Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical Bayesian algorithm.

Olivier Eches, Nicolas Dobigeon, Jean-Yves Tourneret

Estimating the Number of Endmembers in Hyperspectral Images Using the Normal Compositional Model and a Hierarchical Bayesian Algorithm

Olivier Eches, Nicolas Dobigeon, *Member, IEEE*, and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*—This paper studies a semi-supervised Bayesian unmixing algorithm for hyperspectral images. This algorithm is based on the normal compositional model recently introduced by Eismann and Stein. The normal compositional model assumes that each pixel of the image is modeled as a linear combination of an unknown number of pure materials, called *endmembers*. However, contrary to the classical linear mixing model, these endmembers are supposed to be random in order to model uncertainties regarding their knowledge. This paper proposes to estimate the mixture coefficients of the Normal Compositional Model (referred to as *abundances*) as well as their number using a reversible jump Bayesian algorithm. The performance of the proposed methodology is evaluated thanks to simulations conducted on synthetic and real AVIRIS images.**

***Index Terms*—Bayesian inference, hyperspectral images, Monte Carlo methods, normal compositional model, reversible jump, spectral unmixing.**

I. Introduction

FOR several decades, hyperspectral imagery has been receiving growing interest in the signal and image processing literature (see [1] and references therein). This interest can be easily explained by the high spectral resolution of the images provided by the hyperspectral sensors such as AVIRIS [2] and Hyperion [3]. Spectral unmixing is a crucial step in the analysis of these images [4]. It consists of decomposing the measured pixel reflectances into a mixture of pure spectra whose fractions, referred to as abundances, have to be estimated. Most unmixing procedures for hyperspectral images assume that the image pixels are linear combinations of pure materials. More precisely, according to the linear mixing model (LMM) presented in [4], the L-spectrum $\boldsymbol{y} = [y_1, \ldots, y_L]^T$ of a mixed pixel is modeled as a linear combination of R spectra $\boldsymbol{m}_r$ corrupted by additive white Gaussian noise

$$\boldsymbol{y} = \sum_{r=1}^{R} \boldsymbol{m}_r \alpha_r + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ denotes the spectrum of the rth material (referred to as endmember), α_r is the fraction of the rth material in the pixel (referred to as abundance), R is the number of pure materials present in the observed scene, and L is the number of available spectral bands for the image. In addition, due to obvious physical considerations, the abundances satisfy the following positivity and sum-to-one constraints:

$$\begin{cases} \alpha_r \geq 0, \quad \forall r = 1, \ldots, R \\ \sum_{r=1}^{R} \alpha_r = 1. \end{cases} \tag{2}$$

Supervised algorithms assume that the R endmember spectra $\boldsymbol{m}_r$ are known, e.g., extracted from a spectral library. In practical applications, they can be obtained by an endmember extraction procedure such as the well-known N-FINDR algorithm developed by Winter [5] or the vertex component analysis (VCA) proposed by Nascimento *et al.* [6]. However, the LMM has some limitations when applied to real images [4]. For instance, the endmember extraction procedures based on the LMM can be inefficient when the image does not contain enough pure pixels. This problem, also outlined by Nascimento in [6], is illustrated in Fig. 1. This figure shows 1) the projections on the two most discriminant axes identified by a principal component analysis (PCA) of $R = 3$ endmembers corresponding to the red/light stars, i.e., to the vertices of the red/light triangle, 2) the domain containing all linear combinations of the actual endmembers (i.e., the red/light triangle), and 3) the simplex estimated by the N-FINDR algorithm using the black pixels (blue/dark triangle). As there is no pixel close to the vertices of the red/light triangle, the N-FINDR estimates a much smaller simplex (in blue/dark) than the actual one (in red/light).

An alternative model referred to as normal compositional model (NCM) has been recently proposed in [7]. The NCM allows one to alleviate the problems mentioned above by assuming that the pixels of the hyperspectral image are linear combinations of random endmembers (as opposed to deterministic for the LMM) with known means (e.g., resulting from the N-FINDR algorithm). This model provides more flexibility regarding the observed pixels and the endmembers. In particular, the endmembers are allowed to be further from the observed pixels which is clearly an interesting property for the problem illustrated in Fig. 1. The NCM assumes that the spectrum of a pixel is defined by the following mixture:

$$\boldsymbol{y} = \sum_{r=1}^{R} \boldsymbol{\varepsilon}_r \alpha_r \tag{3}$$

Digital Object Identifier 10.1109/JSTSP.2009.2038212

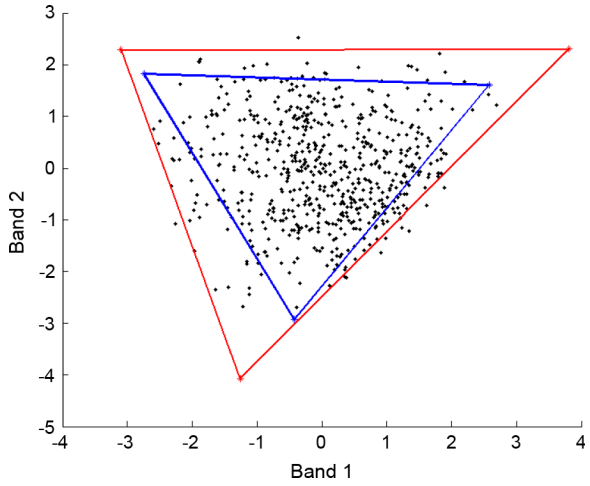

Fig. 1. Scatterplot of dual-band correct (red/light) and incorrect (blue/dark) results of the N-FINDR algorithm.

where the $\boldsymbol{\varepsilon}_r$ are independent Gaussian vectors with known means extracted from a spectral library or estimated by the N-FINDR algorithm. Note that there is no additive noise in (3) since the random nature of the endmembers already models some kind of uncertainty regarding the endmembers. This paper assumes that the covariance matrix of each endmember can be written $\sigma^2\boldsymbol{I}_L$, where $\boldsymbol{I}_L$ is the $L \times L$ identity matrix and σ^2 is the endmember variance in any spectral band. More sophisticated models with different variances in the spectral bands could be investigated[1]. For simplicity, this paper assumes a common variance in all spectral bands that has been considered successfully in many studies [10]–[13].

A supervised spectral unmixing strategy has been proposed in [14] when the number of endmembers R participating in the mixture model (3) is known. This paper addresses the important problem of estimating this number of endmembers R and the nature of these endmembers. It is a model selection problem since different values of R provide abundance vectors with different sizes, i.e., different models. The proposed algorithm concentrates on the NCM because of the interesting properties mentioned above. However, it could be used with appropriate modifications to determine the number of component of the LMM (1) (see [15] for more details). We assume that the endmember means in (3) belong to a given spectral library. However, contrary to the approach in [14], the algorithm does not know the nature and the number of endmember means from this library participating in the NCM. The algorithm is referred to as "semi-supervised" to reflect the partial knowledge about the endmembers present in the NCM mixture. The proposed strategy relies on a hierarchical Bayesian model. Appropriate prior distributions are chosen for the NCM abundances to satisfy the positivity and sum-to-one constraints, as in [15]. A vague conjugate inverse gamma distribution is defined for the endmember variance reflecting the lack of knowledge regarding this parameter. A prior distribution for the number of endmembers R is also introduced. The proposed algorithm is hierarchical since it allows one to estimate the hyperparameter of the NCM. A vague prior is assigned to this hyperparameter. The parameters and hyperparameter of the resulting hierarchical Bayesian model are then jointly estimated using the full posterior distribution. Unfortunately the joint posterior distribution for the NCM is too complex to derive closed-form expressions for the minimum mean square error (MMSE) or maximum *a posteriori* (MAP) estimators. Note in particular that the dimension of the abundance vector depends on the unknown number R of endmembers involved in the mixture. Therefore, we propose to use a reversible jump Markov Chain Monte Carlo (MCMC) algorithm [16] to generate samples distributed according to the posterior of interest. This dimension matching strategy proposes moves between parameter spaces with different dimensions. It has been successfully used in many signal and image processing applications including segmentation [17], [18], analysis of musical audio data [19], and spectral analysis [20].

The paper is organized as follows. Section II derives the posterior distribution of the unknown parameter vector resulting from the NCM. Section III studies the reversible jump MCMC sampling strategy that is used to generate samples distributed according to this posterior and to solve the model selection problem. Simulation results conducted on synthetic and real data as well as a comparison with other model selection techniques are presented in Sections IV and V. Conclusions are reported in Section VI.

II. Hierarchical Bayesian Model

This section studies the likelihood and the priors inherent to the proposed NCM for the spectral unmixing of hyperspectral images. A particular attention is devoted to the unknown number R of endmembers, and to the positivity and sum-to-one constraints about the abundances.

A. Likelihood

The NCM assumes that the endmembers $\boldsymbol{\varepsilon}_r$, $r = 1, \ldots, R$, are independent Gaussian vectors whose means are assumed to belong to a known spectral library $\mathcal{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{R_{\max}}\}$ (where $R_{\max}$ is the number of elements in the library and $\boldsymbol{s}_r$ represents the L-spectrum $[s_{r,1}, \ldots, s_{r,L}]^T$ of the rth endmember). Moreover, this paper assumes that the endmember spectrum components are independent from one band to another and Gaussian such that

$$\boldsymbol{\varepsilon}_r | \boldsymbol{m}_r, \sigma^2 \sim \mathcal{N}(\boldsymbol{m}_r, \sigma^2 \boldsymbol{I}_L)$$

where $\boldsymbol{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the mean vector of $\boldsymbol{\varepsilon}_r$, $\sigma^2\boldsymbol{I}_L$ is its covariance matrix, and $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes the Gaussian distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$. However, the number of components R, as well as the spectra involved in the mixture are unknown. Using the NCM definition (3) and the independence between the endmembers, the likelihood of the observed pixel $\boldsymbol{y}$ can be written as

$$f\left(\boldsymbol{y}|\boldsymbol{\theta}^+\right) \propto \frac{1}{\left[\sigma^2 c(\boldsymbol{\alpha}^+)\right]^{L/2}} \exp\left(-\frac{\|\boldsymbol{y} - \mu(\boldsymbol{\alpha}^+)\|^2}{2\sigma^2 c(\boldsymbol{\alpha}^+)}\right) \quad (4)$$

[1]The case of a structured covariance matrix having K different variances has been considered in [8]. The results have been generalized to any colored noise [9].

where $\boldsymbol{\theta}^+ = \{\boldsymbol{\alpha}^+, \sigma^2, R, M\}$, $||\boldsymbol{x}|| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}$ is the standard ℓ^2 norm, $\boldsymbol{\alpha}^+ = [\alpha_1, \ldots, \alpha_R]^T$, $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R]$ contains the endmember defining the mixture (3), and

$$\mu\left(\boldsymbol{\alpha}^+\right) = \sum_{r=1}^{R} \boldsymbol{m}_r \alpha_r, \quad c\left(\boldsymbol{\alpha}^+\right) = \sum_{r=1}^{R} \alpha_r^2. \tag{5}$$

Note that the mean and variance of this Gaussian distribution depend both on the abundance vector $\boldsymbol{\alpha}^+$ contrary to the classical LMM. Note also that the dimensions of $\boldsymbol{\alpha}^+$ and $\boldsymbol{M}$ (via the quantities $\mu\left(\boldsymbol{\alpha}^+\right)$ and $c\left(\boldsymbol{\alpha}^+\right)$) depend on the unknown number of endmembers R.

B. Parameter Priors

1) Prior Distributions for the Number and Mean of the Endmembers: A discrete uniform distribution on $[1, \ldots, R_{\max}]$ is chosen as prior distribution for the number of mixture components R

$$P(R = k) = \frac{1}{R_{\max}}, \quad k = 1, \ldots, R_{\max} \tag{6}$$

where $R_{\max}$ is the maximum number of pure elements that can be present in a pixel. Note that this uniform prior distribution does not favor any model order among $[1, \ldots, R_{\max}]$.

All combinations of R spectra belonging to the library $\mathcal{S}$ are assumed to be equiprobable conditionally upon the number of endmembers R leading to

$$\begin{aligned} P(\boldsymbol{M} = [\boldsymbol{s}_{i_1}, \ldots, \boldsymbol{s}_{i_R}] \,|R) &= \binom{R_{\max}}{R}^{-1} \\ &= \frac{\Gamma(R+1)\Gamma(R_{\max} - R + 1)}{\Gamma(R_{\max} + 1)} \end{aligned} \tag{7}$$

where $i_1, \ldots, i_R$ is an unordered sequence of R distinct elements of $[1, \ldots, R_{\max}]$.

2) Abundance Prior: Because of the sum-to-one constraint inherent to the mixing model, the abundance vector can be split into two parts as $\boldsymbol{\alpha}^+ = \left[\boldsymbol{\alpha}^T, \alpha_R\right]^T$ with $\alpha_R = 1 - \sum_{r=1}^{R-1} \alpha_r$. To satisfy the positivity constraint, the abundance vector $\boldsymbol{\alpha}$ has to belong to a simplex defined by

$$\mathbb{S} = \left\{ \boldsymbol{\alpha} \;\middle|\; \alpha_r \geq 0,\ \forall r = 1, \ldots, R-1, \sum_{r=1}^{R-1} \alpha_r \leq 1 \right\}. \tag{8}$$

A uniform distribution on this simplex is chosen as prior distribution for the partial abundance vector $\boldsymbol{\alpha}$

$$f(\boldsymbol{\alpha}|R) \propto \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}) \tag{9}$$

where $\propto$ means "proportional to" and $\mathbf{1}_{\mathbb{S}}(.)$ is the indicator function defined on the set $\mathbb{S}$. This prior ensures the positivity and sum-to-one constraints of the abundance coefficients and reflects the absence of other prior knowledge about this parameter. Note that any abundance could be removed from $\boldsymbol{\alpha}^+$ and not only the last one α_R. For symmetry reasons, the algorithm proposed in the next section will remove one abundance coefficient from $\boldsymbol{\alpha}^+$ uniformly drawn in $\{1, \ldots, R\}$. Here, this component is supposed to be α_R to simplify notations. Moreover, for sake of conciseness, the notations $\boldsymbol{\mu}(\boldsymbol{\alpha})$ and $c(\boldsymbol{\alpha})$ will be used in the sequel to denote the quantities in (5) where α_R has been replaced by $1 - \sum_{r=1}^{R-1} \alpha_r$.

3) Endmember Variance Prior: The prior distribution for the variance σ^2 is chosen as a conjugate inverse gamma distribution

$$\sigma^2|\delta \sim \mathcal{IG}(\nu, \delta) \tag{10}$$

where ν and δ are the shape and scale parameters (see [21, p. 581] for definition). This paper classically assumes $\nu = 1$ (as in [17]) and estimates δ within a hierarchical Bayesian framework. Such approach require to define a prior distribution for the hyperparameter δ. This paper assumes that the prior of δ is a non-informative Jeffreys' prior [22] defined by

$$f(\delta) \propto \frac{1}{\delta} \mathbf{1}_{\mathbb{R}^+}(\delta). \tag{11}$$

This prior reflects the lack of knowledge regarding the hyperparameter δ.

C. Posterior Distribution

The posterior distribution of the unknown parameter vector $\boldsymbol{\theta} = \left\{\boldsymbol{\alpha}, \sigma^2, \boldsymbol{M}, R\right\}$ can be easily computed after marginalizing over the hyperparameter δ

$$\begin{aligned} f\left(\boldsymbol{\theta}|\boldsymbol{y}\right) \propto \int f\left(\boldsymbol{y}|\boldsymbol{\alpha}, \sigma^2, \boldsymbol{M}, R\right) f\left(\boldsymbol{\alpha}|R\right) P\left(\boldsymbol{M}|R\right) \\ \ldots P\left(R\right) f\left(\sigma^2|\delta\right) f\left(\delta\right) d\delta. \end{aligned} \tag{12}$$

Straightforward computations lead to the following posterior distribution:

$$\begin{aligned} f(\boldsymbol{\theta}\,|\boldsymbol{y}) \propto \frac{\Gamma(R+1)\Gamma(R_{\max} - R + 1)}{R_{\max}\Gamma(R_{\max} + 1)} \\ \times \frac{1}{\sigma^{2R+L}\left[c(\boldsymbol{\alpha})\right]^{L/2}} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2\sigma^2 c(\boldsymbol{\alpha})}\right) \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}). \end{aligned} \tag{13}$$

The posterior distribution (13) is too complex to derive the MMSE or MAP estimators of $\boldsymbol{\theta}$. In particular, the dimension of $\boldsymbol{\theta}$ is unknown and depends on the number of endmembers R. In such case, it is usual to resort to MCMC methods to generate samples distributed according to the posterior distribution and to use these samples to approximate the MMSE or MAP estimators. To handle the model selection problem resulting from the unknown value of R, we use a reversible jump MCMC method that generate samples according to $f(\boldsymbol{\theta}\,|\boldsymbol{y})$ defined in (3).

III. Reversible Jump MCMC Algorithm

A stochastic simulation algorithm will be employed to sample according to $f(\boldsymbol{\theta}|\boldsymbol{y})$. The vectors to be sampled belong to a space whose dimension depends on R, requiring to use a dimension matching strategy as in [23], [24]. More precisely, the proposed algorithm, detailed in Algorithm 1, consists of four different moves:

1) updating the endmember means contained in $\boldsymbol{M}$;
2) updating the abundance vector $\boldsymbol{\alpha}$;
3) updating the variance σ^2;
4) updating the hyperparameter δ.

These steps are scanned systematically as in [23] and are detailed next.

Algorithm 1: Hybrid Metropolis-within-Gibbs sampler for semi-supervised hyperspectral image unmixing according to the NCM

• Initialization:
  — Sample parameter $R^{(0)}$,
  — Choose $R^{(0)}$ spectra in the library $\mathcal{S}$ to build $\boldsymbol{M}^{(0)}$,
  — Sample parameters $\sigma^{2(0)}$ and $\boldsymbol{\alpha}^{(0)}$,
  — Set $t \longleftarrow 1$,
• Iterations: For $t = 1, 2, \ldots$, do
  — Update $\boldsymbol{M}^{(t)}$:
    draw $u_1 \sim \mathcal{U}_{[0,1]}$,
    if $u_1 \leq b_{R^{(t-1)}}$ **then**
      propose a *birth* move (see Algorithm 2),
    else if $b_{R^{(t-1)}} < u_1 \leq b_{R^{(t-1)}} + d_{R^{(t-1)}}$ **then**
      propose a *death* move see Algorithm 3),
    else if $u_1 > b_{R^{(t-1)}} + d_{R^{(t-1)}}$ **then**
      propose a *switch* move see Algorithm 4),
    else if
    draw $u_2 \sim \mathcal{U}_{[0,1]}$,
    if $u_2 < \rho$ [see Eq. (18)] **then**
      set $(\boldsymbol{\alpha}^{(t)}, \boldsymbol{M}^{(t)}, R^{(t)}) = (\boldsymbol{\alpha}^{*}, \boldsymbol{M}^{*}, R^{*})$,
    else
      set $(\boldsymbol{\alpha}^{(t)}, \boldsymbol{M}^{(t)}, R^{(t)}) = (\boldsymbol{\alpha}^{(t-1)}, \boldsymbol{M}^{(t-1)}, R^{(t-1)})$,
    end if,
  — Sample $\boldsymbol{\alpha}^{(t)}$ from the pdf in Eq. (20),
  — Sample $\sigma^{2(t)}$ from the pdf in Eq. (22),
  — Sample $\delta^{(t)}$ from the pdf in Eq. (23),
  — Set $t \longleftarrow t + 1$.

A. Updating the Endmember Matrix $\boldsymbol{M}$

Three types of moves, referred to as "birth," "death," and "switch" (as in [15]), enable one to update the endmember means involved in the mixture. The two first moves consist of increasing or decreasing the number of pure components R by 1. Therefore, they require the use of the reversible jump MCMC method introduced by Green [16]. In the third move, the dimension of R is not changed, requiring the use of a standard Metropolis–Hastings (MH) acceptance procedure. Assume that at iteration t, the current model is defined by $(\boldsymbol{\alpha}^{(t)}, \boldsymbol{M}^{(t)}, R^{(t)})$. The "birth," "death," and "switch" moves are defined as follows (the step-by-step algorithms are, respectively, presented in Algorithm 2, Algorithm 3, and Algorithm 4).

- birth: a *birth* move $R^{*} = R^{(t)} + 1$ is proposed with the probability $b_{R^{(t)}}$. A new spectrum $\boldsymbol{s}^{*}$ is randomly chosen among the available endmembers means of the library $\mathcal{S}$ to build $\boldsymbol{M}^{*} = [\boldsymbol{M}^{(t)}, \boldsymbol{s}^{*}]$. As the number of pure components has increased by 1, a new abundance coefficient vector $\boldsymbol{\alpha}^{+*}$ is proposed according to a rule inspired by [23]:
  — draw a new abundance coefficient w^{*} from the beta distribution $\mathcal{B}e\left(1, R^{(t)}\right)$;
  — re-scale the existing weights so that all weights sum to 1, using $\alpha_r^{*} = \alpha_r^{(t)}(1 - w^{*}), r = 1, \ldots, R^{(t)}$;
  — build $\boldsymbol{\alpha}^{+*} = \left[\alpha_1^{*}, \ldots, \alpha_{R^{(t)}}^{*}, w^{*}\right]^{T}$.

Algorithm 2: Birth Move

1: set $R^{*} = R^{(t)} + 1$,
2: choose $\boldsymbol{s}^{*}$ in $\mathcal{S}$ such as $\boldsymbol{s}^{*} \neq \boldsymbol{m}_r^{(t)},\ r = 1, \ldots, R^{(t)}$,
3: set $\boldsymbol{M}^{*} = \left[\boldsymbol{m}_1^{(t)}, \ldots, \boldsymbol{m}_{R^{(t)}}^{(t)}, \boldsymbol{s}^{*}\right]$,
4: draw $w^{*} \sim \mathcal{B}e(1, R^{(t)})$,
5: add w^{*} to $\boldsymbol{\alpha}^{+(t)}$ and re-scale the others coefficients, i.e.,

$$\boldsymbol{\alpha}^{+*} = \left[\frac{\alpha_1^{(t)}}{C}, \ldots, \frac{\alpha_{R^{(t)}}^{(t)}}{C}, w^{*}\right]^{T}$$

with $C = 1/(1 - w^{*})$.

- death: a *death* move $R^{*} = R^{(t)} - 1$ is proposed with the probability $d_{R^{(t)}}$. One spectrum of $\boldsymbol{M}^{(t)}$ is removed, as well as the corresponding abundance. The remaining abundances are re-scaled to sum to 1.

Algorithm 3: Death Move

1: set $R^{*} = R^{(t)} - 1$,
2: draw $j \sim \mathcal{U}_{\{1,\ldots,R^{(t)}\}}$,
3: remove $\boldsymbol{m}_j^{(t)}$ from $\boldsymbol{M}^{(t)}$, i.e., set
$\boldsymbol{M}^{*} = [\boldsymbol{m}_1^{(t)}, \ldots, \boldsymbol{m}_{(j-1)}^{(t)}, \boldsymbol{m}_{(j+1)}^{(t)}, \ldots, \boldsymbol{m}_{R^{(t)}}^{(t)}]$
4: remove $\alpha_j^{(t)}$ from $\boldsymbol{\alpha}^{+(t)}$ and re-scale the remaining coefficients, i.e., set

$$\boldsymbol{\alpha}^{+*} = \left[\frac{\alpha_1^{(t)}}{C}, \ldots, \frac{\alpha_{(j-1)}^{(t)}}{C}, \frac{\alpha_{(j+1)}^{(t)}}{C}, \ldots, \frac{\alpha_{R^{(t)}}^{(t)}}{C}\right]^{T}$$

with $C = \sum_{r \neq j} \alpha_r^{(t)}$.

- switch: a *switch* move is proposed with the probability $u_{R^{(t)}}$. A spectrum randomly chosen in $\boldsymbol{M}^{(t)}$ is replaced by another spectrum randomly chosen in the library $\mathcal{S}$.

Algorithm 4: Switch Move

1: **if** $R_{\max} - R^{(t)} \neq 0$ **then**
2: draw $j \sim \mathcal{U}_{\{1,\ldots,R^{(t)}\}}$,
3: choose $\boldsymbol{s}^{*}$ in $\mathcal{S}$ such as $\boldsymbol{s}^{*} \neq \boldsymbol{m}_r^{(t)},\ r = 1, \ldots, R^{(t)}$,
4: replace $\boldsymbol{m}_j^{(t)}$ in $\boldsymbol{M}^{(t)}$ by $\boldsymbol{s}^{*}$, i.e., set
$\boldsymbol{M}^{*} = \left[\boldsymbol{m}_1^{(t)}, \ldots, \boldsymbol{m}_{(j-1)}^{(t)}, \boldsymbol{s}^{*}, \boldsymbol{m}_{(j+1)}^{(t)}, \ldots, \boldsymbol{m}_{R^{(t)}}^{(t)}\right]$
5: set $\boldsymbol{\alpha}^{+*} = \boldsymbol{\alpha}^{+(t)}$ and $R^{*} = R^{(t)}$.
6: **end if**

At each iteration, one of these moves is randomly chosen with probabilities $b_{R^{(t)}}$, $d_{R^{(t)}}$ and $u_{R^{(t)}}$. These probabilities follow three conditions[2]:

- $b_{R^{(t)}} + d_{R^{(t)}} + u_{R^{(t)}} = 1$;
- $d_1 = 0$ (a *death* move is not allowed for $R = 1$);
- $b_{R_{\max}} = 0$ (a *birth* move is impossible for $R = R_{\max}$).

[2]The case $R = 1$ has been accepted for the semi-supervised algorithm as a pixel can be spectrally pure.

As a result, $b_{R^{(t)}} = d_{R^{(t)}} = u_{R^{(t)}} = 1/3$ for $R \in \{2, \ldots, R_{\max} - 1\}$ and $b_1 = d_{R_{\max}} = u_1 = u_{R_{\max}} = 1/2$.

The abundance coefficient vector $\boldsymbol{\alpha}$ has a uniform prior distribution on the simplex defined in (8), which is equivalent to choosing a Dirichlet distribution $\mathcal{D}_R(1, \ldots, 1)$ as prior for the full abundance vector $\boldsymbol{\alpha}^+$. Therefore, the acceptance probability for the *birth* move is $\rho = \min\{1, A_b\}$ with

$$A_b = \left[\frac{c(\boldsymbol{\alpha}^{(t)})}{c(\boldsymbol{\alpha}^*)}\right]^{L/2} \frac{d_{R^{(t)}+1}}{b_{R^{(t)}}} \frac{R^{(t)}}{g_{1,R^{(t)}}(w^*)}(1-w^*)^{R^{(t)}-1} \times \exp\left[\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^{(t)})\|^2}{2c(\boldsymbol{\alpha}^{(t)})} - \frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^*)\|^2}{2c(\boldsymbol{\alpha}^*)}\right] \quad (14)$$

where $g_{a,b}$ denotes the probability density function (pdf) of the beta distribution $\mathcal{B}e(a, b)$ (see the Appendix).

Regarding the acceptance probability $\rho = \min\{1, A_d\}$ for the *death* move, two cases have to be considered. When the current model order is such that $R^{(t)} \neq 2$, the acceptance probability is

$$A_d = \exp\left[\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^{(t)})\|^2}{2c(\boldsymbol{\alpha}^{(t)})} - \frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^*)\|^2}{2c(\boldsymbol{\alpha}^*)}\right] \times \left[\frac{c(\boldsymbol{\alpha}^{(t)})}{c(\boldsymbol{\alpha}^*)}\right]^{L/2} \frac{b_{R^{(t)}-1}}{d_R} \frac{g_{1,R^{(t)}-1}(\alpha^*_{R^{(t)}})}{R^{(t)} - 1} \times (1 - \alpha^*_{R^{(t)}})^{R^{(t)}-2}. \quad (15)$$

When $R^{(t)} = 2$, the death move yields $R^* = 1$, i.e., one single pure element is present in the proposed model, which leads obviously to $\boldsymbol{\alpha}^* = 1$. Thus, $\boldsymbol{\alpha}^*$ is deterministic, and the acceptance rate is

$$A_d = \frac{f(\boldsymbol{M}^*, R^*|\boldsymbol{y})}{f\left(\boldsymbol{M}^{(t)}, R^{(t)}|\boldsymbol{y}\right)} \frac{p_{R^* \longrightarrow R^{(t)}}}{p_{R^{(t)} \longrightarrow R^*}} \frac{q\left(\boldsymbol{M}^{(t)}|\boldsymbol{M}^*\right)}{q\left(\boldsymbol{M}^*|\boldsymbol{M}^{(t)}\right)}. \quad (16)$$

In this case, the Jacobian equals to 1 as $\boldsymbol{\alpha}^*$ is independent from $\boldsymbol{\alpha}^{(t)}$. Consequently, the previous equation yields

$$A_d = \left[\frac{c(\boldsymbol{\alpha}^{(t)})}{c(\boldsymbol{\alpha}^*)}\right]^{L/2} \frac{b_{R^{(t)}-1}}{d_R^{(t)}} \times \exp\left[\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^{(t)})\|^2}{2c(\boldsymbol{\alpha}^{(t)})} - \frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^*)\|^2}{2c(\boldsymbol{\alpha}^*)}\right]. \quad (17)$$

Concerning the *switch* move, the acceptance probability is the standard MH ratio $\rho = \min\{1, A_s\}$ with

$$A_s = \left[\frac{c(\boldsymbol{\alpha}^{(t)})}{c(\boldsymbol{\alpha}^*)}\right]^{L/2} \exp\left[\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^{(t)})\|^2}{2c(\boldsymbol{\alpha}^{(t)})} - \frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^*)\|^2}{2c(\boldsymbol{\alpha}^*)}\right]. \quad (18)$$

Note that the proposal ratio associated to the *switch* move is 1, since in each direction the probability of selecting one spectrum from the library is $1/(R_{\max} - R^{(t)})$. Once the endmember matrix $\boldsymbol{M}$ has been obtained, the abundances, endmember variances, and hyperparameter δ are generated following the procedures detailed below.

B. Generating Samples According to $f(\boldsymbol{\alpha}|\boldsymbol{y}, R, \sigma^2, \boldsymbol{M})$

The Bayes' theorem yields

$$f(\boldsymbol{\alpha}|\boldsymbol{y}, R, \sigma^2, \boldsymbol{M}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\alpha}|R) \quad (19)$$

which easily leads to

$$f(\boldsymbol{\alpha}|\boldsymbol{y}, R, \sigma^2, \boldsymbol{M}) \propto \frac{1}{[\sigma^2 c(\boldsymbol{\alpha})]^{L/2}} \times \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2\sigma^2 c(\boldsymbol{\alpha})}\right) \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}). \quad (20)$$

Thus, the conditional distribution of $\boldsymbol{\alpha}$ is defined on the simplex $\mathbb{S}$ ensuring the abundance vector $\boldsymbol{\alpha}^+$ satisfies the positivity and sum-to-one constraints (2) imposed to the model. The generation of $\boldsymbol{\alpha}$ according to (20) can be achieved using a Metropolis-within-Gibbs move with a uniform prior distribution (9) as proposal distribution.

C. Generating Samples According to $f(\sigma^2|\boldsymbol{y}, R, \boldsymbol{\alpha}, \boldsymbol{M}, \delta)$

The conditional distribution of the variance σ^2 can be determined as follows:

$$f\left(\sigma^2|\boldsymbol{y}, R, \boldsymbol{\alpha}, \boldsymbol{M}, \delta\right) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f\left(\sigma^2|\delta\right). \quad (21)$$

Thus, the conditional distribution of the noise variance is

$$\sigma^2|\boldsymbol{y}, R, \boldsymbol{\alpha}, \boldsymbol{M}, \delta \sim \mathcal{IG}\left(\frac{L}{2} + 1, \frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha})\|^2}{2c(\boldsymbol{\alpha})} + \delta\right). \quad (22).$$

D. Generating Samples According to $f(\delta|\sigma^2, R)$

The conditional distribution of δ is

$$\delta|\sigma^2, R \sim \mathcal{G}\left(R, \frac{R}{\sigma^2}\right) \quad (23)$$

where $\mathcal{G}(a, b)$ is the gamma distribution with shape parameter a and scale parameter b [21, p. 581].

IV. Simulation Results

A. Synthetic Pixel

This section studies the accuracy of the semi-supervised algorithm presented in Sections II and III for unmixing a synthetic pixel. We first consider a pixel resulting from the combination of $R = 3$ random endmembers with variance $\sigma^2 = 0.002$ and abundance vector $\boldsymbol{\alpha}^+ = [0.5,\ 0.3,\ 0.2]^T$. The endmember means are the spectra of construction concrete, green grass and micaceous loam. The results have been obtained for $N_{\mathrm{MC}} = 20\,000$ iterations, including $N_{\mathrm{bi}} = 1500$ burn-in iterations, with the algorithm summarized in Algorithm 1. The spectrum library $\mathcal{S}$ used in this simulation, depicted in Fig. 2, contains six elements defined as construction concrete, green grass, micaceous loam, olive green paint, bare red brick, and galvanized steel metal. These spectra have been extracted from the spectral libraries provided with the ENVI software [25]. The first step of the analysis estimates the number of components R. The posterior distribution of R depicted in Fig. 3 is clearly in good agreement with the actual value of R since its maximum is obtained for $R = 3$. The second step of the analysis estimates the *a posteriori* probabilities of all spectrum combinations, conditioned to $R = 3$. For this simulation example, 100% of these sampled spectrum combinations are composed of the first three spectra of the library which are the actual spectra defining the mixture.

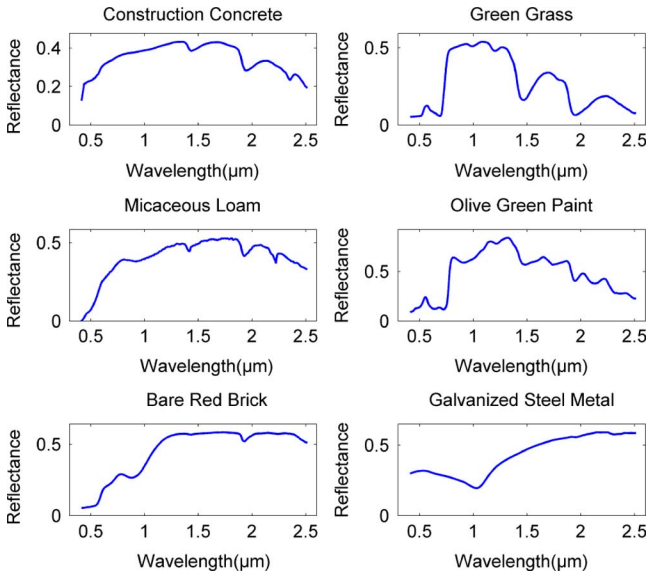

Fig. 2. Endmember spectra of the library.

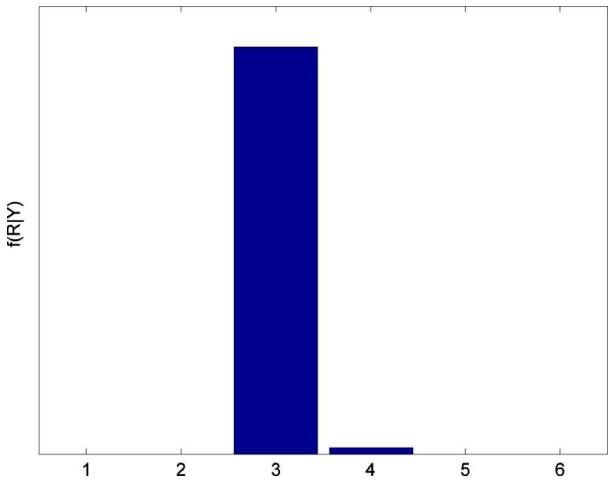

Fig. 3. Posterior distribution of the estimated number of components R.

The posterior distributions of the corresponding abundance coefficients and variance σ^2 are finally estimated and depicted in Figs. 4 and 5. The posteriors are in good agreement with the actual values of these parameters, i.e., $\boldsymbol{\alpha}^+ = [0.5\,0.3\,0.2]^T$ and $\sigma^2 = 0.002$.

B. Comparison With Other Model Selection Strategies

This section compares the Bayesian algorithm developed in this paper with other model selection strategies. First, it is important to note that all model selection methods can be used only when several pixels share the same endmembers, contrary to the proposed algorithm. Moreover, these methods cannot be used to determine which endmembers from the library are actually present in the observed mixture [they just allow one to estimate the number of endmembers R defining (3)]. We suppose here that several synthetic pixels are sharing the same R endmember spectra. The first algorithm studies a maximum-likelihood estimator (MLE) of the intrinsic dimension of the observed data

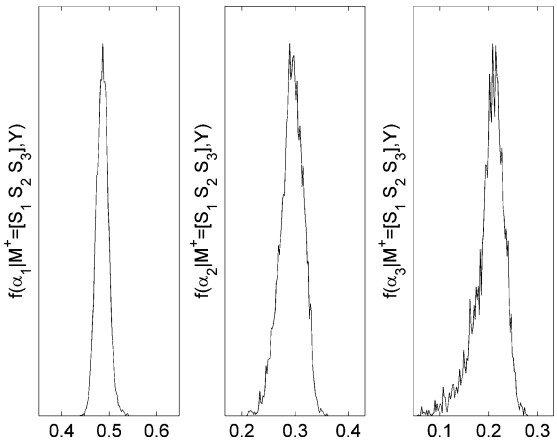

Fig. 4. Posterior distribution of the estimated abundances.

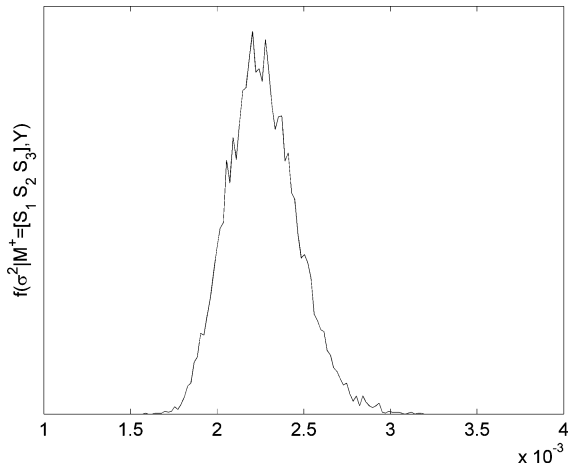

Fig. 5. Posterior distribution of the estimated variance σ^2.

[26], [31]. The next methods mainly rely on the sample covariance matrix of the observed pixels. More precisely, we consider the enhanced versions of the well-known Akaike information criterion (AIC) and minimum description length (MDL initially derived in [32] and [34] and reconsidered in [27] for detecting signals embedded in white noise). The new AIC-based criterion constructed from random matrix theory (RMT) in [28] (referred to as RMT estimator), the Malinowski's F-test advocated in [29] and the PCA-based dimension reduction technique described in [31] (and widely used in the hyperspectral imagery community [4]) are also investigated for comparison. Finally, the hyperspectral signal subspace identification by minimum error (HySime) proposed in [30] is considered for this comparison.

To provide a dimension estimator common to all datasets, several pixels sharing the same R endmember spectra have to be considered. For the algorithms described above, the number of observations (i.e., the number pixels) has to be larger than the observation dimension (i.e., the number of bands). The simulation parameters have been fixed to $L = 207$ bands and $P = 225$ pixels. The $P = 225$ pixels have been simulated according to the NCM in (3) with two different variances ($\sigma^2 = 2.10^{-5}$ and $\sigma^2 = 0.01$) and different numbers of endmember signatures ($R = 3$, $R = 4$, and $R = 5$). For each scenario, the abundance vectors $\boldsymbol{\alpha}^+$ for each pixel have been generated according

TABLE I
ESTIMATED NUMBERS OF PURE COMPONENTS (IN %) OBTAINED BY THE REVERSIBLE-JUMP MCMC METHOD AND OTHER MODEL SELECTION ALGORITHMS WITH $\sigma^2 = 1.10^{-2}$ AND $\sigma^2 = 2.10^{-5}$

	$\sigma^2 = 1.10^{-2}$									$\sigma^2 = 2.10^{-5}$								
	$R=3$			$R=4$			$R=5$			$R=3$			$R=4$			$R=5$		
Estimated R	2	3	4	3	4	5	4	5	6	2	3	4	3	4	5	4	5	6
RJ-MCMC	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0
MLE [26]	0	0	0	0	0	0	0	0	0	0	0	100	0	100	0	100	0	0
AIC [27]	0	100	0	0	95	5	45	55	0	0	100	0	0	99	1	0	95	5
MDL [27]	5	0	0	0	0	0	0	0	0	100	0	0	100	0	0	100	0	0
RMT [28]	32	38	20	1	8	21	0	18	20	30	34	27	0	6	17	0	12	23
F-test [29]	100	0	0	7	0	0	0	0	0	100	0	0	100	0	0	100	0	0
PCA [4]	0	0	0	0	0	0	0	0	0	100	0	0	100	0	0	83	0	0
HySime [30]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 6. Real hyperspectral data: Moffett field acquired by AVIRIS in 1997 (left) and the region of interest shown in true colors (right).

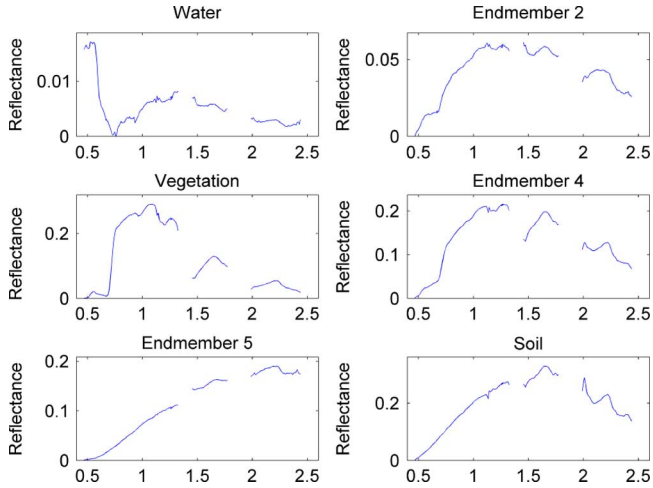

Fig. 7. $R = 6$ endmember spectra built with ENVI.

to Gaussian distributions truncated to the simplex (8) with the following mean vectors:

- 3 endmembers: $\boldsymbol{\alpha}_{R=3} = [0.4,\ 0.25]$;
- 4 endmembers: $\boldsymbol{\alpha}_{R=4} = [0.3,\ 0.15,\ 0.2]$;
- 5 endmembers: $\boldsymbol{\alpha}_{R=5} = [0.3,\ 0.15,\ 0.1,\ 0.1]$.

The threshold default values recommended in [31] have been used for the PCA-based method and the Malinowski's F-test. The estimated number of components R obtained by the proposed reversible-jump MCMC method is compared to the other model selection algorithms in Table I for the different simulation settings. The reversible-jump MCMC method performs significantly better than the other model selection algorithms especially for larger noise variance (i.e., $\sigma^2 = 0.01$). This result can be explained by the fact the proposed Bayesian algorithm fully takes advantage of the NCM model structure, contrary to the other model selection techniques.

V. SPECTRAL UNMIXING OF AN AVIRIS IMAGE

This section considers a real hyperspectral image of size 50×50 depicted in Fig. 6 to evaluate the performance of the different algorithms. This image has been extracted from a larger image acquired in 1997 by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over Moffett Field, CA, and has received much attention in hyperspectral imagery literature (see for instance [35] and [36]). The analyzed area is composed of a lake (top) and a coastal zone (bottom). The data set has been classically reduced from the original 224 bands to $L = 189$ bands by removing water absorption bands.

As detailed in Section III, the reversible jump MCMC algorithm requires the use of a spectral library containing the endmember spectra that might appear in the image to be analyzed. As no ground truth is available for this AVIRIS image, we have built the required library from the image. First, a K-means procedure has been used to classify the image into six regions. Then the purest pixels related to each class, extracted with the pixel purity index algorithm [37], have been chosen as components of the library. This new spectral library whose spectra are represented in Fig. 7 contains six endmembers. Three of them have been clearly identified as follows:

- endmember 1: water;
- endmember 3: vegetation;
- endmember 6: soil.

whereas the endmembers 2, 4, and 5 are more difficult to be interpreted.

The Moffett image is analyzed by the proposed reversible jump MCMC algorithm, using this spectral library. For each pixel, the abundance vector is estimated conditionally upon the endmember matrix. Fig. 8 shows the map of the estimated number of endmembers in the considered area. The lake area includes at least two endmembers defined as "water" and the unidentified "endmember #2." As the water spectrum energy is very low, the algorithm has to balance the measured water

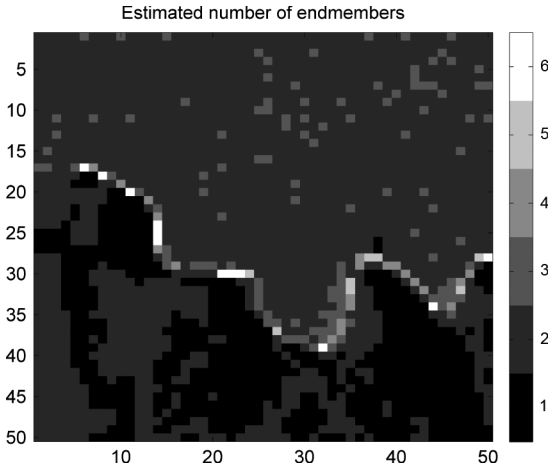

Fig. 8. Number of endmembers estimated by the proposed algorithm (darker (resp. brighter) areas means $R = 1$ [resp. $R = 6$)].

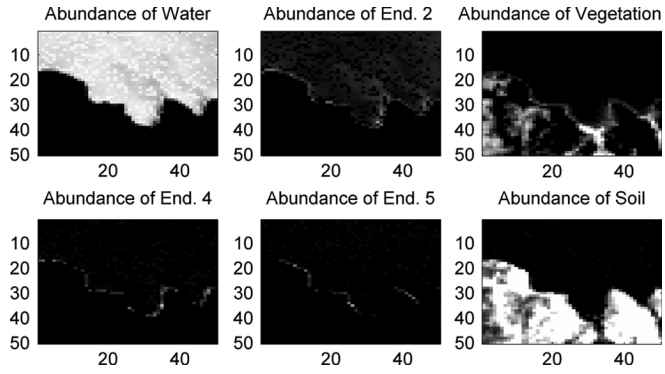

Fig. 9. Fraction maps estimated by the proposed algorithm.

pixel spectrum by adding other contributions, which explains the presence of more than one endmember in this area. The lake-shore area contains the greatest number of endmembers (up to 6), including the unidentified endmembers 4 and 5. The ground area is composed exclusively of soil and vegetation, which is clearly in good agreement with the results obtained for instance in [15] and [14]. The fraction maps for each endmember and the presence maps (with a threshold of 15% for the abundances) obtained by the algorithm are depicted in Figs. 9 and 10, respectively. These figures allow one to appreciate the proportion of endmembers contained in the different regions of the image. To finish, we would like to mention that more results, conducted on the well-known Cuprite data set, can be found in the technical report [38].

VI. Conclusion

A new semi-supervised hierarchical Bayesian unmixing algorithm was derived for hyperspectral images. This algorithm was based on the normal compositional model introduced by Eismann and Stein [7]. The proposed algorithm generated samples distributed according to the joint posterior of the abundances, the endmember variances and one hyperparameter. These samples were then used to estimate the parameters of interest. The proposed algorithm showed several advantages versus the standard model selection strategies. In particular, it allows one to

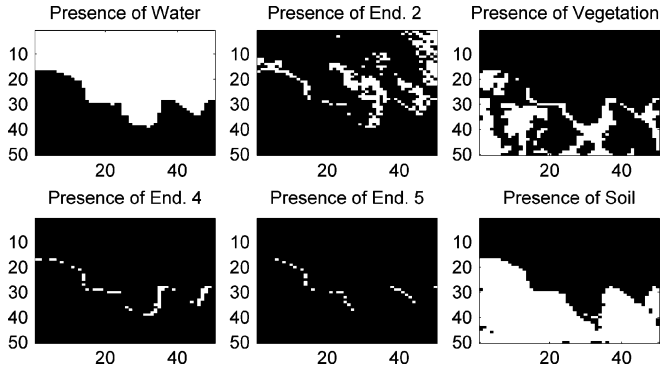

Fig. 10. Presence maps estimated by the proposed algorithm.

estimate which components from a spectral library participate in the mixture for a single given pixel. The simulation results on synthetic and real data showed interesting results.

Perspectives include the extension of the NCM algorithm to more complex scene models. For instance, the hyperspectral images could be considered as a set of homogenous regions, in which common endmembers are present. In this case, spatial correlations for the abundance vectors could be introduced between neighboring pixels to improve unmixing.

Appendix
Acceptance Probabilities for the "Birth" and "Death" Moves

The acceptance probabilities for the "birth" and "death" moves introduced in Section III are derived in this Appendix .

At iteration index t, consider the birth move from the state $\left\{\boldsymbol{\alpha}^{(t)}, \boldsymbol{M}^{(t)}, R^{(t)}\right\}$ to the new state $\{\boldsymbol{\alpha}^*, \boldsymbol{M}^*, R^*\}$ with $\boldsymbol{\alpha}^* = [(1-w^*)\alpha_1, \ldots, (1-w^*)\alpha_{R^{(t)}}]^T$, $\boldsymbol{M}^* = \left[\boldsymbol{M}^{(t)}, \boldsymbol{s}^*\right]$ and $R^* = R^{(t)} + 1$. The acceptance ratio associated to this move is

$$A_b = \frac{f(\boldsymbol{\alpha}^*, \boldsymbol{M}^*, R^*|\boldsymbol{y})}{f(\boldsymbol{\alpha}^{(t)}, \boldsymbol{M}^{(t)}, R^{(t)}|\boldsymbol{y})} \frac{p_{R^* \longrightarrow R^{(t)}}}{p_{R^{(t)} \longrightarrow R^*}} \times \frac{q(\boldsymbol{M}^{(t)}, \boldsymbol{\alpha}^{(t)}|\boldsymbol{M}^*, \boldsymbol{\alpha}^*)}{q(\boldsymbol{M}^*, \boldsymbol{\alpha}^*|\boldsymbol{M}^{(t)}, \boldsymbol{\alpha}^{(t)})} \left|\boldsymbol{J}\left(w^*\right)\right| \quad (24)$$

where $q(\cdot|\cdot)$ refers to the proposal distribution, $\left|\boldsymbol{J}\left(w^*\right)\right|$ is the Jacobian of the transformation and $p._{\longrightarrow}.$ denotes the transition probability, i.e., $p_{R^* \longrightarrow R^{(t)}} = d_{R^*}$ et $p_{R^{(t)} \longrightarrow R^*} = b_{R^{(t)}}$. Using the moves specified in Section III, the proposal ratio is

$$\frac{q(\boldsymbol{M}^{(t)}, \boldsymbol{\alpha}^{(t)}|\boldsymbol{M}^*, \boldsymbol{\alpha}^*)}{q(\boldsymbol{M}^*, \boldsymbol{\alpha}^*|\boldsymbol{M}^{(t)}, \boldsymbol{\alpha}^{(t)})} = \frac{1}{g_{1,R^{(t)}}(w^*)} \frac{R_{\max} - R^{(t)}}{R^{(t)} + 1} \quad (25)$$

where $g_{a,b}(\cdot)$ denotes the pdf of a beta distribution $\mathcal{B}e(a,b)$. Indeed, the probability of choosing a new element in the library ("birth" move) is $1/(R_{\max} - R^{(t)})$ and the probability of removing an element ("death" move) is $1/\left(R^{(t)} + 1\right)$. The Jacobian matrix of the transformation is given as follows

$$\boldsymbol{J}(w^*) = \begin{pmatrix} \frac{\partial \alpha_1^*}{\partial \alpha_1^{(t)}} & \cdots & \frac{\partial \alpha_1^*}{\partial \alpha_{R-1}^{(t)}} & \frac{\partial \alpha_1^*}{\partial w^*} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial \alpha_R^*}{\partial \alpha_1^{(t)}} & \cdots & \frac{\partial \alpha_R^*}{\partial \alpha_{R-1}^{(t)}} & \frac{\partial \alpha_R^*}{\partial w^*} \end{pmatrix}$$

$$= \begin{pmatrix} 1-w^* & 0 & \cdots & 0 & -\alpha_1^{(t)} \\ 0 & 1-w^* & \cdots & 0 & -\alpha_2^{(t)} \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1-w^* & -\alpha_{R-1}^{(t)} \\ -(1-w^*) & -(1-w^*) & \cdots & -(1-w^*) & -\alpha_R^{(t)} \end{pmatrix}$$

which implies

$$|\boldsymbol{J}(w^*)| = (1-w^*)^{R^{(t)}-1}. \quad (26)$$

The posterior ratio is

$$\frac{f(\boldsymbol{\alpha}^*, \boldsymbol{M}^*, R^*|\boldsymbol{y})}{f(\boldsymbol{\alpha}^{(t)}, \boldsymbol{M}^{(t)}, R^{(t)}|\boldsymbol{y})} = \frac{f(\boldsymbol{y}|\boldsymbol{\alpha}^*, \boldsymbol{M}^*, R^*)}{f(\boldsymbol{y}|\boldsymbol{\alpha}^{(t)}, \boldsymbol{M}^{(t)}, R^{(t)})} \times \frac{f(\boldsymbol{\alpha}^*|R^*)}{f(\boldsymbol{\alpha}^{(t)}|R^{(t)})} \frac{f(\boldsymbol{M}^*|R^*)}{f(\boldsymbol{M}^{(t)}|R^{(t)})} \frac{f(R^*)}{f(R^{(t)})}. \quad (27)$$

The prior distribution of the abundance coefficient vector $\boldsymbol{\alpha}^+$ is a Dirichlet distribution $\mathcal{D}_R(1, \ldots, 1)$, which leads to

$$\frac{f(\boldsymbol{\alpha}^*|R^*)}{f(\boldsymbol{\alpha}^{(t)}|R^{(t)})} = \frac{\Gamma(R^{(t)}+1)}{\Gamma(1)\Gamma(R^{(t)})} = R^{(t)}+1. \quad (28)$$

By choosing *a priori* equiprobable configurations for $\boldsymbol{M}$ conditionally upon R, the prior ratio related to the means spectrum matrix is

$$\frac{f(\boldsymbol{M}^*|R^*)}{f(\boldsymbol{M}^{(t)}|R^{(t)})} = \frac{R^{(t)}+1}{R_{\max} - R^{(t)}}. \quad (29)$$

The prior ratio related to the number of mixtures R reduces to 1. Finally, the acceptance ratio for the "birth" move can be written

$$A_b = \exp\left[\frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^{(t)})\|^2}{2c(\boldsymbol{\alpha}^{(t)})} - \frac{\|\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{\alpha}^*)\|^2}{2c(\boldsymbol{\alpha}^*)}\right] \times \left(\frac{c(\boldsymbol{\alpha}^{(t)})}{c(\boldsymbol{\alpha}^*)}\right)^{L/2} \frac{d_{R^{(t)}+1}}{b_{R^{(t)}}} \frac{R^{(t)}}{g_{1,R^{(t)}}(w^*)} (1-w^*)^{R^{(t)}-1}. \quad (30)$$

REFERENCES

[1] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ: Wiley, 2007.

[2] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/ infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.

[3] J. S. Pearlman, P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso, and S. L. Carman, "Hyperion, a space-based imaging spectrometer," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1160–1173, Jun. 2003.

[4] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 50, no. 1, pp. 44–56, Jan. 2002.

[5] M. E. Winter, "Fast autonomous spectral endmember determination in hyperspectral data," in *Proc. 13th Int. Conf. Appl. Geol. Remote Sens.*, Vancouver, BC, Canada, Apr. 1999, vol. 2, pp. 337–344.

[6] J. M. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[7] M. T. Eismann and D. Stein, "Stochastic mixture modeling," in *Hyperspectral Data Exploitation: Theory and Applications*, C.-I. Chang, Ed. New York: Wiley, 2007, ch. 5.

[8] N. Dobigeon and J.-Y. Tourneret, "Bayesian sampling of structured noise covariance matrix for hyperspectral imagery," Univ. of Toulouse, Toulouse, France, 2008 [Online]. Available: http://dobigeon.perso.enseeiht.fr.

[9] N. Dobigeon, J.-Y. Tourneret, and A. O. Hero, III, "Bayesian linear unmixing of hyperspectral images corrupted by colored Gaussian noise with unknown covariance matrix," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 3433–3436.

[10] J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1045–1046, Jul. 1996.

[11] C.-I. Chang, "Further results on relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 1030–1032, May 1998.

[12] 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. Remote Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.

[13] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001.

[14] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model: Application to hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 19, no. 6, Jun. 2010, see also technical report available at http:// eches.perso.enseeiht.fr.

[15] N. Dobigeon, J.-Y. Tourneret, and C.-I. Chang, "Semi-supervised linear spectral using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2696, Jul. 2008.

[16] P. J. Green, "Reversible jump Markov Chain Monte Carlo methods computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.

[17] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 747–758, Mar. 2002.

[18] 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 Process.*, vol. 55, no. 4, pp. 1251–1263, Apr. 2007.

[19] M. Davy, S. Godsill, and J. Idier, "Bayesian analysis of polyphonic western tonal music," *J. Acoust. Soc. Amer.*, vol. 119, no. 4, pp. 2498–2517, Apr. 2006.

[20] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2667–2676, Oct. 1999.

[21] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer-Verlag, 2004.

[22] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proc. R. Soc. London. Ser. A*, vol. 186, no. 1007, pp. 453–461, 1946.

[23] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with an unknown number of components," *J. R. Statist. Soc. Ser. B*, vol. 59, no. 4, pp. 731–792, 1997.

[24] S. Richardson and P. J. Green, "Corrigendum: On Bayesian analysis of mixtures with an unknown number of components," *J. R. Stat. Soc. Ser. B*, vol. 60, no. 3, pp. 661–661, 1998.

[25] "ENVI User's guide Version 4.0,". Boulder, CO, RSI (Research Syst., Inc.), 2003.

[26] E. Levina and P. J. Bickel, "Maximum likelihood estimation of intrinsic dimension," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2005, pp. 777–784.

[27] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.

[28] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, Jul. 2008.

[29] E. R. Malinowski, "Statistical F-tests for abstract factor analysis and target testing," *J. Chemometrics*, vol. 3, no. 1, pp. 49–60, 1988.

[30] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[31] E. Levina, A. S. Wagaman, A. F. Callender, G. S. Mandair, and M. D. Morris, "Estimating the number of pure chemical components in a mixture by maximum likelihood," *J. Chemometrics*, vol. 21, no. 1–2, pp. 24–34, 2007.

[32] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inf. Theory*, 1973, pp. 267–281.
[33] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Contr.*, vol. 19, no. 6, pp. 716–723, Dec. 1974.
[34] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
[35] E. Christophe, D. Léger, and C. Mailhes, "Quality criteria benchmark for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2103–2114, Sep. 2005.
[36] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.
[37] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries 4th Annu. JPL Airborne Geosci. Workshop*, Washington, DC, 1993, vol. 1, pp. 11–14.
[38] 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," Univ. of Toulouse, Toulouse, France, 2009 [Online]. Available: http://eches.perso.enseeiht.fr.

Olivier Eches was born in Villefranche-de-Rouergue, France, in 1984. He received the Eng. degree in electrical engineering from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, both in June 2007. He is currently pursuing the Ph.D. degree at the University of Toulouse (IRIT/INP-ENSEEIHT) on the study of Bayesian algorithms and MCMC methods for the analysis of hyperspectral images.

Nicolas Dobigeon (S'05–M'08) was born in Angoulême, France, in 1981. He received the Eng. degree in electrical engineering from ENSEEIHT, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, both in 2004 and the Ph.D. degree in signal processing from the National Polytechnic Institute of Toulouse in 2007.

From 2007 to 2008, he was a Postdoctoral Research Associate in the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. Since 2008, he has been an Assistant Professor with the National Polytechnic Institute of Toulouse (ENSEEIHT—University of Toulouse), within the Signal and Communication Group of the IRIT Laboratory. His research interests are centered around statistical signal and image processing with a particular interest to Bayesian inference and Markov chain Monte Carlo (MCMC) methods.

Jean-Yves Tourneret (M'94–SM'08) received the Ingénieur degree in electrical engineering from the École Nationale Supérieure d'Électronique, d'Électrotechnique, d'Informatique et d'Hydraulique in Toulouse (ENSEEIHT) in 1989 and the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992.

He is currently a Professor at the University of Toulouse, France (ENSEEIHT), and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal processing with a particular interest to classification and Markov chain Monte Carlo methods.

Dr. Tourneret was the program chair of the European conference on signal processing (EUSIPCO), which was held in Toulouse, France, in 2002. He was also member of the organizing committee for the international conference ICASSP'06 which was held in Toulouse in 2006. He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001–2007, 2010–present). He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.