ denoted as $\mathbf{z} = [z_1, \dots, z_P]^T$ with $z_p \in \{1, \dots, K\}$ is introduced to identify the class to which each pixel p belongs ($p = 1, \dots, P$). In other terms

$$z_p = k \Leftrightarrow p \in \mathcal{I}_k. \quad (3)$$

In each class, the abundance vectors to be estimated are assumed to share the same first and second order statistical moments, i.e., $\forall k \in \{1, \dots, K\}, \forall (p, p') \in \mathcal{I}_k \times \mathcal{I}_k$

$$\begin{aligned} \mathbb{E}[\mathbf{a}_p] &= \mathbb{E}[\mathbf{a}_{p'}] = \boldsymbol{\mu}_k \\ \mathbb{E}[(\mathbf{a}_p - \boldsymbol{\mu}_k)(\mathbf{a}_p - \boldsymbol{\mu}_k)^T] &= \mathbb{E}[(\mathbf{a}_{p'} - \boldsymbol{\mu}_k)(\mathbf{a}_{p'} - \boldsymbol{\mu}_k)^T]. \end{aligned} \quad (4)$$

Therefore, the k th class of the hyperspectral image to be unmixed is fully characterized by the abundance mean vector and the abundance covariance matrix of the pixels belonging to this class.

C. Markov random fields

To describe spatial constraints between pixels, it is important to properly define a neighborhood structure. The neighborhood relation between two pixels i and j , has to be symmetric: if i is a neighbor of j then j is a neighbor of i . This relation is applied to the nearest neighbors of the considered pixel, for example the fourth, eighth or twelfth nearest pixels. Fig. 1 shows two examples of neighborhood structures. The four pixel structure or *1-order neighborhood* will be considered in the rest of the paper. Therefore, the associated set of neighbors, or *cliques*, has only vertical and horizontal possible configurations (see [10], [11] for more details).

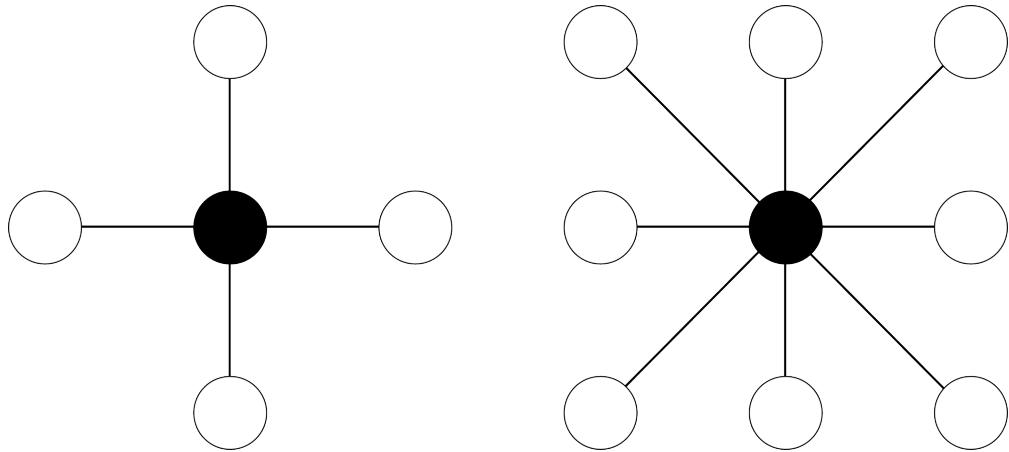


Fig. 1. 4-pixel (left) and 8-pixel (right) neighborhood structures. The considered pixel appear as a black circle whereas its neighbors are depicted in white.

Once the neighborhood structure has been clearly established, the MRF can be easily defined. Let z_p denote a random variable associated to the p th pixel of an image of P pixels. In our context of hyperspectral image unmixing, these variables will refer to the class to which the pixels belong and take their values in a finite set, e.g., $\{1, \dots, K\}$. The whole set of random variables $\{z_1, \dots, z_P\}$ forms a random field. An MRF is then defined when the conditional distribution of z_i given the other pixels z_{-i} only depend on its neighbors $z_{\mathcal{V}(i)}$, i.e.,

$$f(z_i | z_{-i}) = f(z_i | z_{\mathcal{V}(i)}) \quad (5)$$

where $\mathcal{V}(i)$ is the neighborhood structure considered and $\mathbf{z}_{-i} = \{z_j; j \neq i\}$.

Since the pioneer work of Geman [11], MRFs have been widely used in the image processing community as in [26], [27]. The hyperspectral community has also recently exploited the advantages of MRFs for hyperspectral image analysis [16], [19], [28]. However, to our knowledge, MRFs have not been studied for hyperspectral image unmixing. MRFs provide an efficient way of modeling correlations between pixels, which is adapted to the intrinsic properties of most images. Two specific MRFs are appropriate for image analysis: the Ising model for binary random variables and the Potts-Markov model that is a simple generalization to more-than-two variables [14]. This paper focuses on the Potts-Markov model since it is very appropriate to hyperspectral image segmentation [19]. Given a discrete random field \mathbf{z} attached to an image with P pixels, the Hammersley-Clifford theorem yields

$$f(\mathbf{z}) = \frac{1}{G(\beta)} \exp \left[\sum_{p=1}^P \sum_{p' \in \mathcal{V}(p)} \beta \delta(z_p - z_{p'}) \right] \quad (6)$$

where β is the *granularity* coefficient, $G(\beta)$ is the normalizing constant or *partition function* [29] and $\delta(\cdot)$ is the Kronecker function

$$\delta(x) = \begin{cases} 1, & \text{if } x = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Note that drawing a label vector $\mathbf{z} = [z_1, \dots, z_P]$ from the distribution (6) can be easily achieved without knowing $G(\beta)$ by using a Gibbs sampler (the corresponding algorithmic scheme is summarized in Algo. 1). However, a major difficulty with the distribution (6) comes from the partition function that has no closed-form expression and depends on the unknown hyperparameter β . The hyperparameter β tunes the degree of homogeneity of each region in the image. Some simulations have been conducted to show the influence of this parameter on image homogeneity. Synthetic images have been generated from a Potts-Markov model with $K = 3$ (corresponding to three gray levels in the image) and a 1-order neighborhood structure. Fig. 2 indicates that a small value of β induces a *noisy* image with a large number of regions, contrary to a large value of β that leads to few and large homogeneous regions. Note that it is unnecessary to consider values of $\beta \geq 2$ since for the 1-order neighborhood structure adopted here, “*When $\beta \geq 2$, the Potts-Markov model is almost surely concentrated on single-color images*” [30, p. 237]. Note however that for larger neighborhood systems, a smaller value of β would be enough

to obtain uniform patches in Potts realizations since, for example, β is expected to be about twice for an 2-order neighborhood structure [31]. In this work, the granularity coefficient β will be fixed *a priori*. However, it is interesting to mention that the estimation of β might also be conducted by using the methods studied in [32], [33] and [34].

Algorithm 1 MRF Simulation

```

1: for  $n = 1$  to  $N_{\text{MC}}$  do
2:   for  $p = 1$  to  $P$  do
3:     for  $k = 1$  to  $K$  do
4:       Compute  $w_k \propto P[z_p = k | z_{-p}]$  according to (6)
5:     end for
6:   Compute the normalizing constant

```

$$G(\beta) = \sum_{k=1}^K w_k$$

```

7:   Set the probability vector

```

$$\tilde{\mathbf{w}} = \left[\frac{w_1}{G(\beta)}, \dots, \frac{w_K}{G(\beta)} \right]$$

```

8:   Draw  $z_p$  in  $\{1, \dots, K\}$  with proba.  $\{\tilde{w}_1, \dots, \tilde{w}_K\}$ .

```

```

9: end for

```

```

10: end for

```

D. Abundance Reparametrization

As explained before, the fraction vectors \mathbf{a}_p should satisfy positivity and sum-to-one constraints defined as

$$\begin{cases} a_r > 0, \forall r = 1, \dots, R, \\ \sum_{r=1}^R a_r = 1. \end{cases} \quad (7)$$

To ensure that these abundance constraints are satisfied, we have considered a reparametrization for positive parameters summing to one that was introduced in [12] for the spectral unmixing of satellite images. Note that this reparametrization has also shown interesting results for a pharmacokinetic problem [35] and has been recently applied to hyperspectral unmixing [36]. It consists of rewriting the abundances as a function of random variables that will be referred to

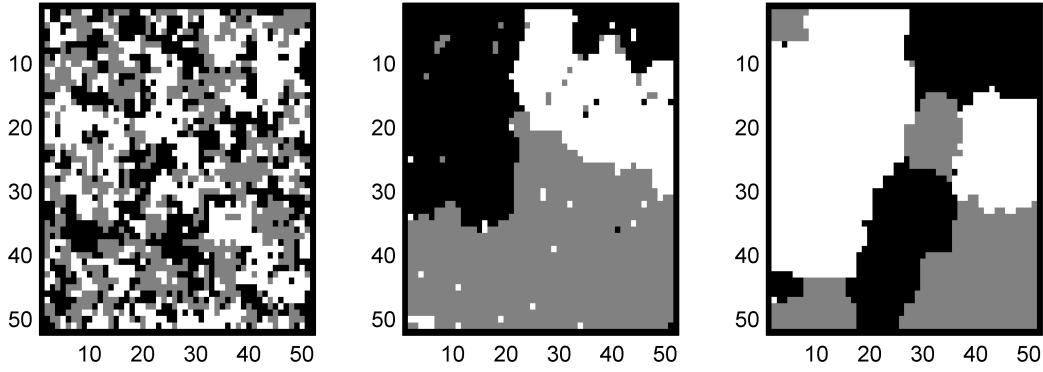


Fig. 2. Synthetic images generated from a Potts-Markov model with (from left to right) $\beta = 0.8, 1.4, 2$.

as *logistic coefficients* in the rest of the paper. A logistic coefficient vector $\mathbf{t}_p = [t_{1,p} \dots, t_{R,p}]^T$ is assigned to each abundance vector \mathbf{a}_p , according to the relationship

$$a_{r,p} = \frac{\exp(t_{r,p})}{\sum_{r=1}^R \exp(t_{r,p})}. \quad (8)$$

Initially, the spatial dependencies resulting from the image partitioning described in II-B are based on the first and second order moments of the abundance vectors \mathbf{a}_p . However, the spatial constraints defined in (4) can be easily adapted when using logistic coefficient vectors. Indeed, in each class, the unknown logistic coefficient vectors are assumed to share the same first and second order moments, i.e., $\forall k \in \{1, \dots, K\}, \forall (p, p') \in \mathcal{I}_k \times \mathcal{I}_k$

$$\begin{aligned} \boldsymbol{\psi}_k &= E[\mathbf{t}_p | z_p = k] = E[\mathbf{t}_{p'} | z_{p'} = k] \\ \boldsymbol{\Sigma}_k &= E[(\mathbf{t}_p - \boldsymbol{\psi}_k)(\mathbf{t}_p - \boldsymbol{\psi}_k)^T | z_p = k] \\ &= E[(\mathbf{t}_{p'} - \boldsymbol{\psi}_k)(\mathbf{t}_{p'} - \boldsymbol{\psi}_k)^T | z_{p'} = k]. \end{aligned} \quad (9)$$

With this reparametrization, the k th class is fully characterized by the unknown hyperparameters $\boldsymbol{\psi}_k$ and $\boldsymbol{\Sigma}_k$.

III. HIERARCHICAL BAYESIAN MODEL

This section investigates the likelihood and the priors inherent to the LMM for the spectral unmixing of hyperspectral images, based on Potts-Markov random fields and logistic coefficients.

A. Unknown parameters

The unknown parameter vector associated to the LMM unmixing strategy is denoted as

$$\Theta = \{\mathbf{T}, \mathbf{z}, s^2\}$$

where s^2 is the noise variance, \mathbf{z} is the label vector and $\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_P]$ with $\mathbf{t}_p = [\mathbf{t}_{1,p}, \dots, \mathbf{t}_{R,p}]^T$ ($p = 1, \dots, P$) is the logistic coefficient matrix used for the abundance reparametrization. Note that the noise variance s^2 has been assumed to be unknown in the present paper, contrary to the model considered in [12].

B. Likelihood

The additive white Gaussian noise sequence of the LMM allows one to write² $\mathbf{y}_p | \mathbf{t}_p, s^2 \sim \mathcal{N}(\mathbf{M}\mathbf{a}_p(\mathbf{t}_p), s^2 \mathbf{I}_L)$ ($p = 1, \dots, P$). Therefore the likelihood function of \mathbf{y}_p is

$$f(\mathbf{y}_p | \mathbf{t}_p, s^2) \propto \frac{1}{s^L} \exp \left[-\frac{\|\mathbf{y}_p - \mathbf{M}\mathbf{a}_p(\mathbf{t}_p)\|^2}{2s^2} \right] \quad (10)$$

where \propto means proportional to and $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$ is the standard ℓ^2 norm. By assuming independence between the noise sequences \mathbf{n}_p ($p = 1, \dots, P$), the likelihood of the P image pixels is

$$f(\mathbf{Y} | \mathbf{T}, s^2) = \prod_{p=1}^P f(\mathbf{y}_p | \mathbf{t}_p, s^2). \quad (11)$$

C. Parameter priors

This section introduces the prior distributions of the unknown parameters and their associated hyperparameters that will be used for the LMM. The directed acyclic graph (DAG) for the parameter priors and hyperpriors for the considered model is represented in Fig. 3.

1) *Label prior:* The prior distribution for the label vector $\mathbf{z} = [z_1, \dots, z_P]^T$ introduced in paragraph II-C is a Potts-Markov random field with a 1-order neighborhood and a known granularity coefficient β (fixed *a priori*). The resulting prior distribution can be written

$$f(\mathbf{z}) \propto \exp \left[\sum_{p=1}^P \sum_{p' \in \mathcal{V}(p)} \beta \delta(z_p - z_{p'}) \right] \quad (12)$$

where $\mathcal{V}(p)$ is the 1-order neighborhood depicted in Fig. 1 and \propto means “proportional to.”

²Note that the dependence of the abundance vector \mathbf{a}_p on the logistic coefficient vector \mathbf{t}_p through (8) has been explicitly mentioned by denoting $\mathbf{a}_p = \mathbf{a}_p(\mathbf{t}_p)$.

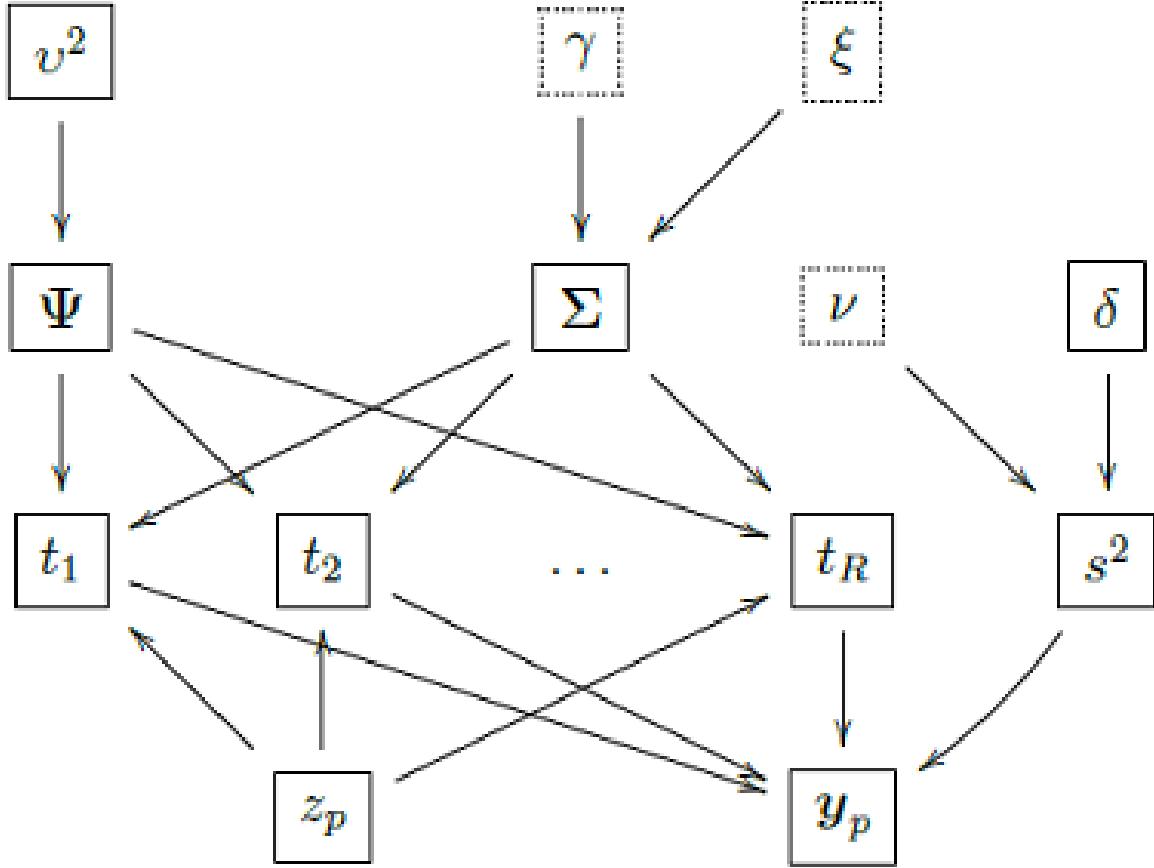


Fig. 3. DAG for the parameter priors and hyperpriors (the fixed parameters appear in dashed boxes) for the LMM.

2) *Logistic coefficient prior:* Following the approach described in paragraph II-B, each component of t_p is assumed to be distributed according to a Gaussian distribution. In addition, as highlighted in II-D (see (9)), the mean and variance of the logistic coefficients depend on the class to which the corresponding pixel belong. Therefore, the prior distribution for the t_p is explicitly defined conditionally upon the pixel label

$$t_{r,p}|z_p = k, \psi_{r,k}, \sigma_{r,k}^2 \sim \mathcal{N}(\psi_{r,k}, \sigma_{r,k}^2) \quad (13)$$

where the hyperparameters $\psi_{r,k}$ and $\sigma_{r,k}^2$ depend on the associated pixel class k . As suggested in Section I, a hierarchical Bayesian algorithm will be used to estimate these hyperparameters. For a given pixel p , by assuming prior independence between the coefficients $t_{1,p}, \dots, t_{R,p}$, the prior distribution for the vector $\mathbf{t} = [t_{1,p}, \dots, t_{R,p}]^T$ is

$$f(\mathbf{t}_p|z_p = k, \psi_k, \Sigma_k) \sim \mathcal{N}(\psi_k, \Sigma_k) \quad (14)$$

where $\boldsymbol{\psi}_k = [\psi_{1,k}, \dots, \psi_{R,k}]^T$ and $\boldsymbol{\Sigma}_k = \text{diag}(\sigma_{r,k}^2)$ is the $R \times R$ diagonal matrix whose diagonal elements are $\sigma_{r,k}^2$.

By assuming prior independence between the P vectors $\mathbf{t}_1, \dots, \mathbf{t}_P$, the full posterior distribution for the logistic coefficient matrix \mathbf{T} is

$$f(\mathbf{T}|\mathbf{z}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) = \prod_{k=1}^K \prod_{p \in \mathcal{I}_k} f(\mathbf{t}_p | z_p = k, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k) \quad (15)$$

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

3) *Noise variance prior:* A conjugate inverse-gamma distribution is assigned to the noise variance

$$s^2 | \nu, \delta \sim \mathcal{IG}(\nu, \delta) \quad (16)$$

where ν and δ are adjustable hyperparameters. This paper assumes $\nu = 1$ (as in [37] or [38]) and estimates δ jointly with the other unknown parameters and hyperparameters (using a hierarchical Bayesian algorithm).

D. Hyperparameter priors

Hierarchical Bayesian algorithms require to define prior distributions for the hyperparameters. A particular attention has been devoted to the hyperparameters $\psi_{r,k}$ and $\sigma_{r,k}^2$ as they fully describe the different classes partitioning the image. The prior distributions for $\psi_{r,k}$ and $\sigma_{r,k}^2$ are conjugate distributions. More precisely, a vague inverse-gamma distribution is chosen for the logistic coefficient variance $\sigma_{r,k}^2$, i.e.,

$$\sigma_{r,k}^2 | \xi, \gamma \sim \mathcal{IG}(\xi, \gamma) \quad (17)$$

where ξ and γ have been tuned to $\xi = 1$ and $\gamma = 5$ (in order to obtain a large variance). Moreover, a centered Gaussian distribution with unknown variance has been chosen as prior for the logistic coefficient mean

$$\psi_{r,k} | v^2 \sim \mathcal{N}(0, v^2) \quad (18)$$

where v^2 is another adjustable hyperparameter. By assuming independence between the different mean vectors $\boldsymbol{\psi}_k$, ($k = 1, \dots, K$), as well as between the covariance matrices $\boldsymbol{\Sigma}_k$, ($k = 1, \dots, K$), the full priors for the two hyperparameters $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ can be expressed as

$$f(\boldsymbol{\Psi}|v^2) \propto \prod_{k=1}^K \prod_{r=1}^R \left(\frac{1}{v^2} \right)^{\frac{1}{2}} \exp \left(-\frac{\psi_{r,k}^2}{2v^2} \right) \quad (19)$$

$$f(\Sigma|\xi, \gamma) \propto \prod_{k=1}^K \prod_{r=1}^R \frac{\gamma^\xi}{\Gamma(\xi)} (\sigma_{r,k}^2)^{-(\xi+1)} \exp\left(-\frac{\gamma}{\sigma_{r,k}^2}\right). \quad (20)$$

Jeffreys' priors are chosen for the hyperparameters δ and v^2 (see, e.g., [39, p. 131] for details including computations)

$$f(\delta) \propto \frac{1}{\delta} \mathbf{1}_{\mathbb{R}^+}(\delta), \quad f(v^2) \propto \frac{1}{v^2} \mathbf{1}_{\mathbb{R}^+}(v^2). \quad (21)$$

where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ denotes the indicator function defined on \mathbb{R}^+ . These choices, also adopted in [37], [40], reflect the lack of knowledge regarding these two hyperparameters. At this last hierarchy level within the Bayesian inference, the hyperparameter vector can be defined as $\Omega = \{\Psi, \Sigma, v^2, \delta\}$.

E. Joint distribution

The joint posterior distribution of the unknown parameters and hyperparameters is classically defined using the hierarchical structure

$$f(\Theta, \Omega | \mathbf{Y}) = f(\mathbf{Y} | \Theta) f(\Theta | \Omega) f(\Omega). \quad (22)$$

Straightforward computations yield the following posterior

$$\begin{aligned} f(\Theta, \Omega | \mathbf{Y}) &\propto \left(\frac{1}{s^2}\right)^{\frac{LP}{2}} \prod_{p=1}^P \exp\left[-\frac{\|\mathbf{y}_p - \mathbf{M}\mathbf{a}_p(\mathbf{t}_p)\|^2}{2s^2}\right] \\ &\times \exp\left[\sum_{p=1}^P \sum_{p' \in \mathcal{V}(p)} \beta\delta(z_p - z_{p'})\right] \\ &\times \frac{\delta^{\nu-1}}{(s^2)^{\nu+1}} \exp\left(-\frac{\delta}{s^2}\right) \prod_{p=1}^P \left(\frac{1}{v^2}\right)^{\frac{RK}{2}+1} \\ &\times \prod_{r,k} \frac{1}{\sigma_{r,k}^{n_k+1}} \exp\left[-\left(\frac{\psi_{r,k}^2}{2v^2} + \frac{2\gamma + \sum_{p \in \mathcal{I}_k} (t_{r,p} - \psi_{r,k})^2}{2\sigma_{r,k}^2}\right)\right] \end{aligned} \quad (23)$$

with $n_k = \text{card}(\mathcal{I}_k)$. The posterior distribution (23) associated to the LMM is too complex to obtain closed-form expressions for the MMSE or MAP estimators of the unknown parameter vector Θ . To alleviate this problem, we propose to use MCMC methods to generate samples that are asymptotically distributed according to (23). The generated samples are then used to approximate the Bayesian estimators. The next section studies a hybrid Gibbs sampler that generates samples asymptotically distributed according to the posterior distribution (23).

IV. HYBRID GIBBS SAMPLERS

This section studies Metropolis-within-Gibbs samplers that generate samples according to the joint posterior $f(\Theta, \Omega | \mathbf{Y})$. The algorithm is summarized in Algo. 2. The Gibbs sampler iteratively generates samples distributed according to the conditional distributions detailed below.

A. Conditional distribution of the label vector \mathbf{z}

For each pixel p ($p = 1, \dots, P$), the class label z_p is a discrete random variable whose conditional distribution is fully characterized by the probabilities

$$P[z_p = k | \mathbf{z}_{-p}, \mathbf{t}_p, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k] \propto f(\mathbf{t}_p | z_p = k, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k) f(z_p | \mathbf{z}_{-p}) \quad (24)$$

where $k = 1, \dots, K$ (K is the number of classes) and \mathbf{z}_{-p} denotes the vector \mathbf{z} whose p th element has been removed. These posterior probabilities can be expressed as

$$\begin{aligned} & P[z_p = k | \mathbf{z}_{-p}, \mathbf{t}_p, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k] \\ & \propto \exp \left[\sum_{p=1}^P \sum_{p' \in \mathcal{V}(p)} \beta \delta(z_p - z_{p'}) \right] \\ & \quad \times |\boldsymbol{\Sigma}_k|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{t}_p - \boldsymbol{\psi}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{t}_p - \boldsymbol{\psi}_k) \right] \end{aligned} \quad (25)$$

where $|\boldsymbol{\Sigma}_k| = \prod_{r=1}^R \sigma_{r,k}^2$. Note that the posterior probabilities of the label vector \mathbf{z} in (25) define an MRF. Consequently, sampling from this conditional distribution can be achieved using the scheme detailed in Algo. 1, i.e., by drawing a discrete value in the finite set $\{1, \dots, K\}$ with the probabilities (25).

B. Conditional distribution of logistic coefficient matrix \mathbf{T}

For each pixel p , the Bayes theorem yields

$$f(\mathbf{t}_p | z_p = k, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k, \mathbf{y}_p) \propto f(\mathbf{y}_p | \mathbf{t}_p, s^2) f(\mathbf{t}_p | z_p = k, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k).$$

Straightforward computations lead to

$$\begin{aligned} & f(\mathbf{t}_p | z_p = k, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k, \mathbf{y}_p, s^2) \\ & \propto \left(\frac{1}{s^2} \right)^{\frac{L}{2}} \exp \left\{ -\frac{1}{2s^2} \|\mathbf{y}_p - \mathbf{M}\mathbf{a}_p(\mathbf{t}_p)\|^2 \right\} \\ & \quad \times |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\mathbf{t}_p - \boldsymbol{\psi}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{t}_p - \boldsymbol{\psi}_k) \right]. \end{aligned} \quad (26)$$

Unfortunately, it is too difficult to generate samples distributed according to (26). Therefore, a Metropolis-Hastings step is used, based on a random walk method [41, p. 245] with a Gaussian distribution $\mathcal{N}(0, u_r^2)$ as proposal distribution. The variance u_r^2 of the instrumental distribution has been fixed to obtain an acceptance rate between 0.15 and 0.5 as recommended in [42].

C. Conditional distributions of the noise variance

The Bayes theorem yields

$$f(s^2 | \mathbf{Y}, \mathbf{T}, \delta) \propto f(s^2 | \delta) \prod_{p=1}^P f(\mathbf{y}_p | \mathbf{t}_p, s^2).$$

As a consequence, $s^2 | \mathbf{Y}, \mathbf{T}, \delta$ is distributed according to the inverse-Gamma distribution

$$s^2 | \mathbf{Y}, \mathbf{T}, \delta \sim \mathcal{IG}\left(\frac{LP}{2} + 1, \delta + \sum_{p=1}^P \frac{\|\mathbf{y}_p - M\mathbf{a}_p(\mathbf{t}_p)\|^2}{2}\right). \quad (27)$$

D. Conditional distribution of Ψ and Σ

For each endmember r ($r = 1, \dots, R$) and each class k ($k = 1, \dots, K$), thanks to Bayes relation the conditional distribution of $\psi_{r,k}$ can be written as

$$f(\psi_{r,k} | \mathbf{z}, \mathbf{t}_r, \sigma_{r,k}^2, v^2) \propto f(\psi_{r,k} | v^2) \prod_{p \in \mathcal{I}_k} f(t_{r,p} | z_p = k, \psi_{r,k}, \sigma_{r,k}^2). \quad (28)$$

Similarly, the conditional distribution of $\sigma_{r,k}^2$ is

$$f(\sigma_{r,k}^2 | \mathbf{z}, \mathbf{t}_r, \psi_{r,k}) \propto f(\sigma_{r,k}^2) \prod_{p \in \mathcal{I}_k} f(t_{r,p} | z_p = k, \psi_{r,k}, \sigma_{r,k}^2). \quad (29)$$

Straightforward computations allow one to obtain the following results

$$\psi_{r,k} | \mathbf{z}, \mathbf{t}_r, \sigma_{r,k}^2, v^2 \sim \mathcal{N}\left(\frac{v^2 n_k \bar{t}_{r,k}}{\sigma_{r,k}^2 + v^2 n_k}, \frac{v^2 \sigma_{r,k}^2}{\sigma_{r,k}^2 + v^2 n_k}\right) \quad (30)$$

$$\sigma_{r,k}^2 | \mathbf{z}, \mathbf{t}_r, \psi_{r,k} \sim \mathcal{IG}\left(\frac{n_k}{2} + 1, \gamma + \sum_{p \in \mathcal{I}_k} \frac{(t_{r,p} - \psi_{r,k})^2}{2}\right) \quad (31)$$

with $\bar{t}_{r,k} = \frac{1}{n_k} \sum_{p \in \mathcal{I}_k} t_{r,p}$

Algorithm 2 Hybrid Gibbs sampler for hyperspectral unmixing using spatial constraints for the LMM

1: % Initialization:

- 2: Sample $\mathbf{z}^{(0)}$ from probability density function (pdf) in Eq. (6),
- 3: Sample $\delta^{(0)}$ and $v^{2(0)}$ from the pdfs in Eq. (21),
- 4: Sample $s^{2(0)}$ from the pdf in Eq. (16),
- 5: Sample $\Psi^{(0)}$ from the pdf in Eq. (19),
- 6: Sample $\Sigma^{(0)}$ from the pdf in Eq. (20),

7: % Iterations:

- 8: **for** $t = 1, 2, \dots$ **do**
- 9: **for** each pixel $p = 1, \dots, P$ **do**
- 10: Sample $\mathbf{z}^{(t)}$ from the pdf in Eq. (25),
- 11: Sample $\mathbf{t}_p^{(t)}$ from the pdf in Eq. (26) using Metropolis-within-Gibbs step,
- 12: Set $a_p^{(t)}$ from Eq. (8),
- 13: **end for**
- 14: Sample $s^{2(t)}$ from the pdf in Eq. (27),
- 15: **for** each class $k = 1, \dots, K$ **do**
- 16: **for** each endmember $r = 1, \dots, R$ **do**
- 17: Sample $\psi_{r,k}^{(t)}$ from the pdf in Eq. (30),
- 18: Sample $\sigma_{r,k}^{2(t)}$ from the pdf in Eq. (31),
- 19: **end for**
- 20: **end for**
- 21: Sample $v^{2(t)}$ from the pdf in Eq. (32)
- 22: Sample $\delta^{(t)}$ from the pdf in Eq. (32),
- 23: **end for**

E. Conditional distribution of v^2 and δ

The conditional distributions of v^2 and δ are the following inverse-gamma and gamma distributions, respectively

$$v^2 | \Psi \sim \mathcal{IG} \left(\frac{RK}{2}, \frac{1}{2} \sum_{k=1}^K \boldsymbol{\psi}_k^T \boldsymbol{\psi}_k \right), \quad \delta | s^2 \sim \mathcal{G} \left(1, \frac{1}{s^2} \right). \quad (32)$$

V. SIMULATION RESULTS ON SYNTHETIC DATA

TABLE I
ACTUAL AND ESTIMATED ABUNDANCE MEAN AND VARIANCE IN EACH CLASS.

		Actual values	Estimated values
Class 1	$\mu_1 = E[a_{p, p \in \mathcal{I}_1}]$	[0.6, 0.3, 0.1] ^T	[0.57, 0.3, 0.13] ^T
	$Var[a_{p,r, p \in \mathcal{I}_1}] (\times 10^{-3})$	[5, 5, 5] ^T	[5.6, 6.7, 6.7] ^T
Class 2	$\mu_2 = E[a_{p, p \in \mathcal{I}_2}]$	[0.3, 0.5, 0.2] ^T	[0.29, 0.49, 0.2] ^T
	$Var[a_{p,r, p \in \mathcal{I}_2}] (\times 10^{-3})$	[5, 5, 5] ^T	[4.5, 5.2, 8.1] ^T
Class 3	$\mu_3 = E[a_{p, p \in \mathcal{I}_3}]$	[0.3, 0.2, 0.5] ^T	[0.3, 0.2, 0.5] ^T
	$Var[a_{p,r, p \in \mathcal{I}_3}] (\times 10^{-3})$	[5, 5, 5] ^T	[4.6, 5.7, 10.2] ^T

Many simulations have been conducted to illustrate the accuracy of the proposed algorithm. The first experiment considers a 25×25 synthetic image with $K = 3$ different classes. The image contains $R = 3$ mixed components whose spectra ($L = 413$ spectral bands) have been extracted from the spectral libraries distributed with the ENVI package [43] and are represented in Fig. 4. A label map shown in Fig. 5 (left) has been generated using (6) with $\beta = 1.1$.

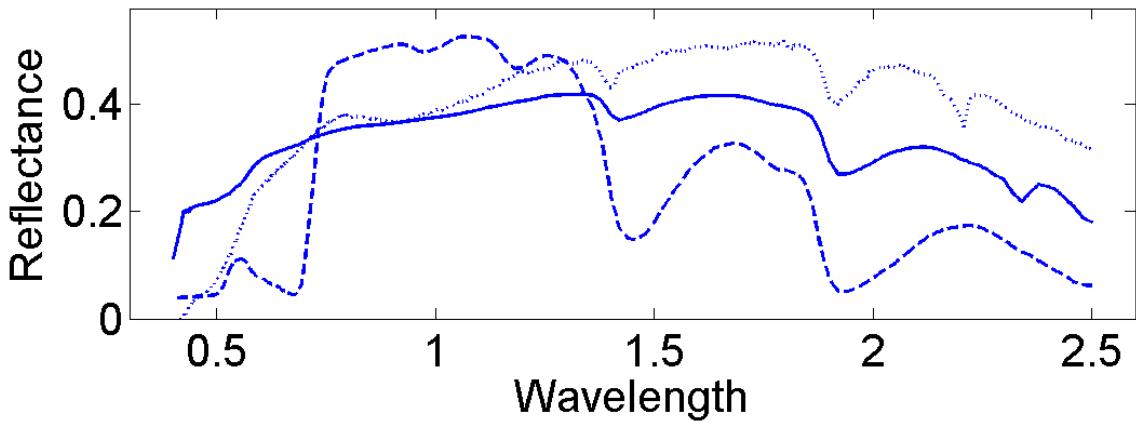


Fig. 4. The $R = 3$ endmember spectra: construction concrete (solid line), green grass (dashed line), micaceous loam (dotted line).

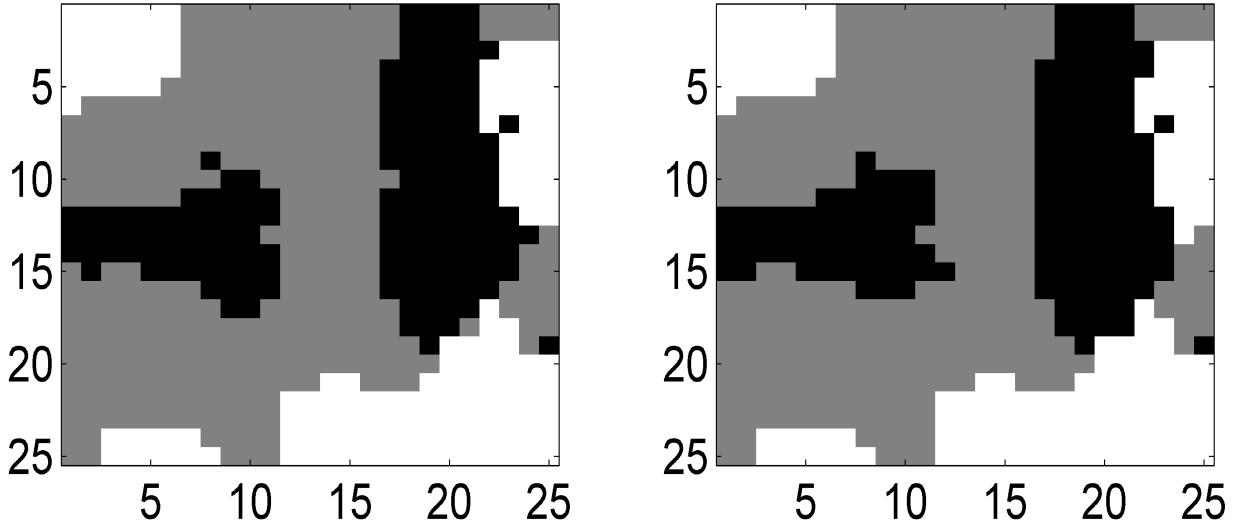


Fig. 5. Left: the actual label map. Right: the label map estimated by the LMM hybrid Gibbs sampler.

The mean and variance of the abundances have been chosen for each class as reported in Table I. These values reflect the fact that the 1st endmember is more present in Class 1 (with average concentration of 60%), the 2nd endmember is more present in Class 2 (with average concentration of 50%) and the 3rd endmember is more present in Class 3 (with average concentration of 50%). In this simulation scenario, the abundance variance has been fixed to a common value 0.005 for all endmembers, pixels and classes. The generated abundance maps for the LMM are depicted in Fig. 6. Note that a white (resp. black) pixel in the fraction map indicates a large (resp. small) value of the abundance coefficient. The noise variance is chosen such as the average signal-to-noise ratio (SNR) is equal to SNR = 19dB, i.e. $s^2 = 0.001$.

The MMSE and MAP estimators of the unknown parameters can be computed from samples generated with the Gibbs samplers presented in Section IV. For instance, the marginal MAP estimates of the label vector \hat{z}_{MAP} are depicted in Fig. 5 for the proposed hybrid Gibbs algorithm (right). The corresponding MMSE estimates of the abundances conditioned upon \hat{z}_{MAP} have been also shown in Fig. 6. A number of $N_{\text{MC}} = 5000$ iterations (with 500 burn-in iterations) has been necessary to obtain these results. Moreover, as mentioned in Section IV, the proposed algorithm generate samples distributed according to the full posterior of interest. Then, these samples can be used to compute, for instance, the posterior distributions of the mean vectors $\mu_k = \mathbb{E}[\mathbf{a}_p]$ ($k = 1, \dots, K, p \in \mathcal{I}_k$). These mean vectors, introduced in (4), are of great interest since they

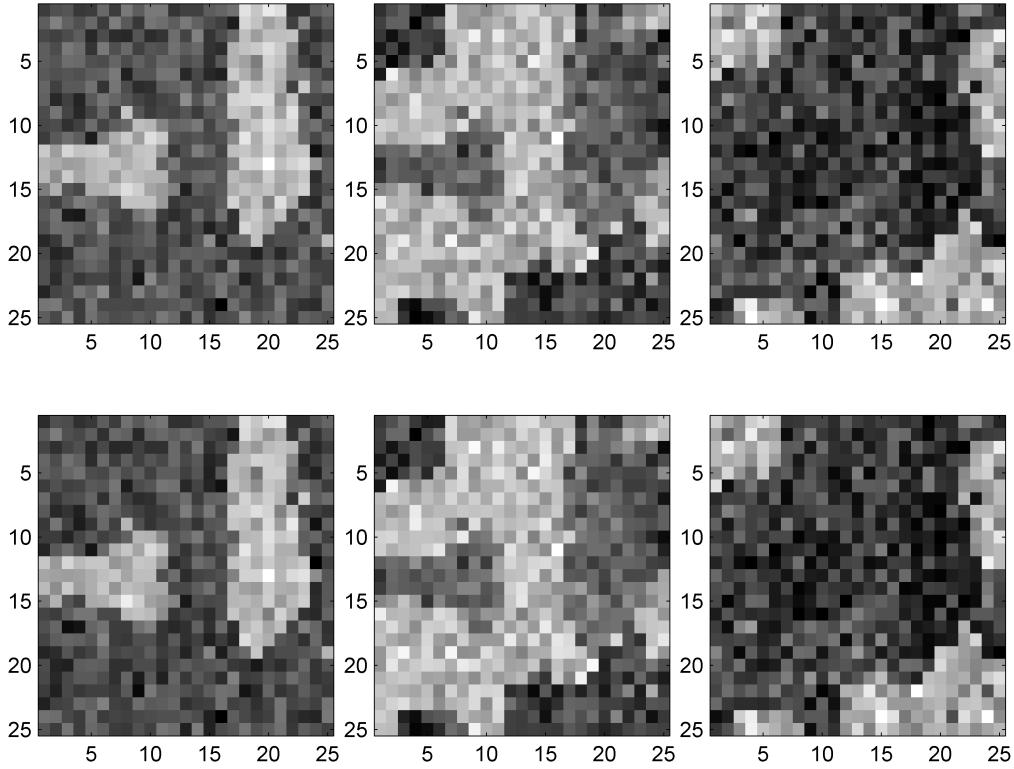


Fig. 6. Top: abundance maps of the 3 pure materials for LMM. Bottom: abundance maps of the 3 pure materials estimated by the hybrid Gibbs sampler (from left to right: construction concrete, green grass, micaceous loam).

are characteristics of each class $k = 1, \dots, K$, according to their definition in paragraph II-B. Therefore, as an additional insight, the histograms of the abundance means μ_k estimated by the proposed algorithm have been depicted in Fig. 7 for the 2nd class, i.e., $k = 2$. Similar results have been obtained for the other classes. They are omitted here for brevity. Finally, the estimated abundance means and variances for each endmember in each class have been reported in Table I (last row). The estimated classes, abundance coefficients and abundance mean vectors estimated by the algorithm are clearly in accordance with the actual ones.

The LMM hybrid Gibbs algorithm is compared respectively to its non-spatial constrained Bayesian counterpart developed in [9]. The synthetic image shown in Fig. 5 has been analyzed by the initial algorithm of [9] with the same number of iterations N_{MC} in addition with the FCLS [7] algorithm. As a criterion, the global mean square error (MSE) of the r th estimated

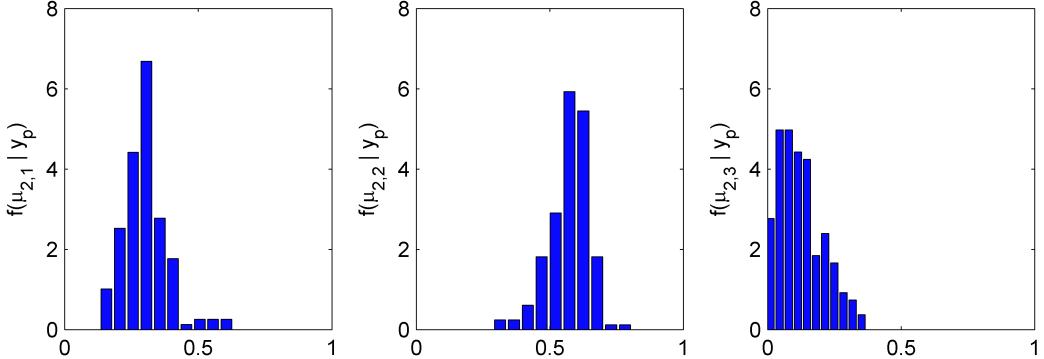


Fig. 7. Top (resp. bottom): histograms of the abundance means $\mu_k = [\mu_{k,1}, \mu_{k,2}, \mu_{k,3}]^T$ estimated by the proposed hybrid Gibbs algorithm for the 2nd class ($k = 2$).

abundances have been computed for each algorithm. This global MSE is defined as

$$\text{MSE}_r^2 = \frac{1}{P} \sum_{p=1}^P (\hat{a}_{r,p} - a_{r,p})^2 \quad (33)$$

where $\hat{a}_{r,p}$ denotes the MMSE estimate of the abundance $a_{r,p}$. Table II reports the different results showing that the algorithm developed in this paper (referred to as “Spatial”) performs better than the non-spatial constrained algorithms (referred to as “Bayesian” and “FCLS”).

TABLE II
GLOBAL MSEs OF EACH ABUNDANCE COMPONENT.

	FCLS	Bayesian	Spatial
MSE_1^2	0.0019	0.0016	$3.1 \times$
MSE_2^2	4.3×10^{-4}	4.1×10^{-4}	8.98×10^{-5}
MSE_3^2	0.0014	0.0013	2.35×10^{-4}

VI. SIMULATION RESULTS ON AVIRIS IMAGES

A. Performance of the proposed algorithm

This section illustrates the performance of the proposed spatial algorithm on a real hyperspectral dataset, acquired over Moffett Field (CA, USA) in 1997 by the JPL spectro-imager AVIRIS.

Many previous works have used this image to illustrate and compare algorithm performance with hyperspectral images [44], [45]. The first region of interest, represented in Fig. 8, is a 50×50 pixel image. The data set has been reduced from the original 224 bands to $L = 189$ bands by removing water absorption bands. As in [9], a principal component analysis has been conducted as a processing step to determine the number of endmembers present in the scene³. Then, the endmembers spectra have been extracted with the help of the endmember extraction procedure N-FINDR proposed by Winter in [47]. The $R = 3$ extracted endmembers, shown in Fig. 9, corresponds to soil, vegetation and water. The algorithm proposed in Section IV have been applied on this image with a number of $K = 4$ classes and $N_{MC} = 5000$ iterations (with 500 burn-in iterations). The number of classes has been fixed to $K = 4$ since prior knowledge on the scene allows one to identify 4 areas in the image: water point, lake shore, vegetation and soil.



Fig. 8. Real hyperspectral data: Moffett field acquired by AVIRIS in 1997 (left) and the region of interest shown in true colors (right).

³This number could have been estimated using the method presented in [46] or by setting R to a given maximum value where the missing endmembers in the analyzed image would have abundances close to 0

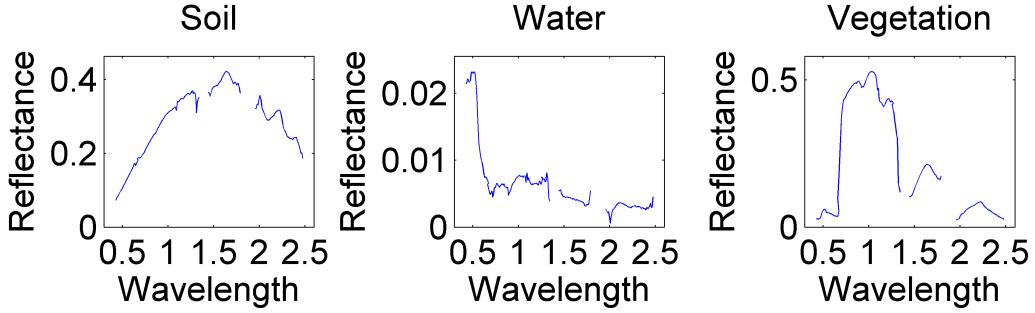


Fig. 9. The $R = 3$ endmember spectra obtained by the N-FINDR algorithm.

The estimated classification and abundance maps for the proposed hybrid Gibbs algorithm are depicted in Fig. 10 (left) and 11 (top). The results provided by the algorithm are very similar and in good agreement with results obtained on this image with an LMM-based Bayesian algorithm [9] (Fig. 11, middle) or with the well-known FCLS algorithm [7] (Fig. 11, bottom).

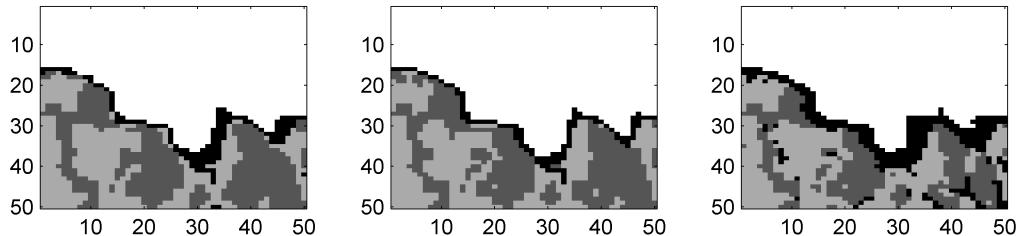


Fig. 10. Classification maps obtained for $K = 4$ with $R = 3$ (left), $R = 4$ (middle) and $R = 5$ (right).

The computational time of the proposed method (combined with the N-FINDR procedure) has been compared with the computational times of two other unmixing algorithms when applied on this image: the FCLS algorithm also combined with the N-FINDR, and the constrained nonnegative matrix factorization (cNMF) algorithm that jointly estimates the endmember matrix and the abundances [48]. The results are reported in Table III. Note these simulations have been carried out with an unoptimized MATLAB 2007b 32bit implementation on a Core(TM)2Duo 2.66GHz computer. The proposed method (referred to as “Spatial”) has the longest computational time when compared to the two others, mainly due to the joint estimation of the labels and the

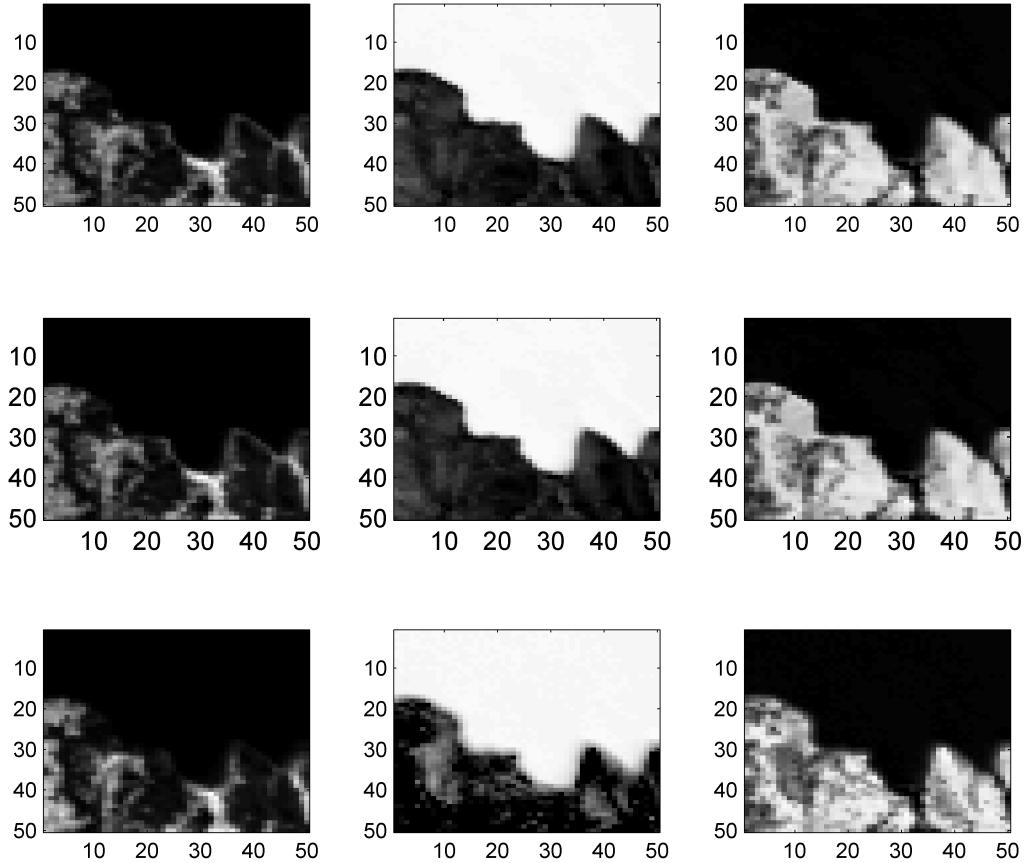


Fig. 11. Top: abundance maps estimated by the proposed algorithm (from left to right: vegetation, water and soil). Middle: abundance maps estimated by the LMM-based Bayesian algorithm (from [9]). Bottom: fraction maps estimated by the FCLS algorithm [7].

abundance vectors.

TABLE III
COMPUTATIONAL TIMES OF LMM-BASED UNMIXING ALGORITHMS.

	FCLS	cNMF	Spatial
Times (s.)	0.388	2.5×10^3	8.4×10^3

B. Influence of R and K

This section studies the influence of the number of endmembers R and the number of classes K on the performance of the proposed algorithm. First, the proposed algorithm has been applied on the Moffett field image with larger numbers of endmembers ($R = 4$ and $R = 5$) and a number of classes fixed to $K = 4$ (as in the previous paragraph). The estimated classification and abundance maps obtained with these two different values of R are depicted in Fig. 10 (middle and right) and 12, respectively.

By comparing these results with those obtained previously in Fig. 10 (left) and 11, we can conclude that the label estimation is robust with respect to the number of endmembers R . Moreover, increasing the number of endmembers does not drastically modify the “water” and “vegetation” abundance maps. Conversely, the “soil” abundance map have been split into different materials when $R = 5$.

The influence of the number of classes K has also been investigated by applying the proposed algorithm on the Moffett image with larger numbers of classes ($K = 5$ and $K = 10$) and a number of endmembers equal to $R = 3$ (as in the previous paragraph). The estimated classification and abundance maps obtained with these two different values of K are given in Fig. 13 (middle and right) and 14, respectively.

Again, by comparing these results with those given in Fig. 13 (left) and 11, we can notice that changing the number of classes does not significantly affect the abundance estimation. However, changes can be noticed on the classification maps when the number of classes increases. For both scenarios ($K = 5$ and $K = 10$), the continental area seems to suffer from an over-segmentation. However, by examining the class characteristics, especially the mean vectors ψ_k that define the classes (as specified in (4)), we can notice that some of these classes share the same statistics. As an example, the estimated histograms of the abundance means are depicted in Fig. 15 for the “soil” class (dark gray) obtained with $K = 4$ (top) and for two merged “continental” classes (dark and light gray) obtained with $K = 5$ (bottom). These histograms are very similar, which means that the corresponding regions are of identical macroscopic composition. Similarly, when K increases from $K = 4$ to $K = 5$ and 10, the lake area is mainly divided into two different water regions. Therefore, the effect of increasing the number of classes is mainly to divide classes of pure pixels into classes containing similar materials.

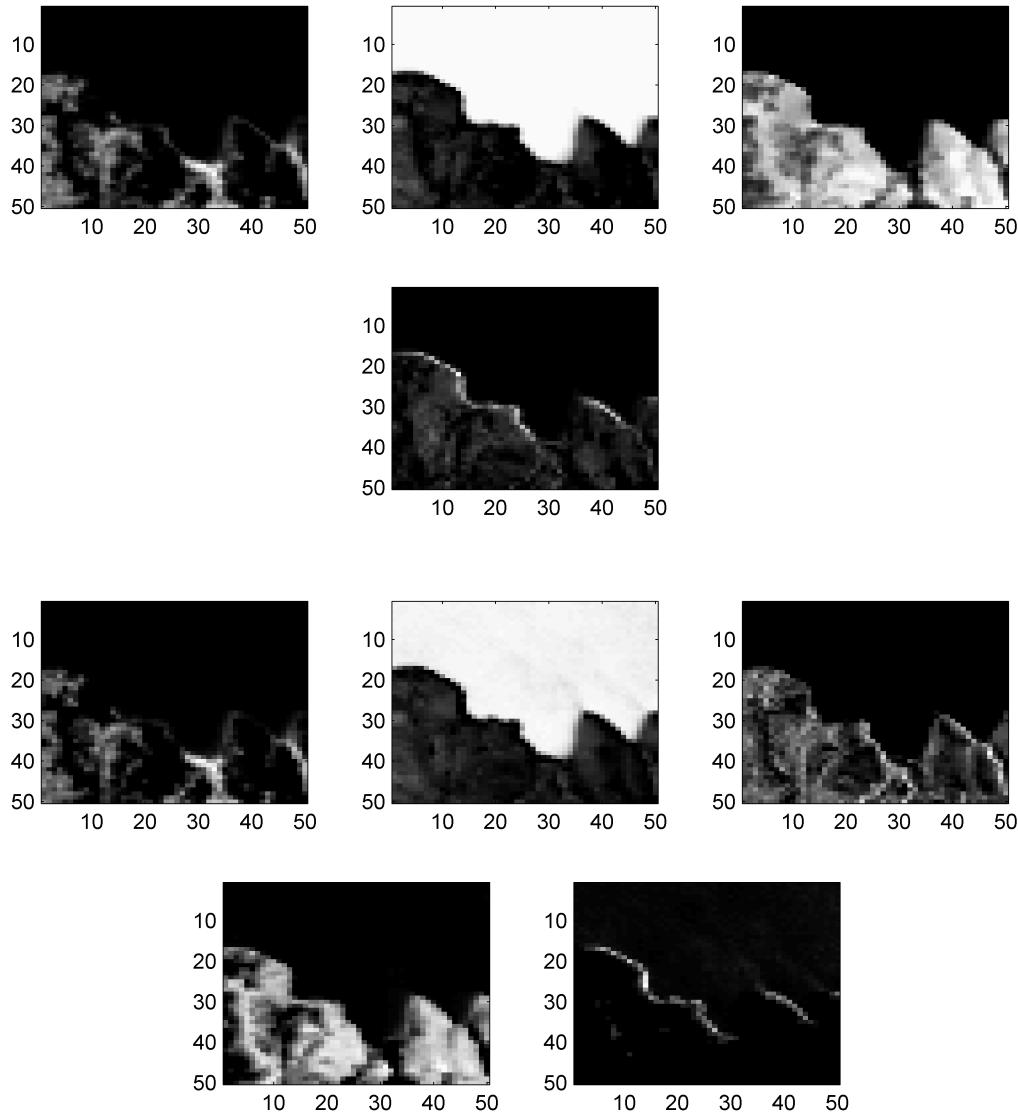


Fig. 12. Abundance maps obtained with $R = 4$ (4 top maps) and $R = 5$ (5 bottom maps).

C. Influence of the endmember extraction algorithm

In this paragraph, the influence of the EEA used to perform the endmember extraction is analyzed. The proposed algorithm is coupled with several endmember extraction methods. The first EEA is the well-known N-FINDR, used in our previous experiments. The second EEA, known as MVSA, consist of fitting a simplex of minimum volume to the hyperspectral data, constraining the abundance fractions to belong to this simplex and solving the resulting problem

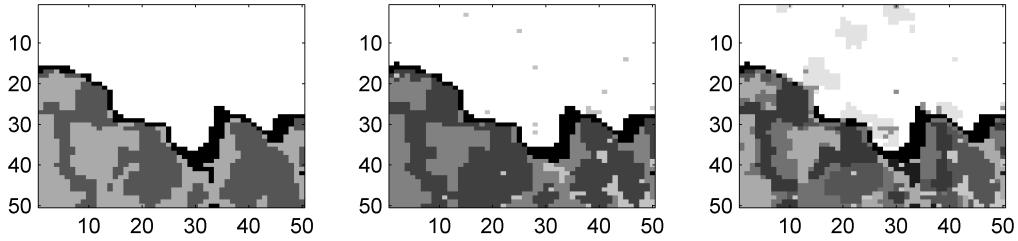


Fig. 13. Classification maps obtained for $R = 3$ with $K = 4$ (left), $K = 5$ (middle) and $K = 10$ (right).

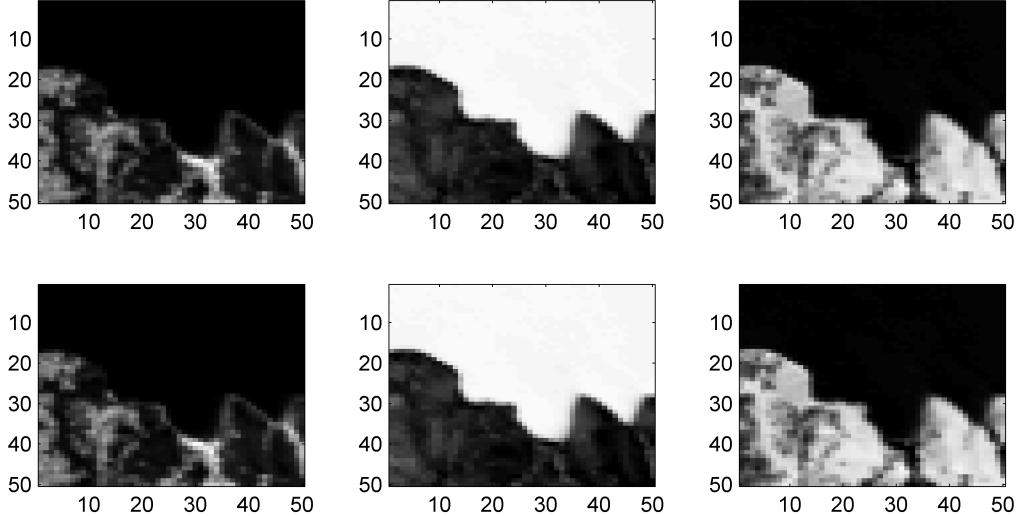


Fig. 14. Abundance maps obtained with $K = 5$ (top) and $K = 10$ (bottom).

by using a sequence of quadratically constrained subproblems [49]. The third method, referred to as “spatial-spectral preprocessing” (SSP) in the sequel of this paper, consists of several preprocessing steps designed to improve the accuracy of volume-based EEAs [50]. Our proposed algorithm has been applied on the Moffett field image with a number of endmembers $R = 3$. The resulting spectra extracted by the three algorithms are represented in Fig. 9 (N-FINDR), 16 (MVSA, top) and 16 (SSP, bottom). It can be noticed that spectra associated to a given material extracted by the three EEAs are very similar. The corresponding abundance maps estimated by the proposed Bayesian algorithm coupled with these EEAs are given in Fig. 11 (top), 18 (MVSA,

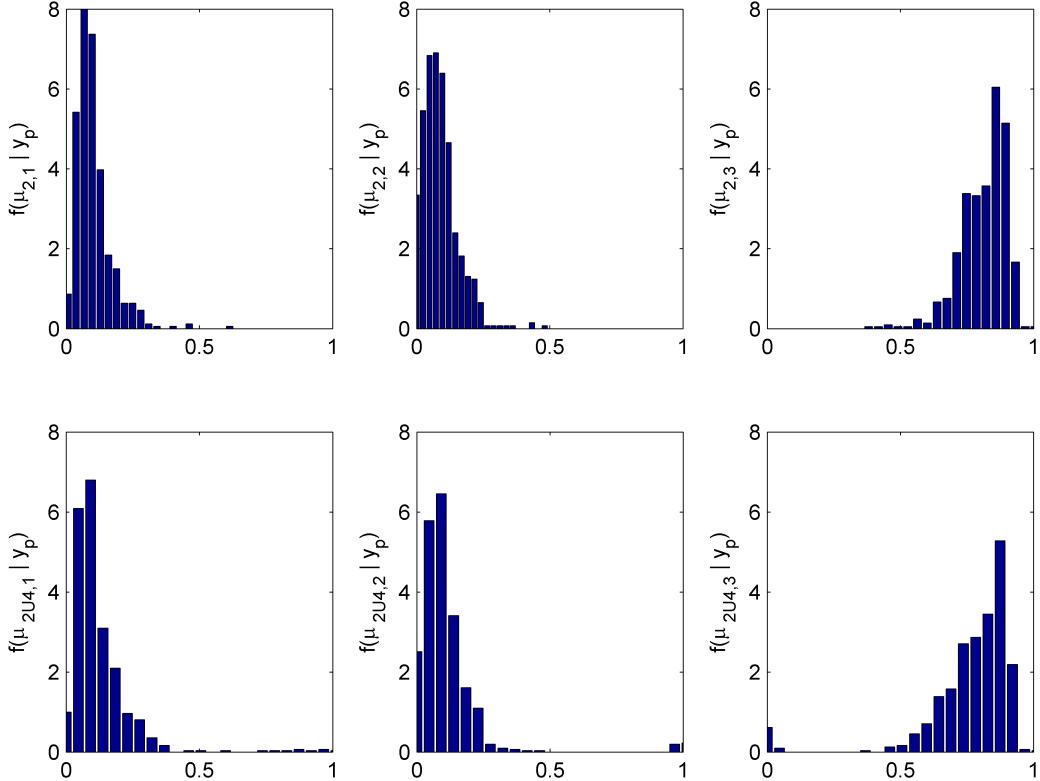


Fig. 15. Top: histograms for the “continental” class obtained for $K = 4$ (from left to right: vegetation, water and soil). Bottom: histograms for two merged “continental” classes obtained for $K = 5$.

top) and 18 (SSP, bottom). Similarly, the estimated classification maps are depicted in Fig. 10 (left) and 17 (with a number of classes fixed to $K = 4$). As expected, the spectra extracted by the three EEAs are very similar, resulting in close abundance estimation and classification results.

D. Simulation on a larger image

The performance proposed Bayesian algorithm has also been evaluated on a larger real hyperspectral image. The sample chosen has been extracted from the AVIRIS Cuprite scene, acquired over a mining site in Nevada, in 1997. The geologic characteristics of the complete data have been mapped in [51], [52]. The area of interest, of size 190×250 and represented in Fig. 19, has been previously studied in [53] to test the VCA algorithm with $R = 14$. Therefore,

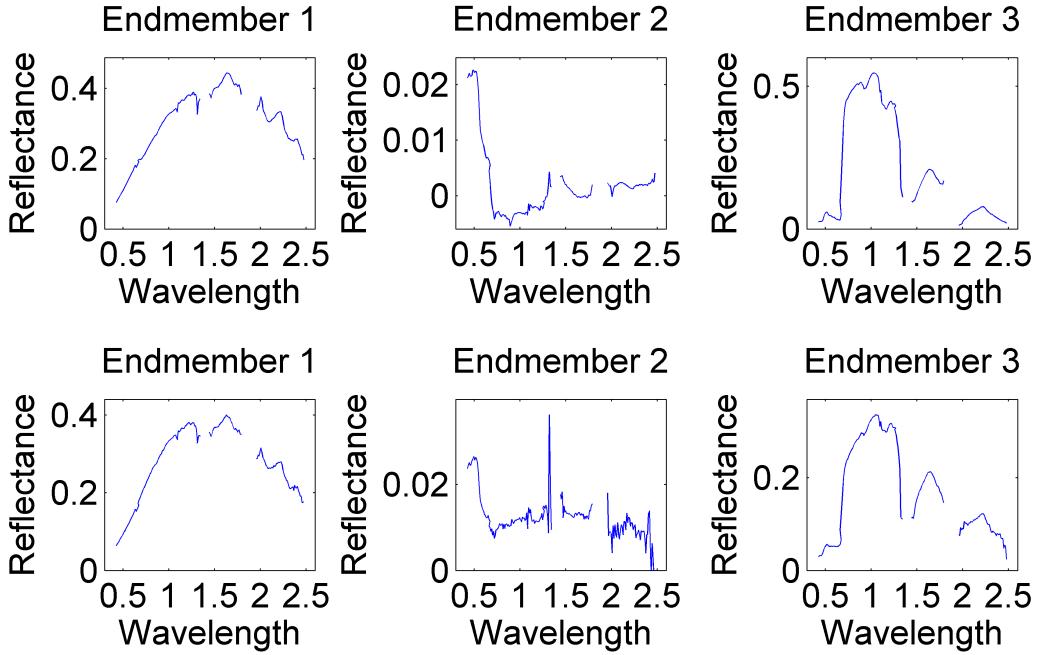


Fig. 16. Estimated endmember signatures obtained with MVSA algorithm (top) and with the spatial-spectral preprocessing (bottom).

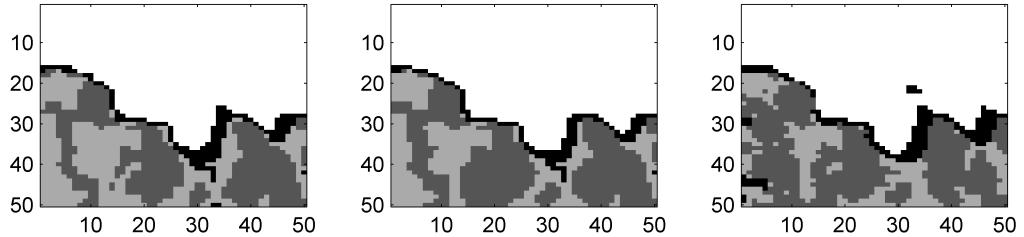


Fig. 17. Classification maps obtained for $R = 3$ and $K = 4$ with N-FINDR (left), MVSA algorithm (middle) and the spatial-spectral preprocessing (right).

in this experiment, the same number of endmembers has been extracted by the VCA algorithm. The number of classes has also been set to $K = 14$, which seems to be a sufficient value to capture the natural diversity of the scene. Then, the proposed algorithm has been used to estimate the abundance and label maps related to the analyzed scene. These maps are depicted in Fig. 20 and 22, respectively.

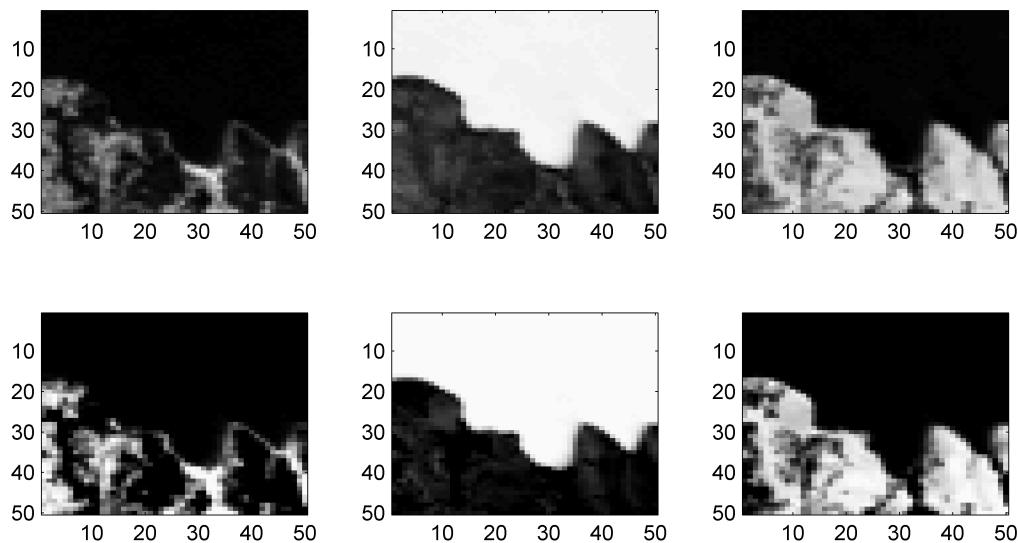


Fig. 18. Fraction maps obtained for $R = 3$ with the MVSA algorithm (top) and with the spatial-spectral preprocessing (bottom).

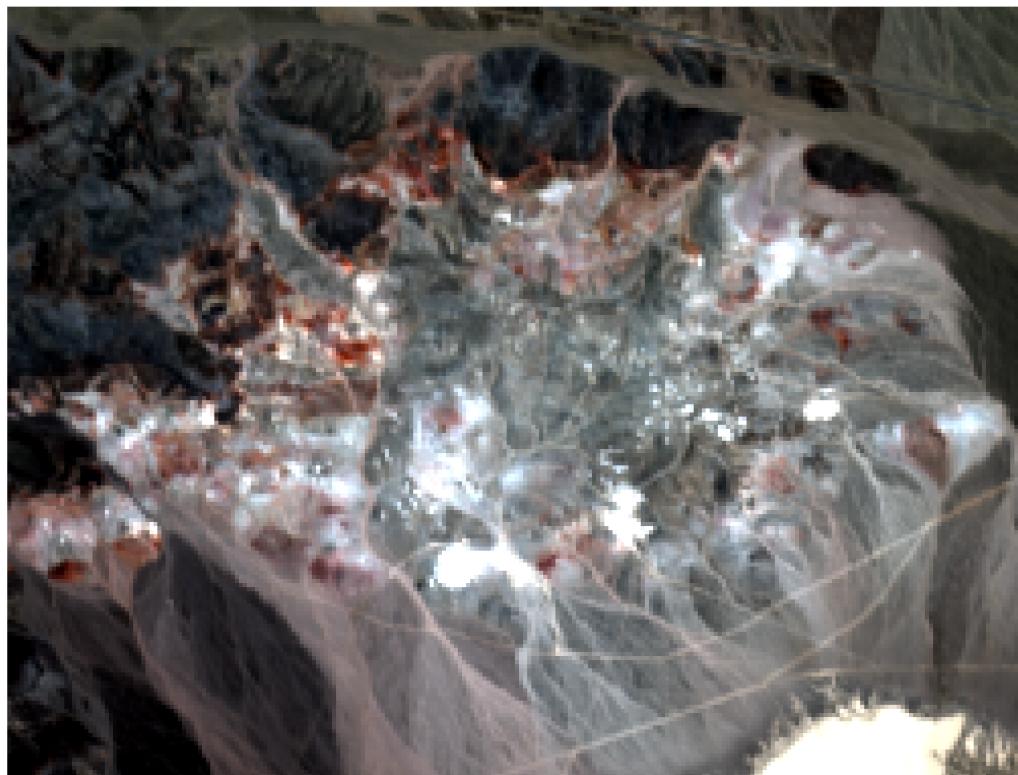


Fig. 19. AVIRIS image of 190×250 pixels extracted from Cuprite scene observed in composite natural colors.

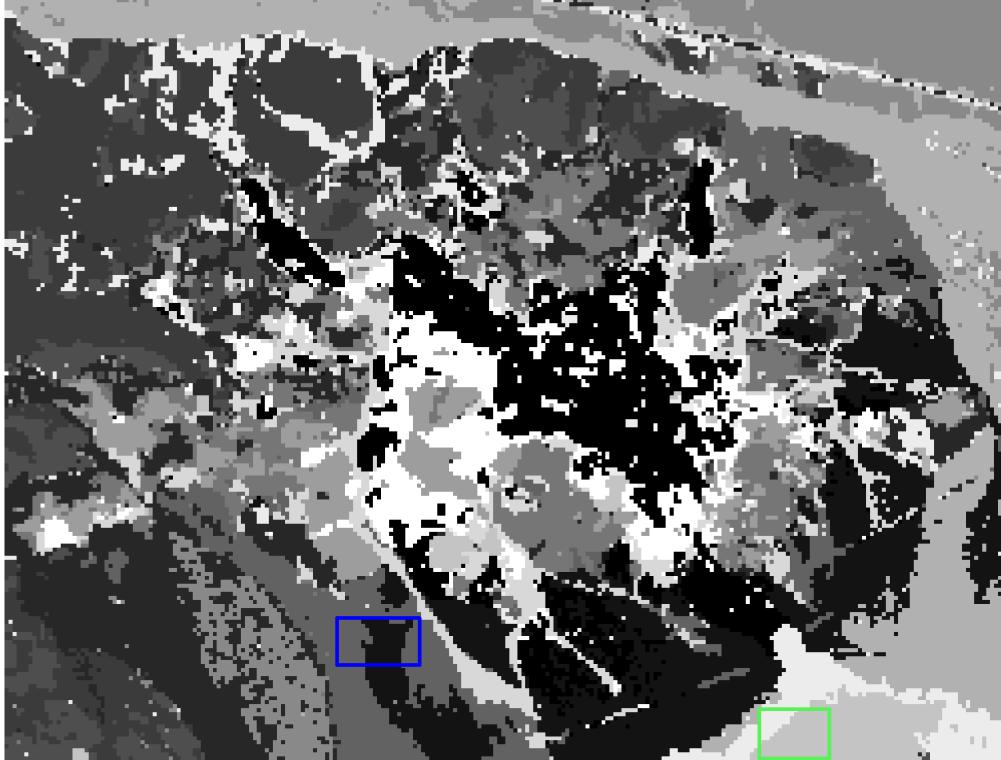


Fig. 20. Classification map obtained on the 190×250 Cuprite area with $K = 14$.

The proposed Bayesian inversion algorithm has been able to identify some regions similar to those recovered in [53]. To illustrate, the composition of two particular areas (marked as colored rectangles in Fig. 20) is investigated. Tables IV report the abundance means for the most significant endmembers that appear in the two highlighted regions. From these tables, we can conclude that the two classes represented in black and dark gray of the “blue” area are composed of very mixed pixels (the abundance of the most significant endmember is 0.201). On the other hand, both classes in the “green” area are clearly dominated by the 6th endmember. By comparing its corresponding signature with the materials included in the USGS library spectra, this 6th endmember matches the Montmorillonite spectrum (see Fig. 21). This result is in good agreement with the ground truth. Indeed, from [52], Montmorillonite is the most commonly found material in this area.

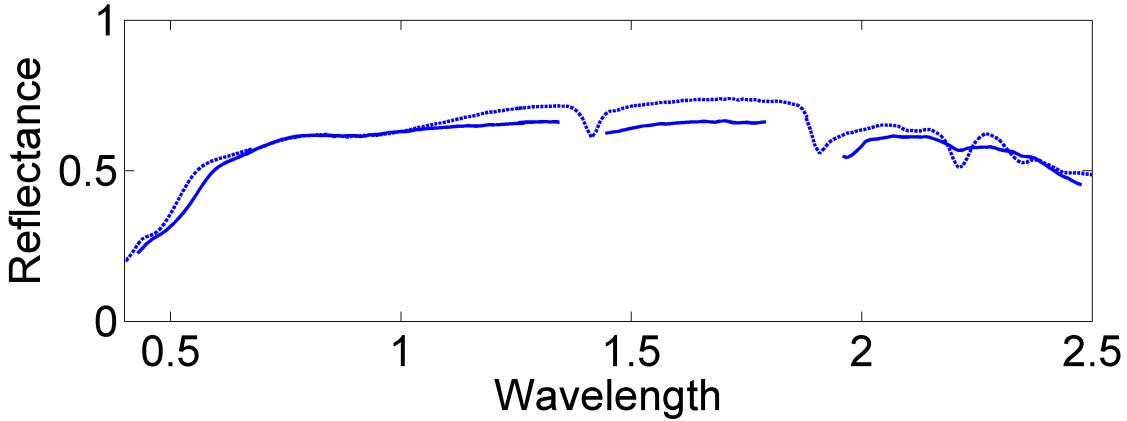


Fig. 21. Comparison of the 6th endmember spectrum extracted by the VCA algorithm (solid line) with the Montmorillonite signature extracted from the USGS spectral library (dashed line).

TABLE IV
ABUNDANCE MEANS FOR THE MOST SIGNIFICANT ENDMEMBERS IN EACH HIGHLIGHTED REGION.

Green area		Blue area			
	light gray		dark gray		
Endm. 1	0.001	0.225	Endm. 1	0.135	0.044
Endm. 3	0.045	0.000	Endm. 9	0.155	0.158
Endm. 5	0.098	0.027	Endm. 10	0.159	0.127
Endm. 6	0.839	0.528	Endm. 13	0.187	0.206

VII. CONCLUSIONS

A new hierarchical Bayesian algorithm was proposed for hyperspectral image unmixing. Markov random fields were introduced to model spatial correlations between the pixels of the image. A hidden discrete label was introduced for each pixel of the image to identify several classes defined by homogeneous abundances (with constant first and second order statistical moments). The positivity and sum-to-one constraints on the abundances were handled by using an appropriate reparametrization defined by logistic coefficient vectors. We derived the joint posterior distribution of the unknown parameters and hyperparameters associated to the pro-

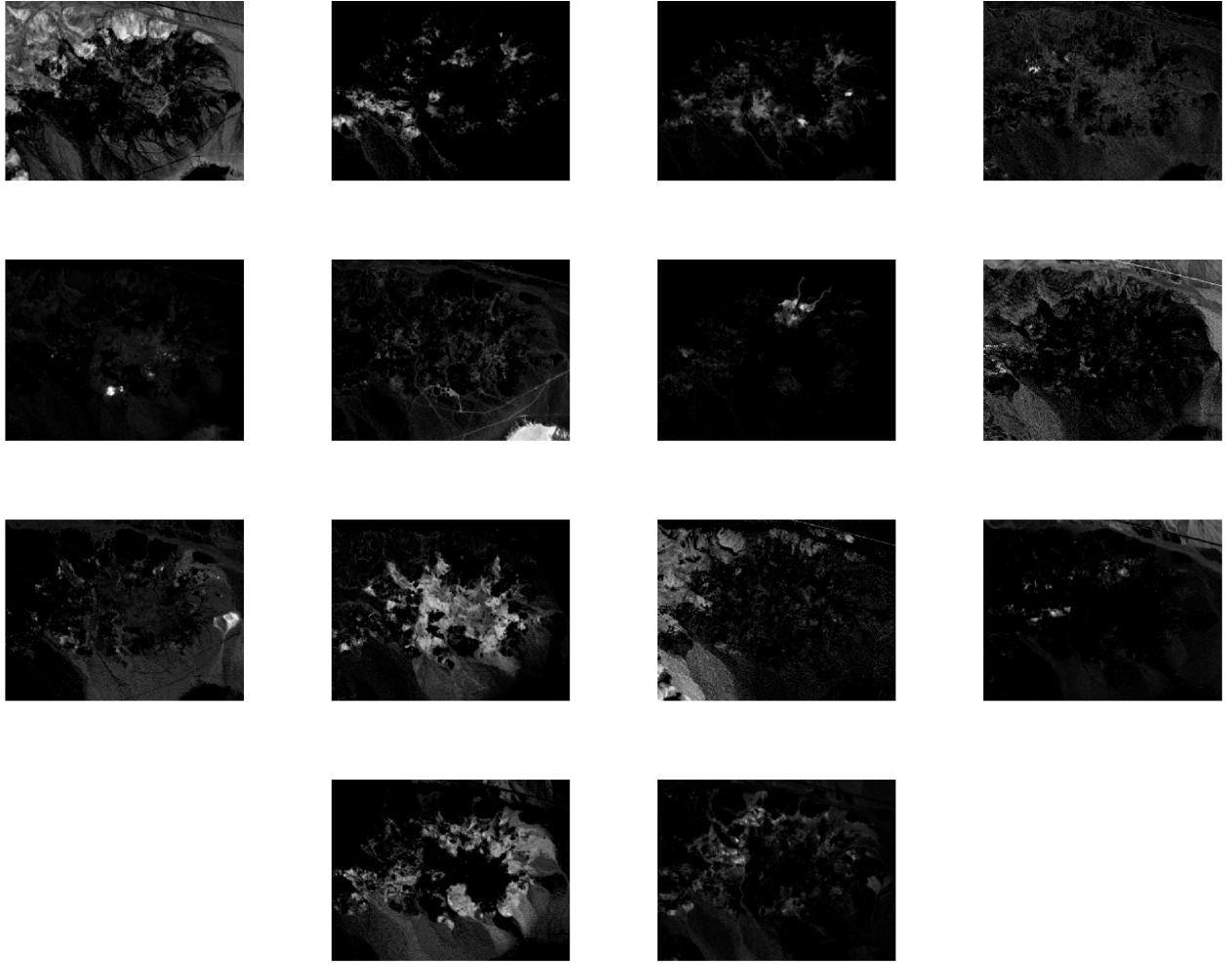


Fig. 22. Fraction maps of the 190×250 Cuprite area.

posed Bayesian linear mixing model. An MCMC method was then studied to generate samples asymptotically distributed according to this posterior. The generated samples were then used to estimate the abundance maps as well as the underlying image labels. The results obtained on simulated data and on real AVIRIS images are very promising. Future works include the estimation of the granularity coefficient involved in Potts-Markov random fields.

ACKNOWLEDGMENTS

The authors would like to thank one of the reviewers for pointing out the relevant paper [12] and for his valuable suggestions that helped to improve the manuscript.

REFERENCES

- [1] K. Jusoff, "Precision forestry using airborne hyperspectral imaging sensor," *Journal of Agricultural Science*, vol. 1, no. 1, pp. 142–147, June 2009.
- [2] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, pp. 44–56, Jan. 2002.
- [3] C.-I Chang, *Hyperspectral data exploitation: theory and applications*. Hoboken, NJ: Wiley, 2007.
- [4] M. T. Eismann and D. Stein, "Stochastic mixture modeling," in *Hyperspectral Data Exploitation: Theory and Applications*, C.-I Chang, Ed. Wiley, 2007, ch. 5.
- [5] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "NCM-based Bayesian algorithm for hyperspectral unmixing," in *Proc. IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Grenoble, France, Aug. 2009, pp. 1–4.
- [6] P. J. Martinez, R. M. Perez, A. Plaza, P. L. Aguilar, M. C. Cantero, and J. Plaza, "Endmember extraction algorithms from hyperspectral images," *Annals of Geophysics*, vol. 49, no. 1, pp. 93–101, Feb. 2006.
- [7] D. C. Heinz and C.-I Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 39, no. 3, pp. 529–545, March 2001.
- [8] C. Theys, N. Dobigeon, J.-Y. Tourneret, and H. Lantéri, "Linear unmixing of hyperspectral images using a scaled gradient method," in *Proc. IEEE-SP Workshop Stat. and Signal Processing (SSP)*, Cardiff, UK, Aug. 2009, pp. 729–732.
- [9] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 2684–2696, July 2008.
- [10] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc. Ser. B*, vol. 36, no. 2, pp. 192–236, 1974.
- [11] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.
- [12] J. T. Kent and K. V. Mardia, "Spatial classification using fuzzy membership models," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 10, no. 5, pp. 659–671, 1988.
- [13] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.
- [14] F. Wu, "The Potts model," *Rev. Modern Phys.*, vol. 54, no. 1, pp. 235–268, Jan. 1982.
- [15] A. Mohammadpour, O. Féron, and A. Mohammad-Djafari, "Bayesian segmentation of hyperspectral images," in *Proc. Workshop on Bayesian Inference and Maximum Entropy Methods (MaxEnt)*, ser. AIP Conference Series, R. Fischer, R. Preuss, and U. V. Toussaint, Eds., vol. 735, Max Plank Inst., Germany, Nov. 2004, pp. 541–548.
- [16] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia, "Texture feature analysis using a Gauss-Markov model in hyperspectral image classification," *IEEE Trans. Geosci. and Remote Sensing*, vol. 42, no. 7, pp. 1543–1551, July 2004.
- [17] R. Neher and A. Srivastava, "A Bayesian MRF framework for labeling terrain using hyperspectral imaging," *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, no. 6, pp. 1363–1374, June 2005.
- [18] O. Feron and A. Mohammad-Djafari, "Image fusion and unsupervised joint segmentation using a HMM and MCMC algorithms," *J. Electron. Imaging*, vol. 14, no. 2, May 2005.
- [19] N. Bali and A. Mohammad-Djafari, "Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis," *IEEE Trans. Image Processing*, vol. 17, no. 2, pp. 217–225, Feb. 2008.

- [20] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. and Remote Sensing*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [21] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
- [22] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. and Remote Sensing Letters*, vol. 7, no. 4, pp. 736–740, Oct. 2010.
- [23] J. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. and Remote Sensing*, vol. 32, pp. 779–785, July 1994.
- [24] C.-I Chang, "Further results on relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. and Remote Sensing*, vol. 36, no. 3, pp. 1030–1032, May 1998.
- [25] C.-I Chang, X.-L. Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. and Remote Sensing*, vol. 36, no. 3, pp. 898–912, May 1998.
- [26] C. Kevrann and F. Heitz, "A Markov random field model-based approach to unsupervised texture segmentation using local and global statistics," *IEEE Trans. Image Processing*, vol. 4, no. 6, pp. 856–862, June 1995.
- [27] A. Tonazzini, L. Bedini, and E. Salerno, "A Markov model for blind image separation by a mean-field EM algorithm," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 473–481, Feb. 2006.
- [28] R. S. Rand and D. M. Keenan, "Spatially smooth partitioning of hyperspectral imagery using spectral/spatial measures of disparity," *IEEE Trans. Geosci. and Remote Sensing*, vol. 41, no. 6, pp. 1479–1490, June 2003.
- [29] R. Kindermann and J. L. Snell, *Markov random fields and their applications*. Providence, RI: Amer. Math. Soc., 1980.
- [30] J.-M. Marin and C. P. Robert, *Bayesian core: a practical approach to computational Bayesian statistics*. New-York: Springer, 2007.
- [31] B. D. Ripley, *Statistical inference for spatial processes*. Cambridge: Cambridge University Press, 1988.
- [32] Z. Zhou, R. Leahy, and J. Qi, "Approximate maximum likelihood hyperparameter estimation for Gibbs prior," *IEEE Trans. Image Processing*, vol. 6, no. 6, pp. 844–861, June 1997.
- [33] X. Descombes, R. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood," *IEEE Trans. Image Processing*, vol. 8, no. 7, pp. 945–963, July 1999.
- [34] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131–144, 2003.
- [35] A. Gelman, F. Bois, and J. Jiang, "Physiological pharmacokinetic analysis using population modeling and informative prior distributions," *J. Amer. Math. Soc.*, vol. 91, no. 436, pp. 1400–1412, Dec. 1996.
- [36] K. Themelis and A. A. Rontogiannis, "A soft constrained MAP estimator for supervised hyperspectral signal unmixing," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.
- [37] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 747–758, March 2002.
- [38] N. Dobigeon, J.-Y. Tourneret, and M. Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1251–1263, April 2007.
- [39] C. P. Robert, *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*, 2nd ed., ser. Springer Texts in Statistics. New York: Springer-Verlag, 2007.

- [40] N. Dobigeon, J.-Y. Tourneret, and M. Davy, “Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach,” *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1251–1263, April 2007.
- [41] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer Verlag, 2004.
- [42] G. O. Roberts, “Markov chain concepts related to sampling algorithms,” in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London: Chapman & Hall, 1996, pp. 259–273.
- [43] RSI (Research Systems Inc.), *ENVI User’s guide Version 4.0*, Boulder, CO 80301 USA, Sept. 2003.
- [44] E. Christophe, D. Léger, and C. Mailhes, “Quality criteria benchmark for hyperspectral imagery,” *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, no. 9, pp. 2103–2114, Sept. 2005.
- [45] T. Akgun, Y. Altunbasak, and R. M. Mersereau, “Super-resolution reconstruction of hyperspectral images,” *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.
- [46] O. Eches, N. Dobigeon, and J.-Y. Tourneret, “Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical Bayesian algorithm,” *IEEE J. Sel. Topics Signal Processing*, vol. 3, no. 3, pp. 582–591, Jun. 2010.
- [47] M. E. Winter, “Fast autonomous spectral endmember determination in hyperspectral data,” in *Proc. 13th Int. Conf. on Applied Geologic Remote Sensing*, vol. 2, Vancouver, April 1999, pp. 337–344.
- [48] P. Sajda, S. Du, T. R. Brown, R. Stoyanova, D. C. Shungu, X. Mao, and L. C. Parra, “Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain,” *IEEE Trans. Medical Imaging*, vol. 23, no. 12, pp. 1453–1465, 2004.
- [49] J. Li and J. M. Bioucas-Dias, “Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data,” in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, vol. 3, Boston, MA, Jul. 2008, pp. 250–253.
- [50] G. Martin and A. Plaza, “Spatial-spectral preprocessing for volume-based endmember extraction algorithm using unsupervised clustering,” in *Proc. IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS’10)*, Reykjavik, Iceland, Jun. 2010.
- [51] R. N. Clark, G. A. Swayze, and A. Gallagher, “Mapping minerals with imaging spectroscopy, U.S. Geological Survey,” *Office of Mineral Resources Bulletin*, vol. 2039, pp. 141–150, 1993.
- [52] R. N. Clark *et al.*, “Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems,” *J. Geophys. Res.*, vol. 108, no. E12, pp. 5–1–5–44, Dec. 2003.
- [53] J. M. Nascimento and J. M. Bioucas-Dias, “Vertex component analysis: a fast algorithm to unmix hyperspectral data,” *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005.