

# **Unsupervised Unmixing of Hyperspectral Images Accounting for Endmember Variability – Complementary results and supporting materials**

Abderrahim Halimi, Nicolas Dobigeon and Jean-Yves Tournet  
E-mail : {Abderrahim.Halimi,Nicolas.Dobigeon, Jean-Yves.Tournet}@enseeiht.fr

**TECHNICAL REPORT – 2014, December**  
University of Toulouse, IRIT/INP-ENSEEIHT  
2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

## **Abstract**

This paper presents an unsupervised Bayesian algorithm for hyperspectral image unmixing accounting for endmember variability. The pixels are modeled by a linear combination of endmembers weighted by their corresponding abundances. However, the endmembers are assumed random to take into account their variability in the image. An additive noise is also considered in the proposed model generalizing the normal compositional model. The proposed algorithm exploits the whole image to benefit from both spectral and spatial information. It estimates both the mean and the covariance matrix of each endmember in the image. This allows the behavior of each material to be analyzed and its variability to be quantified in the scene. A spatial segmentation is also obtained based on the estimated abundances. In order to estimate the parameters associated with the proposed Bayesian model, we propose to use a Hamiltonian Monte Carlo algorithm. The performance of the resulting unmixing strategy is evaluated via simulations conducted on both synthetic and real data.

## **Index Terms**

Hyperspectral imagery, endmember variability, image classification, spectral unmixing, Bayesian algorithm, Hamiltonian Monte-Carlo, MCMC methods.

## I. INTRODUCTION

Hyperspectral imaging is a remote sensing technology that collects **three** dimensional data cubes composed of 2D spatial images acquired in numerous contiguous spectra bands. Due to the limited spatial resolution of the observed image, each pixel generally consists of several physical elements that are linearly [1], [2] or nonlinearly [3]–[5] mixed. Spectral unmixing (SU) consists of decomposing the pixel spectrum to recover these materials, known as *endmembers*, and estimating the corresponding proportions or *abundances* [6]. The linear mixture model (LMM) has received great interest in the literature and has been used intensively for SU. The unmixing is generally performed using two distinct steps: (i) identifying the endmembers using an endmember extraction algorithm (EEA) such as vertex component analysis (VCA) [7], pixel purity index (PPI) [8] and N-FINDR [9], (ii) estimating the abundances under physical non-negativity and sum-to-one constraints using algorithms such as the fully constrained least squares [2]. Some algorithms also tackle the SU problem in an unsupervised manner, i.e., by jointly estimating the endmembers and the abundances. This is generally achieved under a statistical framework using optimization techniques [10] or Markov chain Monte Carlo (MCMC) simulation methods [6], [11]. The unsupervised algorithms generally provide more sophisticated results and appear to be less sensitive to the absence of pure pixels [3].

The previous described algorithms provide one endmember spectrum for each physical component present in the image (see Fig. 1(a)). This appears as a clear simplification since in many cases, the endmember spectra vary along the image causing what is known as *spectral variability*. Spectral variability has been identified as one of the most profound sources of error in abundance estimation and is knowing growing interest in the hyperspectral community [12], [13]. Many algorithms have been proposed in the literature to describe this **variability**. **A detailed discussion about these algorithms, their advantages and challenges is available in [12], [13]. Most of these methods can be gathered into two main classes. The first approaches consider each physical material as a set or bundle of spectra (see Fig. 1(b)). One can distinguish between algorithms assuming a known spectral library [14], [15] and those estimating it from the data [16], [17]. SU resulting from these approaches is generally sensitive to the quality of the available or extracted endmember libraries. The second class of methods relies on a statistical representation of the endmembers that are assumed to be random vectors with given probability distributions (see**

Fig. 1(c)). These approaches provide a flexible way to incorporate some uncertainties regarding the endmembers [18], [19] and, within a unsupervised context, this choice makes the SU more robust in absence of pure pixels [20]. Two main statistical models of the endmembers have been considered in the literature. The Beta compositional model [21] exploits the physically realistic range of the endmember reflectances by assigning them a Beta distribution. Earlier, the normal compositional model (NCM) was proposed to describe the endmember variability by a Gaussian distribution (see Appendix E for an empirical justification of the Gaussian choice) [18], [20]. An alternative of these parametric models was introduced in [22] for the specific issue of estimating the vegetation fractions in urban environments. In this work, empirical learning of the endmember distributions is conducted from a set of pixels identified as belonging to vegetation and non-vegetation areas. In that sense, it consists of an hybrid method between the two main classes of approaches introduced above, based on a statistical description of the endmember variabilities derived from bundles of spectra.

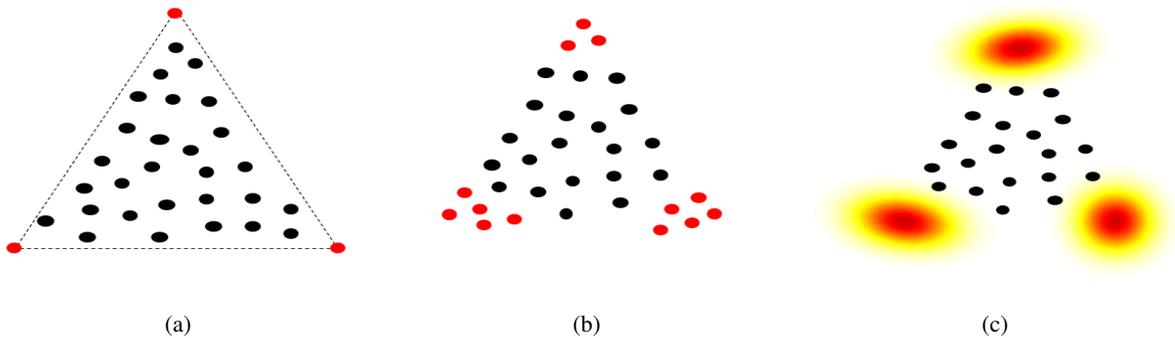


Fig. 1. Simplex representation for (a) endmembers without variability, (b) endmembers as a finite set (or bundle) and (c) endmembers as a distribution.

Adopting a Bayesian perspective, our paper introduces a generalization of this NCM by also considering Gaussian variability for the endmembers (as for the NCM) while incorporating an additive Gaussian noise modeling fitting errors (which was not present in the NCM but in the LMM). Moreover, the proposed model considers a different mean and covariance matrix for each endmember to analyze each component separately. These parameters are both estimated to generalize the works of [18] and [20] that estimated the endmembers means and covariances, respectively. Moreover, the endmember fluctuation with respect to the spectral bands is quantified by considering non-identically distributed endmember variances.

Another important point concerning hyperspectral unmixing is the spatial correlation between pixels. Indeed, even if many algorithms consider a pixel-by-pixel context, recent studies have shown the interest of considering spatial information to improve the unmixing quality [23]–[25]. Within a Bayesian framework, this spatial correlation can be introduced using Markov random fields (MRFs) as already shown in [23], [24], [26]. In this work, a Potts model is considered since it has already shown good performance when processing hyperspectral images [23], [24]. The image is then segmented into regions sharing similar abundance characteristics. Note that this segmentation was also achieved in [24] and [10] by considering Gaussian and Dirichlet distributions for the abundances.

This paper proposes an unsupervised Bayesian algorithm to estimate the parameters associated with endmembers and abundances. In addition to the abundance Dirichlet priors, it assumes appropriate priors for the remaining parameters/hyperparameters to satisfy the known physical constraints. The joint posterior distribution of the proposed Bayesian model is then derived. However, the classical minimum mean square error (MMSE) and maximum a posteriori (MAP) estimators cannot be easily computed from this joint posterior. A classical way of alleviating this problem is to generate samples distributed according to the posterior using MCMC methods. This goal is achieved in this paper using a Gibbs sampler coupled with a Hamiltonian Monte Carlo (HMC) method. HMC is well adapted for large scale problems, i.e., with a large number of parameters to be estimated [27]. Moreover, this method presents good mixing properties when compared to the classical Metropolis-Hasting algorithm. This paper considers a constrained-HMC (CHMC) that has been introduced in [27, Chap. 5] and successfully used for hyperspectral SU in [11]. This CHMC accounts for inequality constraints which is required to satisfy the physical constraints related to the proposed SU problem.

### *Main contributions*

The main objective of the paper is to provide a spectral/spatial algorithm to analyze hyperspectral images accounting for endmember variability. The first contribution of the paper is the generalization of the works [18], [20] by estimating both the mean and the variance of the endmembers. This provides important information such as the sensitivity of each physical material in each spectral wavelength. The second contribution is the generalization of the NCM model by introducing an additive noise that accounts for mismodeling such as non-linearity

effects. These contributions provide a good spectral analysis of the image. In addition to that, and to exploit the hyperspectral image spatial information, we consider an MRF that accounts for spatial correlation between adjacent pixels. Again we consider the Potts model that has shown its efficiency when analyzing hyperspectral images [24], [28]. Moreover, in each spatial class, we consider a Dirichlet distribution as abundance prior which allows the abundance constraints to be satisfied.

The paper is structured as follows. The unmixing problem considered in this study is formulated in Section II. The different components of the proposed Bayesian model are studied in Section III. Section IV introduces the Gibbs sampler and the CHMC method which will be used to generate samples asymptotically distributed according to the joint posterior of the unknown parameters and hyperparameters. Section V analyzes the performance of the proposed algorithm when applied to synthetic images. Results on real hyperspectral images are presented in Section VI whereas conclusions and future works are reported in Section VII.

## II. PROBLEM FORMULATION

The variables used in this paper are described in Table I.

TABLE I  
NOTATION TABLE.

$N$	number of pixels
$R$	number of endmembers
$L$	number of spectral bands
$K$	number of spatial classes
$\mathbf{Y} \in \mathbb{R}^{L \times N}$	spectra of the pixels
$\mathbf{A} \in \mathbb{R}^{R \times N}$	abundance matrix
$\mathbf{M} \in \mathbb{R}^{L \times R}$	endmember means
$\mathbf{\Sigma} \in \mathbb{R}^{R \times L}$	matrix containing the diagonal of endmember covariances
$\mathbf{\Psi} \in \mathbb{R}^{1 \times N}$	noise variances
$\mathbf{z} \in \mathbb{R}^{1 \times N}$	labels
$\mathbf{C} \in \mathbb{R}^{R \times K}$	Dirichlet parameters

### A. Mixing model and endmember variability

This section introduces the proposed mixture model. The classical LMM assumes the pixel spectrum  $\mathbf{y}_n$ ,  $n \in \{1, \dots, N\}$ , where  $N$  is the number of pixels in the image, is a linear combination of  $R$  deterministic endmembers  $\mathbf{s}_r$ ,  $r \in \{1, \dots, R\}$ , corrupted by an additive noise as follows

$$\mathbf{y}_n = \sum_{r=1}^R a_{rn} \mathbf{s}_r + \mathbf{e}_n = \mathbf{S} \mathbf{a}_n + \mathbf{e}_n \quad (1)$$

with

$$\mathbf{e}_n \sim \mathcal{N}(\mathbf{0}_L, \psi_n^2 \mathbf{I}_L) \quad (2)$$

where  $\sim$  means “is distributed according to”,  $R$  is the number of endmembers,  $\mathbf{y}_n$  is an  $(L \times 1)$  vector representing the  $n$ th observed pixel,  $L$  is the number of spectral bands,  $\mathbf{0}_L$  is an  $(L \times 1)$  vector of 0,  $\mathbf{I}_L$  is the  $(L \times L)$  identity matrix,  $\mathbf{a}_n = [a_{1n}, \dots, a_{Rn}]^T$  is the  $(R \times 1)$  abundance vector of the  $n$ th pixel,  $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_R]$  is an  $(L \times R)$  matrix of endmembers and  $\mathbf{e}_n$  is a centered additive, independent and identically distributed Gaussian noise.

The endmembers are generally variable in the observed image due to environmental conditions or inherent variability [12], [13]. The normal compositional model (NCM) has been widely used in the literature to take into account this variability [18], [20], [29] by assuming random endmembers distributed according to Gaussian distributions

$$\mathbf{y}_n = \sum_{r=1}^R a_{rn} \mathbf{s}_{rn} = \mathbf{S}_n \mathbf{a}_n \quad (3)$$

with

$$\mathbf{s}_{rn} \sim \mathcal{N}(\mathbf{m}_r, \text{diag}(\boldsymbol{\sigma}_r^2)) \quad (4)$$

where  $\mathbf{S}_n = [\mathbf{s}_{1n}, \dots, \mathbf{s}_{Rn}]$ ,  $\boldsymbol{\sigma}_r^2 = [\sigma_{r1}^2, \dots, \sigma_{rL}^2]$  is the variance vector of the  $r$ th endmember and  $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_R]$  is the  $(L \times R)$  matrix containing the endmember means of the image.

In this paper, we introduce a new model taking endmember variability into account. More precisely, the proposed model can be seen as a generalization of the NCM model (GNCM) since it introduces an additional residual Gaussian noise  $\mathbf{e}_n$  as follows

$$\mathbf{y}_n = \sum_{r=1}^R a_{rn} \mathbf{s}_{rn} + \mathbf{e}_n = \mathbf{S}_n \mathbf{a}_n + \mathbf{e}_n. \quad (5)$$

This model can also be seen as a generalization of the LMM model used in [6], since the endmember matrix  $\mathbf{S}_n$  depends on each observed pixel in order to introduce the spectral variability while

it is fixed in the LMM. Thus, the GNCM reduces to the LMM for  $\sigma_{r\ell}^2 = 0, \forall \ell, \forall r$ . Considering the GNCM, each physical element  $\#r$  is represented in a given pixel by an endmember  $\mathbf{s}_{rn}$  that has its own Gaussian distribution whose variances  $\sigma_r^2$  change from one band to another. This allows the GNCM to capture the spectral variations of each physical element with respect to each spectral band. The GNCM also includes an additional Gaussian noise  $e_n \sim \mathcal{N}(\mathbf{0}_L, \psi_n^2 \mathbf{I}_L)$  (that is independent from the variables  $\mathbf{s}_{1n}, \dots, \mathbf{s}_{Rn}$ ) whose goal is to make the proposed model more robust with respect to mismodeling. Moreover, we consider that the endmember variability is the main source of randomness in the observed pixel, which is ensured by assigning a prior enforcing small values for the noise variance (see Eq. (17)). Note that the proposed GNCM model reduces to the NCM for  $\psi_n^2 = 0, \forall n$ . Thus, it generalizes the model of [18] by considering a non-isotropic covariance matrix for each endmember. Note finally that the abundance vector  $\mathbf{a}_n$  usually represents spatial coverage of the material in a given pixel. Therefore, it should satisfy the physical positivity and sum-to-one (PSTO) constraints associated with both LMM and GNCM

$$a_{rn} \geq 0, \forall r \in \{1, \dots, R\} \quad \text{and} \quad \sum_{r=1}^R a_{rn} = 1. \quad (6)$$

### III. HIERARCHICAL BAYESIAN MODEL

This section introduces a hierarchical Bayesian model for GNCM-based unsupervised hyperspectral SU accounting for spectral variability. The Bayesian approach assigns prior distributions to the unknown parameters summarizing the prior knowledge about these parameters. This approach is interesting to alleviate the indeterminacy resulting from ill-posed problems and have been successfully applied for HU [6], [30], [31]. More precisely, if  $f(\boldsymbol{\theta})$  denotes the prior distribution assigned to the parameter  $\boldsymbol{\theta}$ , the Bayesian approach computes the posterior distribution of  $\boldsymbol{\theta}$  using the Bayes rule

$$f(\boldsymbol{\theta}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \quad (7)$$

where  $\propto$  means ‘‘proportional to’’ and  $f(\mathbf{Y}|\boldsymbol{\theta})$  is the likelihood of the observation vector  $\mathbf{Y}$ . The vector  $\boldsymbol{\theta}$  is then estimated from this posterior distribution by computing its mean (MMSE estimator) or its maximum (MAP estimator). The following paragraphs introduce the likelihood and the considered prior distributions for  $\boldsymbol{\theta}$ . The unknown parameters of our model include the  $(L \times R)$  endmember mean matrix  $\mathbf{M}$ , the  $(R \times L)$  matrix  $\boldsymbol{\Sigma}$  gathering the endmember variances

(with  $\Sigma_{r,l} = \sigma_{rl}^2$ ), the  $(R \times N)$  abundance matrix  $\mathbf{A}$  (whose  $n$ th column is  $\mathbf{A}_{:n} = \mathbf{a}_n$ ), and the  $(1 \times N)$  vector  $\Psi = [\psi_1, \dots, \psi_N]$ .

### A. Likelihood

Using the observation model (5), the Gaussian properties of both the noise sequence  $e_n$  and the endmembers, and exploiting independence between the observations in different spectral bands, yield **the following likelihood (see Appendix A)**

$$f(\mathbf{y}_n | \mathbf{A}, \mathbf{M}, \Sigma, \Psi) \propto \left( \prod_{\ell=1}^L \Lambda_{\ell n} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \Lambda_{:n}^T [(\mathbf{y}_n - \mathbf{M} \mathbf{a}_n) \odot (\mathbf{y}_n - \mathbf{M} \mathbf{a}_n)] \right\} \quad (8)$$

where  $\Lambda$  is an  $(L \times N)$  matrix whose elements are given by  $\Lambda_{\ell n} = \left( \sum_{r=1}^R a_{rn}^2 \sigma_{r\ell}^2 + \psi_n^2 \right)^{-1}$ ,  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N]$  is an  $(R \times N)$  abundance matrix, and  $\odot$  denotes the Hadamard (termwise) product. Moreover, contrary to the LMM, Eq. (8) shows that the elements<sup>1</sup> of  $\Lambda$  depend on the pixel abundances and **thus** on the pixel index  $\#n$ . This property was also satisfied by the NCM model as previously shown in [18], [20]. Note finally that the joint likelihood of the observation matrix  $\mathbf{Y}$  can be obtained by exploiting independence between the observed pixels

$$f(\mathbf{Y} | \mathbf{A}, \mathbf{M}, \Sigma, \Psi) \propto \prod_{n=1}^N f(\mathbf{y}_n | \mathbf{A}, \mathbf{M}, \Sigma, \Psi). \quad (9)$$

### B. Parameter priors

This section introduces the prior distributions that we have chosen for the parameters of interest.

1) *Classification prior modeling*: Many recent works related to hyperspectral imaging have been considering spatial correlation between the image pixels to partition the image into homogeneous regions with similar abundances [10], [24]. In this paper, we propose to exploit this correlation by dividing the observed image into  $K$  classes sharing the same abundance properties [24]. Each pixel is assigned to a specific class by using a latent label variable  $z_n$  that takes its value into a finite set  $\{1, \dots, K\}$ . The whole set of random variables  $\{z_n\}_{n=1, \dots, N}$  forms a random field. The correlation between neighboring pixels is then introduced by considering a Markov random field prior for  $z_n$  as follows

$$f(z_n | \mathbf{z}_{\setminus n}) = f(z_n | \mathbf{z}_{\nu(n)}) \quad (10)$$

<sup>1</sup>The matrix  $\Lambda$  depends on the noise and endmember variances.

where  $\nu(n)$  denotes the pixel neighborhood as in [24] (a four neighborhood structure will be considered in the rest of the paper),  $\mathbf{z}_{\nu(n)} = \{z_i, i \in \nu(n)\}$  and  $\mathbf{z}_{\setminus n} = \{z_i, i \neq n\}$ . As in [23], [24], [28], this paper considers a Potts-Markov model which is appropriate for hyperspectral image segmentation. The prior of  $\mathbf{z}$  is then obtained using the Hammersley-Clifford theorem

$$f(\mathbf{z}) = \frac{1}{G(\beta)} \exp \left[ \beta \sum_{n=1}^N \sum_{n' \in \nu(n)} \delta(z_n - z_{n'}) \right] \quad (11)$$

where  $\beta > 0$  is the granularity coefficient,  $G(\beta)$  is a normalizing (or partition) constant and  $\delta(\cdot)$  is the Dirac delta function. The parameter  $\beta$  controls the degree of homogeneity of each region in the image. It is assumed known a priori in this paper. However, it could be also included within the Bayesian model and estimated using the strategy described in [32].

2) *Abundance matrix A*: In order to satisfy the constraints (6), the abundance vector should live in the following simplex  $\mathcal{S}$

$$\mathcal{S} = \left\{ \mathbf{a}_n \mid a_{rn} \geq 0, \forall r \text{ and } \sum_{r=1}^R a_{rn} = 1 \right\}. \quad (12)$$

Thus, a natural choice for the prior of  $\mathbf{a}_n$  is a uniform distribution on  $\mathcal{S}$  [5], [33]. However, we want to define a prior enforcing stronger correlations for **spatially** close pixels. Therefore, we propose to assign a Dirichlet prior to the abundances of the  $k$ th class of the image with Dirichlet parameters  $\mathbf{c}_k = (\mathbf{c}_{1k}, \dots, \mathbf{c}_{Rk})^T$  as follows

$$\mathbf{a}_n \mid z_n = k, \mathbf{c}_k \sim \text{Dir}(\mathbf{c}_k), \text{ for } n \in \mathcal{I}_k \quad (13)$$

where  $\text{Dir}(\cdot)$  denotes the Dirichlet distribution, and  $n \in \mathcal{I}_k$  means that  $\mathbf{y}_n$  belongs to the  $k$ th class (which is also equivalent to  $z_n = k$ ). This prior allows the data to be located in several different clusters inside the simplex [10]. **Moreover, the Dirichlet prior is well suited for modeling the abundances since it takes into account the abundance physical PSTO constraints.**

3) *Endmember means*: The endmember mean matrix  $\mathbf{M}$  contains reflectances that should satisfy the following constraints [11]

$$0 < \mathbf{m}_{rl} < 1, \forall r \in \{1, \dots, R\}, \forall l \in \{1, \dots, L\}. \quad (14)$$

Moreover, it makes sense to assume that the reflectances are close to estimates identified by an EEA. Therefore, we choose a truncated Gaussian prior for each endmember as follows [11], [20]

$$\mathbf{m}_r \sim \mathcal{N}_{[0,1]^L}(\widetilde{\mathbf{m}}_r, \epsilon^2 \mathbf{I}_L) \quad (15)$$

where  $\widetilde{\mathbf{m}}_r$  denotes an estimated endmember (resulting from an EEA such as VCA<sup>2</sup>) and  $\epsilon^2$  is a variance term defining the confidence that we have on this estimated endmember  $\widetilde{\mathbf{m}}_r$ .

4) *Endmember variances*: As in [18], [33], the endmember variances have been assigned the following non informative prior

$$f(\boldsymbol{\Sigma}_{:l}) \propto \prod_{r=1}^R \frac{1}{\sigma_{rl}^2} \mathbf{1}_{\mathbb{R}^+}(\sigma_{rl}^2) \quad (16)$$

where we have assumed prior independence between the endmember variances. This distribution introduces some prior knowledge such as the positivity of the endmember variances (the reader is invited to consult [34] for motivations about this prior for scale parameters such as the noise variances).

5) *Noise variance prior*: As stated in [12], endmember variability represents the main source of error in spectral mixture analysis. Moreover, hyperspectral images are generally corrupted by a reduced noise level. Therefore, we assume that the noise effect is smaller than the effect of endmember variability<sup>3</sup>. This can be achieved by choosing an exponential prior

$$f(\psi_n^2 | \lambda) = \lambda \exp(-\lambda \psi_n^2) \mathbf{1}_{\mathbb{R}^+}(\psi_n^2) \quad (17)$$

where  $\lambda$  is a large coefficient imposing sparsity for  $\psi_n^2$  ( $\lambda = 10^7$  in our simulations). We furthermore assume prior independence between the random variables  $\psi_n^2, \forall n \in \{1, \dots, N\}$ . One interest of choosing this prior (17) is that it avoids identifiability problems between the noise and endmember variances (see appendix D for the proof of the problem identifiability). Note that the considered exponential prior includes positivity constraints to the Laplace prior that has been widely used in Bayesian contexts to ensure sparsity [35], [36]. Note also that the estimation of  $\psi_n^2$  can be removed from the proposed Bayesian algorithm without changing significantly the estimation performance (see Section V-D). In particular, the proposed model is sufficiently general since the noise effect can be easily removed by setting to zero the noise variances  $\psi_n^2$ .

<sup>2</sup>We consider in this paper the VCA algorithm even if other algorithms such as N-FINDR [9] and *pixel purity index* (PPI) [8] could also be investigated.

<sup>3</sup>This assumption is no longer satisfied in absence of endmember variability. However, even in this case, we show in Section V that the proposed algorithm provides good results for both abundance and endmember estimates.

### C. Dirichlet parameters

The Dirichlet parameters  $\mathbf{c}_k$  are assigned the following conjugate prior [37]

$$f(\mathbf{c}_k | z_n = k) = \left[ \frac{\Gamma\left(\sum_{r=1}^R c_{rk}\right)}{\prod_{r=1}^R \Gamma(c_{rk})} \right]^\gamma \exp\left(-\alpha \sum_{r=1}^R c_{rk} + R\alpha\right) \prod_{r=1}^R \mathbf{1}_{\mathbb{R}^+}(c_{rk}) \quad (18)$$

where  $\alpha$  and  $\gamma$  are fixed constants that have been chosen to ensure a non-informative prior (flat distribution).

### D. Posterior distribution

The parameters of the proposed Bayesian model are included in the vector  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_p, \boldsymbol{\theta}_h\}$  where  $\boldsymbol{\theta}_p = \{\mathbf{A}, \mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}\}$  (parameters) and  $\boldsymbol{\theta}_h = \{\mathbf{C}, \mathbf{z}\}$  (hyperparameters). This Bayesian model is summarized in the directed acyclic graph (DAG) displayed in Fig. 2.

The joint posterior distribution of the unknown parameter/hyperparameter vector  $\boldsymbol{\theta}$  can be computed from the following hierarchical structure

$$f(\boldsymbol{\theta}_p, \boldsymbol{\theta}_h | \mathbf{Y}) \propto f(\mathbf{Y} | \boldsymbol{\theta}_p, \boldsymbol{\theta}_h) f(\boldsymbol{\theta}_p, \boldsymbol{\theta}_h) \quad (19)$$

where  $f(\mathbf{Y} | \boldsymbol{\theta}_p, \boldsymbol{\theta}_h) = f(\mathbf{Y} | \boldsymbol{\theta}_p)$  has been defined in (9) and  $f(\boldsymbol{\theta}_p, \boldsymbol{\theta}_h)$  is the joint prior of the unknown parameters. Assuming prior independence between the parameters yields

$$\begin{aligned} f(\boldsymbol{\theta}_p, \boldsymbol{\theta}_h) &= f(\boldsymbol{\theta}_p | \boldsymbol{\theta}_h) f(\boldsymbol{\theta}_h) \\ &= f(\mathbf{A} | \mathbf{C}, \mathbf{z}) f(\mathbf{M}) f(\boldsymbol{\Sigma}) f(\boldsymbol{\Psi}) f(\mathbf{C} | \mathbf{z}) f(\mathbf{z}). \end{aligned} \quad (20)$$

The joint posterior distribution  $f(\boldsymbol{\theta}_p, \boldsymbol{\theta}_h | \mathbf{Y})$  can be computed up to a multiplicative constant after replacing (9) and (20) in (19). Unfortunately, it is difficult to obtain closed form expressions for the standard Bayesian estimators associated with (19). In this paper, we propose to use MCMC methods to generate samples asymptotically distributed according to (19) and to build estimators of  $\boldsymbol{\theta}$  from these generated samples. Due to the large number of parameters to be sampled, we use an HMC algorithm which improves the mixing properties of the sampler and reduces the required number of iterations to approximate the target distribution [27]. The parameters are finally estimated using the minimum mean square error (MMSE) estimator for  $\{\mathbf{A}, \mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \mathbf{C}\}$  and the maximum a posteriori (MAP) estimator for the labels  $\mathbf{z}$ . The next section defines the proposed sampling procedure based on a hybrid Gibbs sampler including a CHMC method.

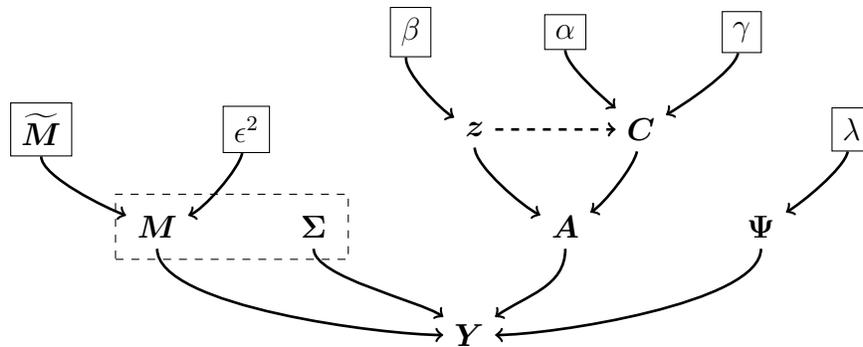


Fig. 2. DAG for the parameter and hyperparameter priors (the fixed parameters appear in boxes). Note that the dashed box defines the statistical distribution of the endmember matrix  $S$ .

#### IV. HYBRID GIBBS ALGORITHM

The principle of the Gibbs sampler is to generate samples according to the conditional distributions of the target distribution (here the posterior (19)) [38]. When a conditional distribution cannot be sampled directly, sampling techniques such as the Metropolis-Hasting (MH) algorithm can be applied. In this paper, we consider HMC as the proposal strategy since it provides better mixing properties than independent or random walk MH moves especially for high-dimensional problems. The next section describes the CHMC algorithm followed by the description of the sampling procedure of the conditional distributions.

##### A. Constrained Hamiltonian Monte Carlo method

HMC is used to sample the high dimensional parameter vector of the proposed Bayesian model. It exploits the gradient of the target distribution to improve the quality of the generated samples. Denoting as  $f(\mathbf{q})$  (resp.  $\mathbf{q}$ ) the distribution (resp.  $d$ -dimensional variable) to be sampled from, HMC defines the Hamiltonian function after introducing a Gaussian momentum variable  $\mathbf{p}$  (that is independent on  $\mathbf{q}$ ) as follows

$$H(\mathbf{p}, \mathbf{q}) = U(\mathbf{q}) + K(\mathbf{p}) \quad (21)$$

where  $U(\mathbf{q}) = -\log[f(\mathbf{q})]$  is the potential energy related to the target distribution  $f(\mathbf{q})$  and  $K(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T\mathbf{p}$  is the momentum energy which results from an independent centered Gaussian distribution for  $\mathbf{p}$  [11]. The evolution of the  $(\mathbf{q}, \mathbf{p})$  samples is determined using the partial derivatives of the Hamiltonian referred to as Hamiltonian equations [27], [39]. For computer

implementations, these equations should be discretized which can be done using the leapfrog method that ensures volume preservation and reversibility of the chains [27], [40]. This leapfrog discretization scheme moves the samples by an  $\epsilon$  stepsize, i.e., from the  $n$ th state  $(\mathbf{q}^n, \mathbf{p}^n)$  to the  $(n + 1)$ th state  $(\mathbf{q}^{(n+1)}, \mathbf{p}^{(n+1)})$  using  $N_L$  iteration steps defined by

$$\mathbf{p}^{(i,n+1/2)} = \mathbf{p}^{(i,n)} - \frac{\epsilon}{2} \frac{\partial U}{\partial \mathbf{q}^T} [\mathbf{q}^{(i,n)}] \quad (22)$$

$$\mathbf{q}^{(i,n+1)} = \mathbf{q}^{(i,n)} + \epsilon \mathbf{p}^{(i,n+1/2)} \quad (23)$$

$$\mathbf{p}^{(i,n+1)} = \mathbf{p}^{(i,n+1/2)} - \frac{\epsilon}{2} \frac{\partial U}{\partial \mathbf{q}^T} [\mathbf{q}^{(i,n+1)}]. \quad (24)$$

The resulting samples are accepted with probability  $\rho$  given by

$$\rho = \min \{1, \exp [H(\mathbf{q}^n, \mathbf{p}^n) - H(\mathbf{q}^{(n+1)}, \mathbf{p}^{(n+1)})]\}. \quad (25)$$

This procedure ensures the resulting samples to be asymptotically distributed according to the target distribution.

In the presence of inequality constraints  $(\mathbf{q}^{(i,n\epsilon)} \in [q_l, q_u])$ , we adopt the procedure presented in [11] and [27, Chap. 5]. This procedure replaces a sample that violates the constraints at each leapfrog iteration by its symmetric to the bound (see [11] for more details). For example, the candidate  $\mathbf{q}^{(i,n)} = q_u + h$  with  $0 < h < (q_u - q_l)$  will be replaced by  $\mathbf{q}^{(i,n)} = q_u - h$  (and similarly  $\mathbf{q}^{(i,n)} = q_l - h$  will be replaced by  $\mathbf{q}^{(i,n)} = q_l + h$ ) when a constraint is not satisfied.

### B. Sampling the parameters/hyperparameters

Sampling according to the joint posterior (19) is achieved by considering a Gibbs sampler that iteratively generates samples distributed according to the conditional distributions. The obtained hybrid Gibbs sampler consists of six steps that are summarized in Algo. 1, where the conditional distributions associated with the parameters/hyperparameters are derived in Appendix B. Note that Algo. 1 generates  $N_{\text{MC}}$  samples for each parameter of interest. However, the MMSE or MAP estimators are computed after removing the first  $N_{\text{bi}}$  samples belonging to the so-called burn-in period (the length of the burn-in period has been determined using appropriate convergence diagnoses [41]). To accelerate the convergence, the abundances  $\mathbf{A}^{(0)}$  and endmember means  $\mathbf{M}^{(0)}$  have been initialized using the results of FCLS and VCA, respectively. The labels have been assigned a random integer in the set  $\{1, \dots, K\}$ . The other parameters have been initialized as

follows  $\sigma_{r\ell}^2 = 10^{-3}$ ,  $\psi_n = 10^{-6}$ ,  $c_{rk} = 1$ ,  $\forall r, \forall \ell, \forall n$  and  $\forall k$ . Note finally that the MCMC approach is robust to local minima and that the results do not depend in the considered initialization. The interested reader is invited to consult [27], [38], [41] for more details about Gibbs sampler and HMC algorithm, including the proofs of convergence of these algorithms.

---

**Algorithm 1** Hybrid Gibbs sampler

---

- 1: Initialization t=0
  - 2: Initialize parameters  $\mathbf{A}^{(0)}$ ,  $\mathbf{M}^{(0)}$ ,  $\Sigma^{(0)}$ ,  $\Psi^{(0)}$ ,  $\mathbf{C}^{(0)}$ , and  $\mathbf{z}^{(0)}$
  - 3: Iterations
  - 4: **for**  $t = 1 : N_{MC}$  **do**
  - 5:   Parameter update
  - 6:   Sample  $\mathbf{T}^{(t)}$  from the pdf (37) using a CHMC procedure
  - 7:   Sample  $\mathbf{M}^{(t)}$  from the pdf (41) using a CHMC procedure
  - 8:   Sample  $\Sigma^{(t)}$  from the pdf (44) using a CHMC procedure
  - 9:   Sample  $\Psi^{(t)}$  from the pdf (49) using a CHMC procedure
  - 10:   Hyperparameter update
  - 11:   Sample  $\mathbf{C}^{(t)}$  from the pdf (52) using a CHMC procedure
  - 12:   Sample  $\mathbf{z}^{(t)}$  from the pdf (48)
  - 13: **end for**
- 

## V. SIMULATION RESULTS ON SYNTHETIC DATA

This section evaluates the performance of the proposed algorithm with synthetic data. It is divided into four parts whose objectives are: 1) introducing the criteria used for the evaluation of the unmixing quality, 2) presenting the different parameters that are estimated in the proposed unmixing approach, 3) analyzing the behavior of the proposed algorithm as a function of the number of endmembers and the size of the image, 4) comparing the proposed strategy with other state-of-the-art algorithms from the literature.

### A. Evaluation criteria

Abundances and endmembers are known for synthetic images. In this case, the quality of the unmixing strategy can be measured by comparing the estimated and actual abundances by using

the average root mean square error (aRMSE) defined by

$$\text{aRMSE}(\mathbf{A}) = \sqrt{\frac{1}{NR} \sum_{n=1}^N \|\mathbf{a}_n - \hat{\mathbf{a}}_n\|^2} \quad (26)$$

where  $\|\cdot\|$  denotes the standard  $l_2$  norm such that  $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$ . The mean of the  $r$ th estimated endmember can be compared with the actual one by using  $\text{RMSE}(\mathbf{m}_r)$  or the spectral angle mapper  $\text{SAM}(\mathbf{m}_r)$  defined as follows

$$\text{RMSE}(\mathbf{m}_r) = \frac{1}{\sqrt{L}} \|\hat{\mathbf{m}}_r - \mathbf{m}_r\|, \quad \text{SAM}(\mathbf{m}_r) = \arccos\left(\frac{\hat{\mathbf{m}}_r^T \mathbf{m}_r}{\|\mathbf{m}_r\| \|\hat{\mathbf{m}}_r\|}\right) \quad (27)$$

where  $\arccos(\cdot)$  is the inverse cosine operator. Moreover, the global endmember error is evaluated by the averaged RMSE (aRMSE) and averaged SAM (aSAM) given by

$$\text{aRMSE}(\mathcal{M}) = \sqrt{\frac{1}{R} \sum_{r=1}^R [\text{RMSE}(\mathbf{m}_r)]^2}, \quad \text{aSAM}(\mathcal{M}) = \frac{1}{R} \sum_{r=1}^R \text{SAM}(\mathbf{m}_r). \quad (28)$$

Note that the RE and SAM criteria can also be evaluated for the  $\#p$ th measured and estimated pixel spectra  $\mathbf{y}_n, \hat{\mathbf{y}}_n$  as follows

$$\text{RE} = \sqrt{\frac{1}{NL} \sum_{n=1}^N \|\hat{\mathbf{y}}_n - \mathbf{y}_n\|^2}, \quad \text{SAM} = \frac{1}{N} \sum_{n=1}^N \arccos\left(\frac{\hat{\mathbf{y}}_n^T \mathbf{y}_n}{\|\mathbf{y}_n\| \|\hat{\mathbf{y}}_n\|}\right). \quad (29)$$

Finally, the Earth movers distance (EMD) criterion (based on the Euclidean distance) have also been considered to simultaneously evaluate the estimated endmembers and abundances [42]. For synthetic data, this criterion compares the estimated parameters to the true ones, while it provides a mutual comparison of the different algorithms for real data (see [42] for more details about EMD).

### B. Performance of the proposed algorithm

This section considers a  $50 \times 50$  synthetic image generated according to (5) with  $R = 3$  physical elements (construction concrete, green grass and micaceous loam) corresponding to spectral signatures available in the ENVI software library [43]. For each pixel, we generate  $R = 3$  endmembers whose means are these ENVI-like spectral signatures and whose variances are band-dependent and represented in Fig. 3 (dashed lines). This image is assumed to be partitioned into  $K = 3$  classes whose label maps have been generated using the Potts model (11) with  $\beta = 1.5$  (see Fig. 4). The abundances corresponding to the pixels belonging to a common class share the

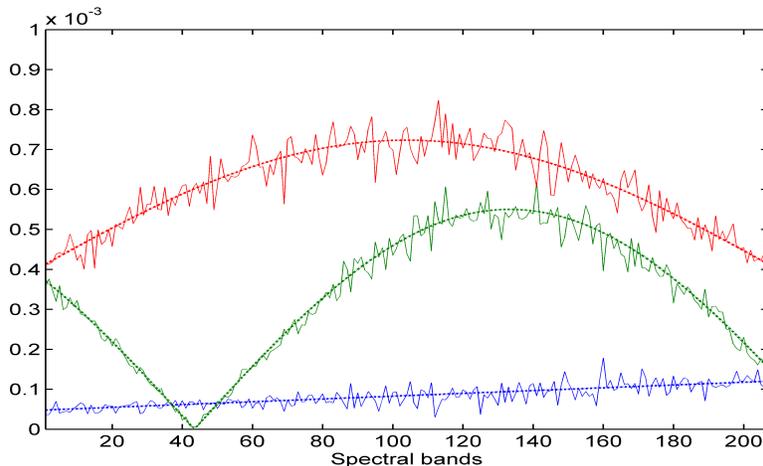


Fig. 3. Actual endmember variances (dashed line) and estimated variances by the proposed UsGNM (continuous line) for the considered  $R = 3$  endmembers.

same Dirichlet parameters (that are reported in Table II) leading to the observed pixels displayed in Fig. 5. Note that the generated abundances have been truncated ( $a_r < 0.9, \forall r$ ) to avoid the presence of pure pixels in the image, **which makes the unmixing problem more challenging**. Finally, we have considered a noise variance equal to  $10^{-7}$  for all pixels (note that the noise variance has to be smaller than the endmember variances). The proposed unsupervised GNCM-based algorithm, denoted by UsGNM, has been run using  $N_{\text{bi}} = 11000$  burn-in iterations and  $N_{\text{MC}} = 12000$  iterations<sup>4</sup>. Fig. 4 (right) displays the estimated classification map obtained

TABLE II  
ACTUAL AND ESTIMATED DIRICHLET PARAMETERS IN EACH SPATIAL CLASS.

	Dirichlet parameters					
	$c_{1k}$	$c_{2k}$	$c_{3k}$	$\hat{c}_{1k}$	$\hat{c}_{2k}$	$\hat{c}_{3k}$
$k = 1$	15	15	1	14.97	14.85	1.00
$k = 2$	1	8	8	1.05	8.24	8.19
$k = 3$	3	1	3	3.12	1.02	3.03

<sup>4</sup> $N_{\text{MC}}$  represents the total number of samples that have been generated. The MMSE or MAP estimators are computed after removing the first  $N_{\text{bi}}$  burn-in iterations. The length of the burn-in period has been determined using appropriate convergence diagnoses [41].

with the proposed algorithm. This map is in a very good agreement with the ground truth shown in Fig. 4 (left). Note that the Dirichlet parameters used in this simulation correspond to three distinguishable classes that are well separated using the proposed algorithm. The obtained classification results can also be observed with the data projected in the plane associated with the two most discriminant principle components as shown in Fig. 5. The proposed algorithm also allows the Dirichlet parameters to be estimated accurately as shown in Table II.

A significant advantage of the proposed algorithm is its ability to estimate the endmember means and variances. Fig. 5 shows the estimated endmembers obtained using the VCA algorithm (diamonds) [7], the Bayesian unsupervised LMM-based unmixing algorithm (UsLMM, circles) [6] and the proposed UsGNM approach (triangles). Contrary to the VCA algorithm that provides bad endmember estimates because of the absence of pure pixels in the image, both UsLMM and UsGNM strategies yield good endmember estimations. As explained before, the good performance of the UsGNM algorithm can be explained by the fact that it is able to mitigate the endmember variability. Fig. 6 displays the endmember means (continuous lines), the endmember distributions (colored areas in Figs. 6(a), (b) and (c)) and the associated variability intervals defined by  $\text{mean} \pm 3\sigma$  (Fig. 6 (d)). Fig. 3 displays the actual and estimated endmember variances for the three endmembers that are clearly in good agreement. These results show the good performance of the proposed approach that fully exploits the spatial (segmentation map, abundances and noise variances) and spectral (endmember means and variances) correlations. The next section studies the robustness of the proposed approach with respect to the number of endmembers and pixels (i.e., image size).

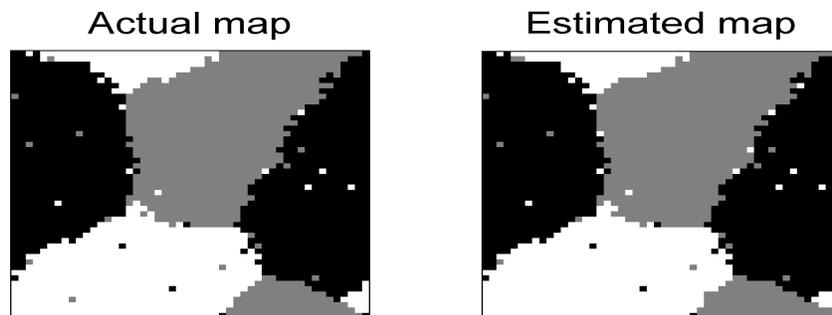


Fig. 4. Actual (left) and estimated (right) classification maps of a synthetic image.

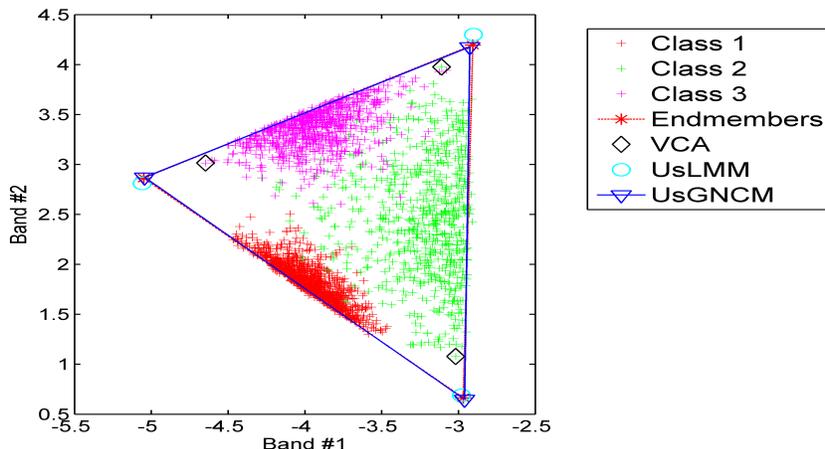


Fig. 5. Classified projected pixels (colored crosses), actual endmembers (red stars), endmembers estimated by VCA (black diamonds), endmembers estimated by UsLMM (cyan circle) and endmembers estimated by UsGNCM (blue triangles).

### C. Performance as a function of the number of endmembers and the image size

The UsGNCM algorithm estimates many parameters which might require a lot of observations in order to obtain acceptable performance. The first part of this section deals with this problem by analyzing the proposed algorithm when varying the number of observed pixels. The considered image has been generated using the three endmember **means and variances considered in Section V-B, the same noise variance,  $K = 1$  spatial class and** abundances uniformly distributed in the truncated simplex  $\mathcal{S}$  (i.e., the abundance are truncated with  $a_i < 0.9, \forall i \in 1, \dots, R$  and the Dirichlet parameters are  $c_{rk} = 1, \forall r, \forall k$ ). Fig. 7 shows the obtained  $\text{aRMSE}(\mathbf{A})$ , RE and SAM when varying the size of the observed image. As expected, the unmixing performance improves by increasing the number of observations. This figure also shows that the  $\text{aRMSE}(\mathbf{A})$  converges to a constant value for  $\sqrt{N} > 50$  while RE and SAM continue to improve when increasing  $N$ . Note, however, that the obtained results are quite good for  $N \geq 100$ . The second part of this section analyzes the behavior of UsGNCM with respect to the number of endmembers. Table III shows the obtained  $\text{aRMSE}(\mathbf{A})$ ,  $\text{aRMSE}(\mathbf{M})$ ,  $\text{aSAM}(\mathbf{M})$ ,  $\text{aRMSE}(\mathbf{\Sigma})$ ,  $\text{aSAM}(\mathbf{\Sigma})$  and **EMD** criteria for  $R = \{3, 4, 5, 6\}$ . The considered endmember **means** are 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 [43].

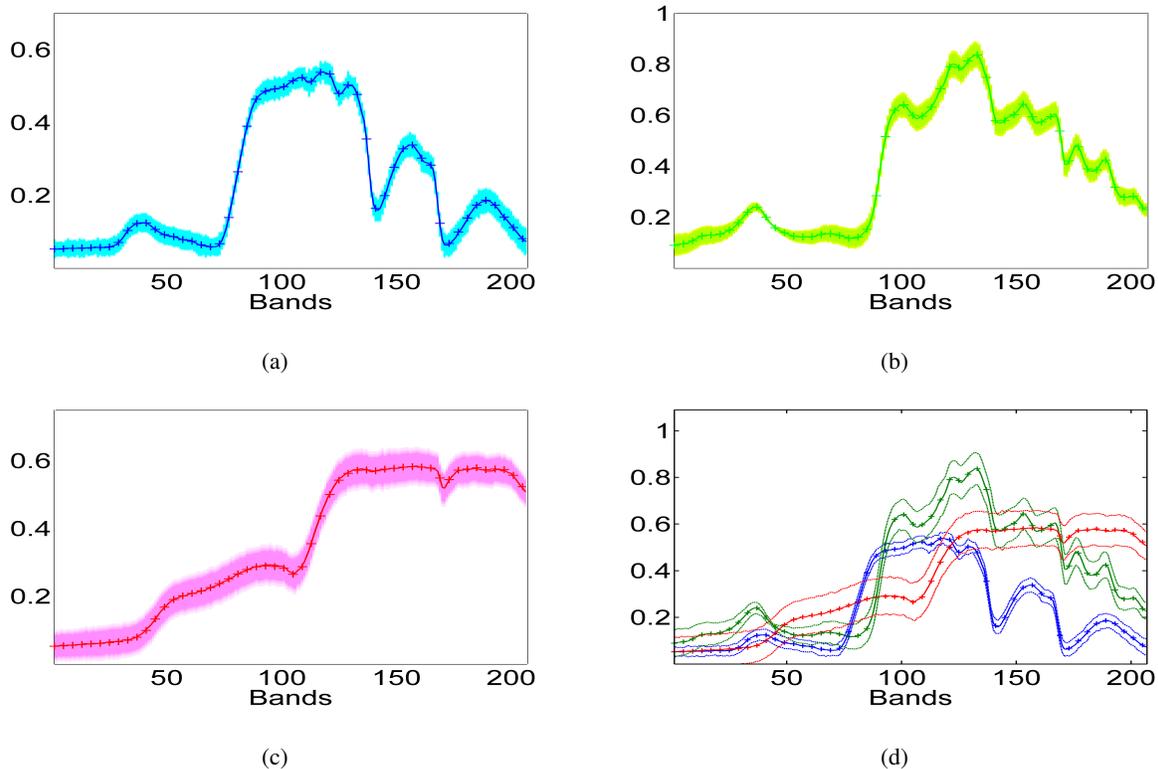


Fig. 6. Actual endmembers (crosses) and endmember means estimated by UsGNM (continuous lines). The estimated endmember distributions are represented in (a), (b), (c) by colored areas. The bottom-right figure (d) shows the endmembers estimated by UsGNM  $\pm 3\sigma$  (dashed lines).

As previously, the images associated with  $R = \{3, 4, 5, 6\}$  have been generated with  $K = 1$  spatial class and abundances uniformly distributed in the truncated simplex  $\mathcal{S}$  with  $a_r < 0.9, \forall r$ . The endmember variances of the first three physical elements are represented in Fig. 3 and we have considered  $(\sigma_4^2, \sigma_5^2, \sigma_6^2) = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$  in this experiment. As expected, increasing the number of endmembers (i.e., increasing  $R$ ) reduces the estimation performance. However, the obtained results are still acceptable confirming the robustness of UsGNM with respect to the number of endmembers  $R$ . Note finally that more simulations, when considering the presence of pure pixels, have also been conducted and showed good performance of UsGNM as shown in Table IV. Moreover, table V shows the obtained results obtained without considering spatial correlation which shows the benefit of this spatial correlation.

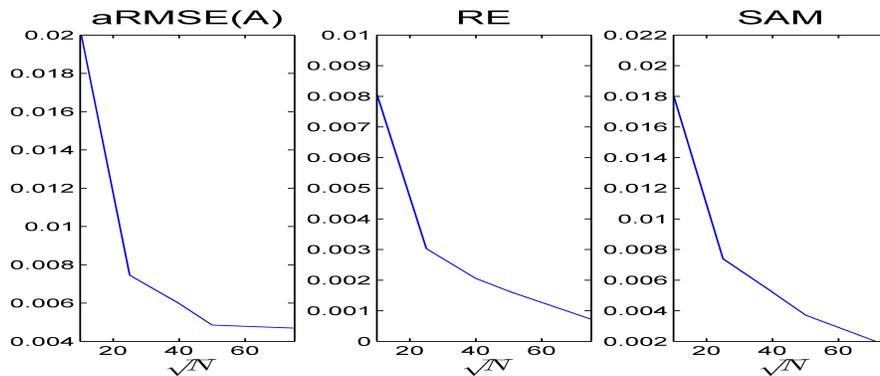


Fig. 7. UsGNM performance for different numbers of pixels.

TABLE III

USGNM PERFORMANCE FOR DIFFERENT NUMBER OF ENDMEMBERS (TRUNCATED SIMPLEX).

	aRMSE( $\mathbf{A}$ ) ( $\times 10^{-2}$ )	aRMSE( $\mathbf{M}$ ) ( $\times 10^{-2}$ )	aSAM( $\mathbf{M}$ ) ( $\times 10^{-2}$ )	aRMSE( $\mathbf{\Sigma}$ ) ( $\times 10^{-4}$ )	aSAM( $\mathbf{\Sigma}$ ) ( $\times 10^{-2}$ )	EMD
$R = 3$	0.48	0.12	0.28	0.23	7.43	152.3
$R = 4$	0.54	0.21	0.51	0.22	9.16	165.1
$R = 5$	0.80	0.28	0.68	0.31	11.7	198.5
$R = 6$	1.40	0.84	1.65	0.52	14.2	478.8

#### D. Comparison with state-of-the-art algorithms

This section evaluates the performance of the proposed UsGNM algorithm for different images. All images have been constructed using  $R = 3$  endmembers with truncated abundances (with  $a_i < 0.9, \forall i \in 1, \dots, R$ ) to avoid the presence of pure pixels, **which makes the unmixing problem more challenging**. The remaining parameters have been defined as follows

- the image  $I_1$  has been generated according to the GNCM model with  $K = 1$  class and abundances uniformly distributed in the simplex  $\mathcal{S}$ . The endmember variances have been adjusted as in Fig. 3. The noise variance is  $\psi_n^2 = 10^{-7}$ .
- the image  $I_2$  is the GNCM image used in Section V-B.
- the image  $I_3$  has been generated according to the LMM model with  $K = 3$  classes, the labels were generated using the Potts model with  $\beta = 1.5$  (the same labels as  $I_2$ ) and the Dirichlet parameters of Table II. The noise variances vary linearly with respect to the spectral bands

TABLE IV

USGNCM PERFORMANCE FOR DIFFERENT NUMBER OF ENDMEMBERS (WITH PURE PIXELS AND SPATIAL CORRELATION).

	aRMSE( $\mathbf{A}$ ) ( $\times 10^{-2}$ )	aRMSE( $\mathbf{M}$ ) ( $\times 10^{-2}$ )	aSAM( $\mathbf{M}$ ) ( $\times 10^{-2}$ )	aRMSE( $\mathbf{\Sigma}$ ) ( $\times 10^{-4}$ )	aSAM( $\mathbf{\Sigma}$ ) ( $\times 10^{-2}$ )
$R = 3$	0.49	0.12	0.29	0.23	7.24
$R = 4$	0.51	0.15	0.36	0.23	8.83
$R = 5$	0.86	0.24	0.53	0.38	11.16
$R = 6$	1.00	0.52	1.30	0.43	13.7

TABLE V

USGNCM PERFORMANCE FOR DIFFERENT NUMBER OF ENDMEMBERS (WITH PURE PIXELS AND NO CORRELATION  $\beta = 0$ ).

	aRMSE( $\mathbf{A}$ ) ( $\times 10^{-2}$ )	aRMSE( $\mathbf{M}$ ) ( $\times 10^{-2}$ )	aSAM( $\mathbf{M}$ ) ( $\times 10^{-2}$ )	aRMSE( $\mathbf{\Sigma}$ ) ( $\times 10^{-4}$ )	aSAM( $\mathbf{\Sigma}$ ) ( $\times 10^{-2}$ )
$R = 3$	0.48	0.14	0.29	0.23	7.44
$R = 4$	0.60	0.26	0.58	0.23	9.11
$R = 5$	0.73	0.27	0.63	0.31	11.44
$R = 6$	1.46	0.95	1.81	0.58	14.01

with

$$\psi_l^2 = 10^{-4} \left( \frac{4}{L-1} l + \frac{L+3}{L-1} \right), \text{ for } l \in [1, \dots, L].$$

These images are processed using different unmixing strategies that are compared to the proposed UsGNCM algorithm. More precisely, we have considered the following unmixing algorithms

- VCA+FCLS: the endmembers are extracted from the whole image using VCA and the abundances are estimated using the FCLS algorithm [2].
- UsLMM: the unsupervised Bayesian algorithm of [6] is used to jointly estimate the endmembers and abundances.
- AEB: this is the automated endmember bundles algorithm proposed in [17]. We consider a 10% image subset and the VCA algorithm to extract the endmembers. For each pixel, the 3 endmembers that provide the smallest RE are selected.
- UsNCM: the proposed unmixing strategy with  $\psi_n = 0$  (i.e., the additive noise  $e_n$  of (5) is removed). Note that the resulting model reduces to the NCM and the corresponding unmix-

ing algorithm can be considered as an unsupervised counterpart of the method introduced in [18].

The first two algorithms provide one estimate for each endmember while the other algorithms estimate endmember variability. Note that the UsNCM is introduced to study the effect of the additive noise. Table VI reports the quality of the estimated abundances and endmembers by unmixing the three images with the different algorithms. This table shows bad performance for VCA+FCLS and AEB algorithms which is mainly due to the absence of pure pixels in the considered images **and to the variation of the endmember/noise variances with respect to the spectral band**. The UsLMM provides good results for the three images. However, it appears to be sensitive to the variation of endmember/noise variances with respect to the spectral band and to the spatial correlations between adjacent pixels. Indeed, the UsLMM did not consider spatial correlation which leads to a performance reduction when processing the images  $I_2$  and  $I_3$ . Note also that the UsLMM algorithm provides one estimate for each endmember and does not take into account the spatial variability of endmembers in the processed images. The best performance is generally obtained by the proposed UsNCM and UsGNCM strategies that provide almost similar results. However, the UsGNCM algorithm is more robust than UsNCM when processing the LMM image  $I_3$ . **Note that when processing  $I_3$ , both UsNCM and UsGNCM consider the effect of the colored noise in the LMM to be due to endmember variability. In fact, this effect was expected since the noise variance in (5) does not depend on spectral bands while the endmember variances do. However, this effect does not affect the performance of UsGNCM that provides the best results in terms of abundance and endmember estimation as highlighted by the criteria aRMSE, aSAM and EMD.** All these results confirm the superiority of the proposed approach in presence of endmember variability, spatial correlation between pixels and in absence of pure pixels in the observed scene.

## VI. SIMULATION RESULTS ON REAL DATA

### A. Description of the Hyperspectral Data

This section illustrates the performance of the proposed UsGNCM algorithm when applied to a real hyperspectral data set. The real image used in this section was acquired in 2010 by the Hypspec hyperspectral scanner over Villelongue, France (00 03'W and 4257'N). The dataset contains  $L = 160$  spectral bands recorded from the visible to near infrared **(400 – 1000nm)**

TABLE VI  
RESULTS ON SYNTHETIC DATA.

		Criteria ( $\times 10^{-2}$ )					EMD ( $\times 10^3$ )
		aRMSE ( $\mathbf{A}$ )	aRMSE ( $\mathbf{M}$ )	aSAM ( $\mathbf{M}$ )	aRMSE ( $\mathbf{\Sigma}$ )	aSAM ( $\mathbf{\Sigma}$ )	
image $I_1$ (GNMCM, $K = 1$ )	VCA+FCLS	4.78	2.20	4.53	-	-	1.31
	UsLMM	0.52	0.18	0.43	-	-	<b>0.15</b>
	AEB	3.73	2.25	4.90	-	-	1.46
	UsNCM	<b>0.48</b>	0.14	0.31	0.23	7.4	0.17
	UsGNMCM	<b>0.48</b>	<b>0.12</b>	<b>0.28</b>	0.23	7.4	<b>0.15</b>
image $I_2$ (GNMCM, $K = 3$ )	VCA+FCLS	3.71	2.68	6.74	-	-	1.39
	UsLMM	0.76	0.49	0.94	-	-	0.27
	AEB	9.46	4.20	8.72	-	-	2.54
	UsNCM	0.56	0.19	0.43	0.27	10	0.16
	UsGNMCM	<b>0.48</b>	<b>0.16</b>	<b>0.41</b>	0.26	10	<b>0.15</b>
image $I_3$ (LMM, $K = 3$ )	VCA+FCLS	9.51	4.42	8.51	-	-	2.59
	UsLMM	1.01	0.49	1.22	-	-	0.36
	AEB	9.30	5.13	10.92	-	-	2.88
	UsNCM	0.86	0.48	1.15	-	-	0.30
	UsGNMCM	<b>0.74</b>	<b>0.34</b>	<b>0.74</b>	-	-	<b>0.26</b>

with a spatial resolution of 0.5m [44]. It has already been studied in [11], [44] and is mainly composed of forested and urban areas. The proposed unmixing algorithm has been applied to two subimages: scene #1 of size  $100 \times 100$  which is composed of  $R = 4$  components: tree, grass, soil and shadow (see Fig. 8 (right)), and scene #2 of size  $31 \times 31$  which is composed of  $R = 3$  components: grass, road and ditch (see Fig. 8 (left)). In addition to the previous studied algorithms, the UsGNMCM has been also compared to the supervised spectral/spatial BCM based algorithms proposed in [21] and that are denoted by BCM-MH and BCM-QP when considering Metropolis-Hastings (MH) sampler and quadratic programming (QP), respectively. For both BCM algorithms, we have selected the pure pixels manually resulting in a set of spectra for each endmember, as suggested in [21].

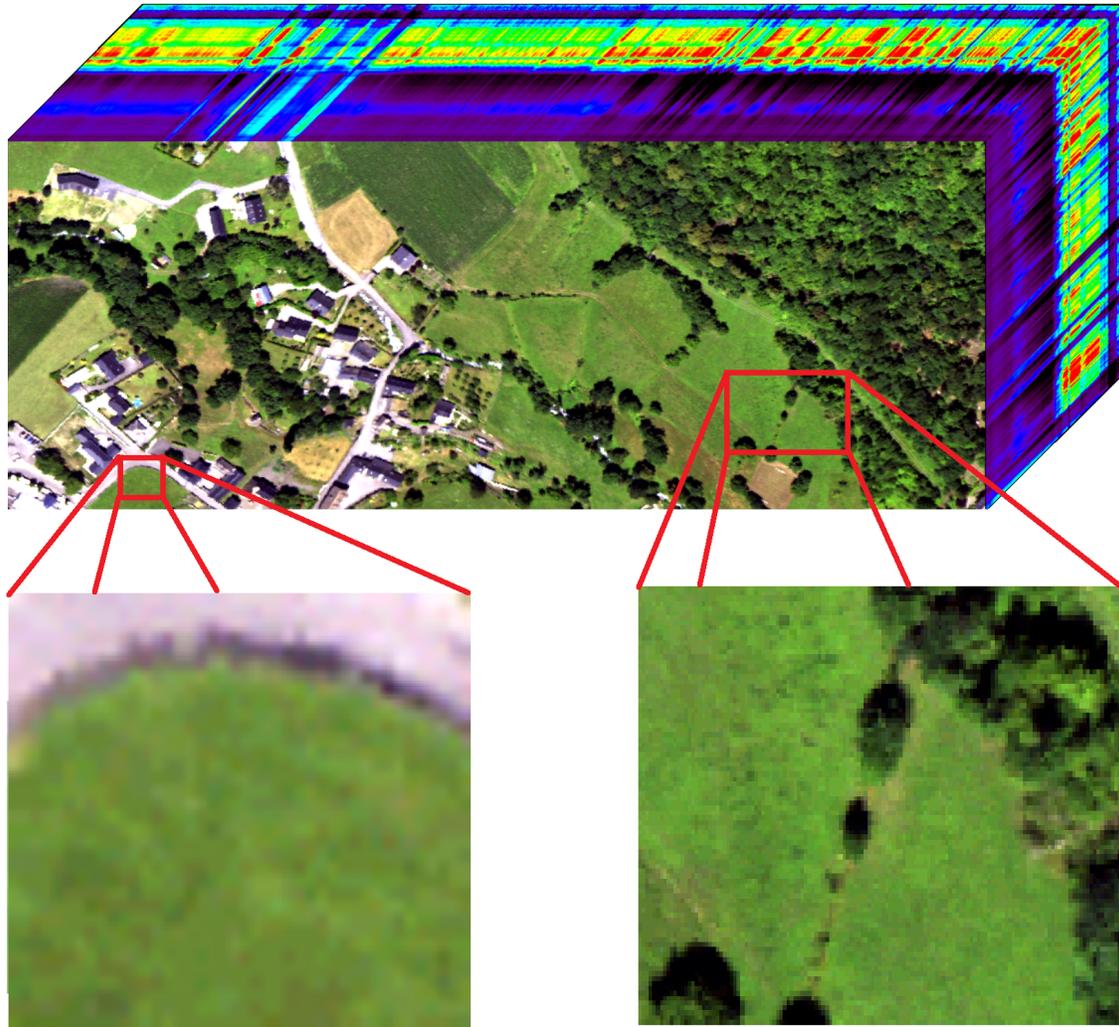


Fig. 8. Real Madonna image and the considered subimages shown in true colors. (Right) scene 1, (left) scene 2

### *B. Endmember and variability estimation*

The proposed UsGNM algorithm can estimate both the endmember means and variances. Fig. 9 compares the endmember estimates of this algorithm with those obtained with VCA, UsLMM and AEB when considering scene #1. The estimated endmembers are globally in good agreement especially for UsGNM and UsLMM. Note that VCA (resp. AEB) provides a different shadow endmember because it extracts the endmember as the purest pixel in the image (resp. each sub-image) while UsLMM and UsGNM estimate both the abundances and

endmembers resulting in a better shadow estimate (lower amplitude). Moreover, the proposed algorithm provides endmember distributions (blue level areas in Fig. 9) which measure the endmember variability in the considered image. **The difference between the estimated UsGNCM interval and the AEB spectra is mainly due to the fact that AEB selects its spectra from the image pixels while the UsGNCM is not limited by this constraint.** It can be seen from Fig. 9 that the higher relative variation is obtained for the shadow spectrum because of its low amplitude. Moreover, the variation is more pronounced for high spectral bands ( $l > 80$ ) which is in agreement with the results presented in [11]. Fig. 10 shows the obtained endmembers when considering scene #2. This figure presents similar results between UsGNCM and UsLMM, especially for capturing spectral components having low amplitudes as for ditch.

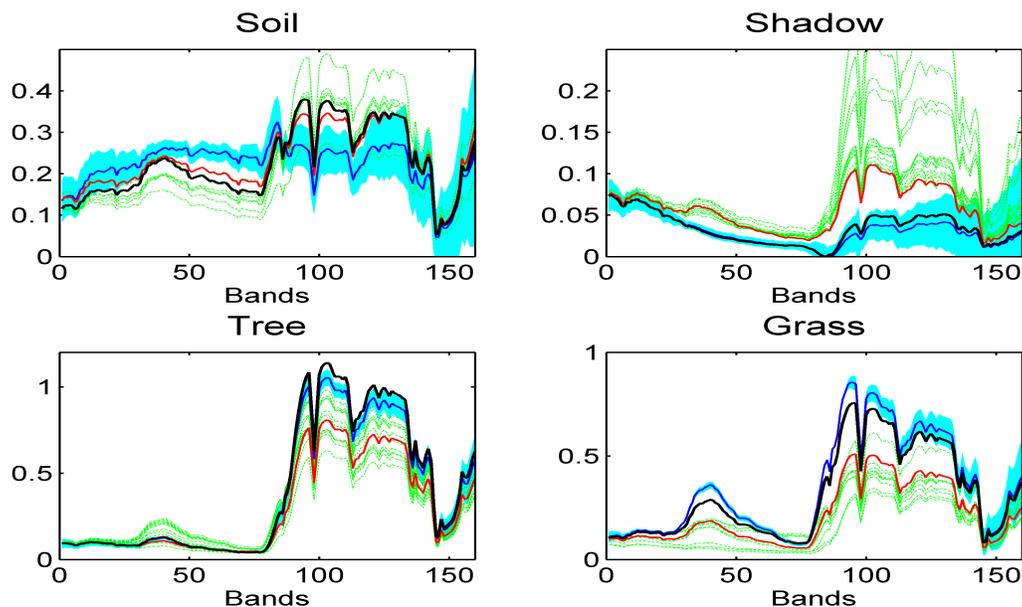


Fig. 9. The  $R = 4$  endmembers estimated by VCA (continuous red lines), UsLMM (continuous black lines), UsGNCM (continuous blue lines) and the estimated endmember distribution (blue level areas) for scene #1 of the Madonna image.

### C. Abundance Estimation and Image Classification

The fraction maps of scene #1 estimated by the studied methods are shown in Fig. 12. Note that a white (black) pixel indicates a large (small) proportion of the corresponding materials.

**These maps lead to the following conclusions**

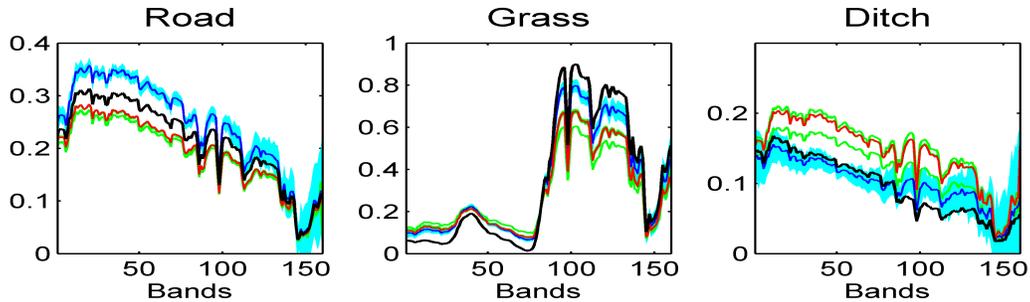


Fig. 10. The  $R = 3$  endmembers estimated by VCA (continuous red lines), UsLMM (continuous black lines), UsGNCM (continuous blue lines) and the estimated endmember distribution (blue level areas) for scene #2 of the Madonna image.

- UsLMM and UsGNCM present similar abundance estimates with a smoother behavior for the second algorithm (because of spatial correlation)
- The abundance maps of VCA-FLCS, AEB, BCM-QP and BCM-MH are higher than those of statistical methods (UsLMM and UsGNCM) especially for the shadow and Tree. This is due to the assumption of presence of pure pixel which is not always true as for the shadow (the shadow pixels in the image are tree-shadowed pixels, thus, they can not be considered as pure shadow pixels)
- AEB (resp. BCM-MH) is sensitive to the similarity between tree and grass spectra (resp. soil and grass spectra) leading to bad grass maps.

Considering scene #2, the compared algorithms provide similar abundance maps with a lower abundances for UsLMM and UsGNCM. These results are not presented here for brevity (see [45]). The ground truth is not available for these real image, thus, we adopted the same procedure as in [13], [21] to quantitatively compare the abundance results (this procedure can be seen as a comparison tool that detects the similarity between the algorithm results). The abundance RMSEs are evaluated when considering the median abundance of all algorithms as a reference. Table VII shows the obtained results for the two images. These results show similar figures for UsLMM and UsGNCM which quantitatively confirm the previous conclusions. The EMD can also be considered to mutually compare the studied six algorithms. The obtained results are shown in Fig. 11 where for each image, the element of the  $i$ th row and  $j$ th column represents the comparison between the  $i$ th and  $j$ th algorithms. For example, the diagonal represents the smallest distance (black color) since we compare each algorithm with itself, while a white color

represents a high distance. These images confirm the previous conclusions and highlight the similarity between the results of UsGNCM and UsLMM.

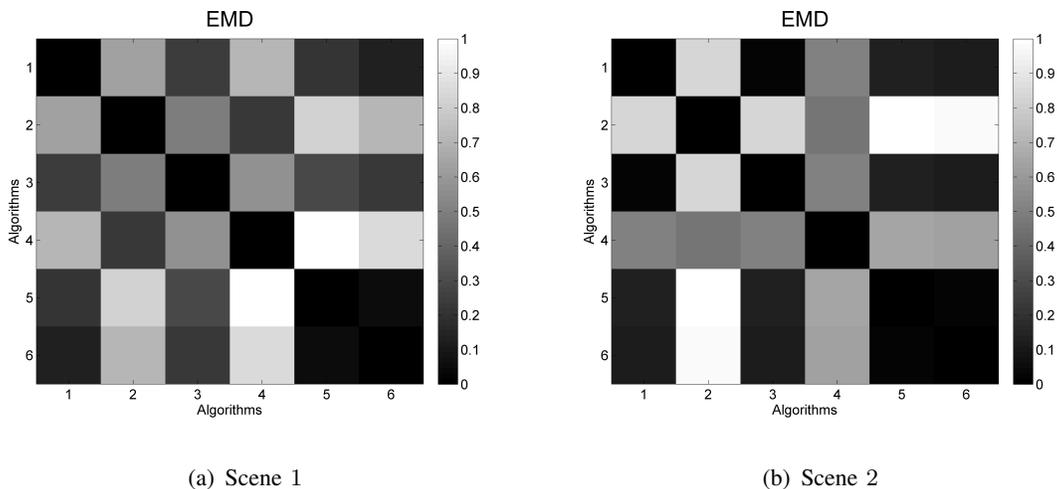


Fig. 11. Normalized EMD when comparing the studied six algorithms for (a) scene #1 of Madonna, (b) scene #2 of Madonna. The algorithms order is: VCA+FCLS, UsLMM, AEB, UsGNCM, BCM-QP and BCM-MH.

In addition to unmixing, UsGNCM also provides a spatial segmentation of the considered scenes as shown in Fig. 13(a) for scene #1 and Fig. 14(a) for scene #2. These classifications clearly highlight the area of each physical element in the scene. Indeed, for scene #1 we have 5 classes that represent tree, soil, shadow, and grass (2 classes) zones while for scene #2 we have 3 classes representing road, ditch and grass areas. Table VIII finally reports the estimated Dirichlet parameters and the number of pixels for each spatial class when considering scene #1. These parameters suggest a highly non uniform distribution over the simplex which promote the use of the proposed approach.

#### D. Reconstruction errors

This section compares the proposed UsGNCM strategy to state-of-the-art algorithms by considering the reconstructed signal. Table VII shows the obtained RE and SAM for the studied strategies<sup>5</sup>. As expected, the AEB algorithm, whose objective is to minimize the RE criteria, shows better results when compared to the other approaches for scene #1. However, both

<sup>5</sup>The estimated endmember mean is considered for UsGNCM. The endmember mean of each set is considered for the BCM approaches.

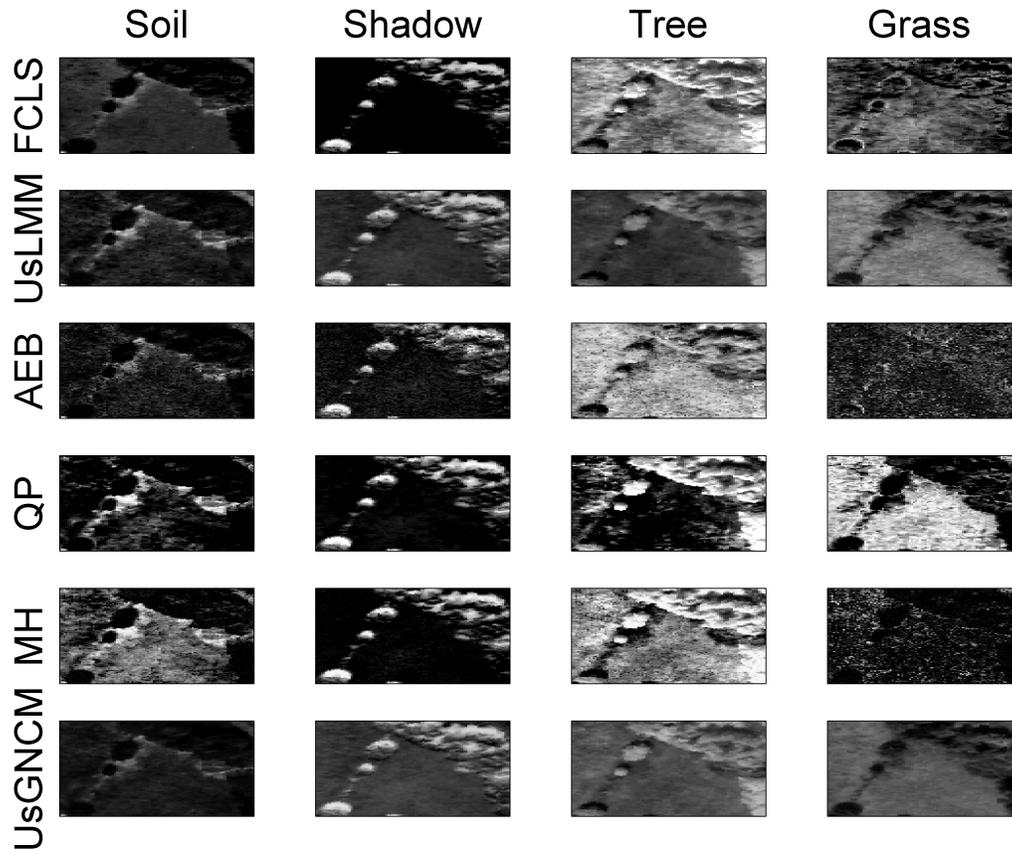


Fig. 12. Abundance maps estimated by FCLS (first row), UsLMM (second row), AEB (third row), BCM-QP (fourth row), BCM-MH (fifth row) and the proposed UsGNCM (sixth row) for the Madonna image.

UsLMM and UsGNCM present good results especially for scene #2 where they outperform AEB. Indeed, UsLMM and UsGNCM are statistical algorithms which estimate endmember that are not necessarily present in the image, which makes them more flexible.

### E. Residual Components

The proposed algorithm also provides a measure of the noise variance for each observed pixel. This parameter brings an information about pixels that are inaccurately described by a linear formulation, i.e., allows modeling errors to be quantified. Fig. 13(b) shows the obtained noise variances for the scene #1. This figure shows a higher error in the shadow area and around trees, i.e., for regions where possible interactions between physical components might occur (e.g.,

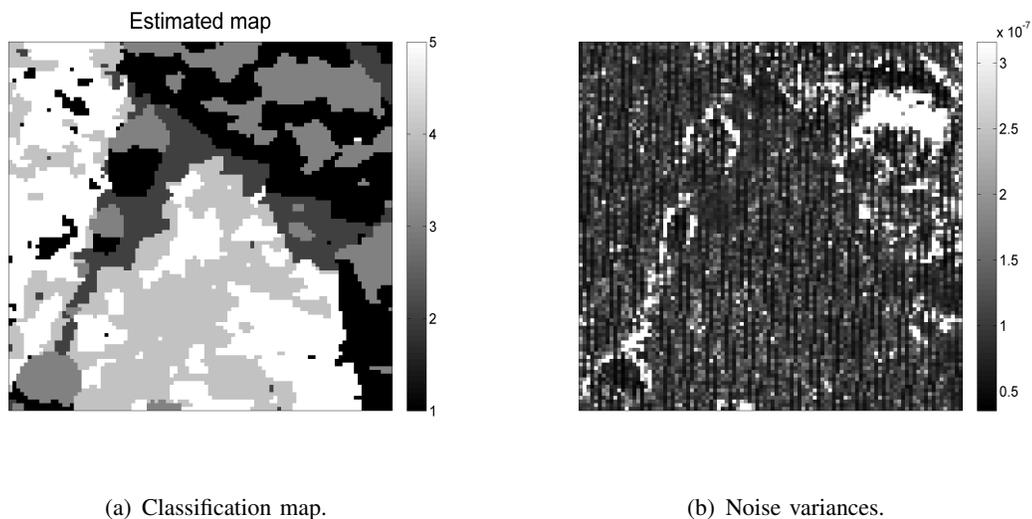


Fig. 13. Estimated maps with the UsGNCM algorithm for the scene #1 of Madonna image. (a) Classification map and (b) noise variances.

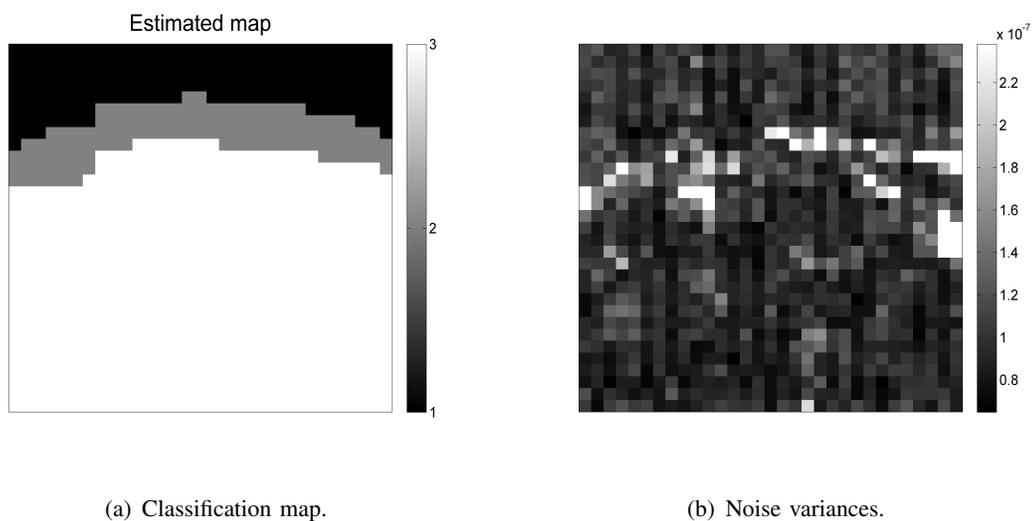


Fig. 14. Estimated maps with the UsGNCM algorithm for the scene #2 of Madonna image. (a) Classification map and (b) noise variances.

tree/soil) resulting in a more complex model than the proposed linear one. The noise variances associated with the scene #2 are shown in Fig. 14(b). This figure shows a higher error near the ditch area which might be due to the presence of nonlinearities as explained in [11]. Note finally that both Fig. 13(b) and Fig. 14(b) highlight the presence of regular vertical patterns that have also been observed in [46] and were associated with a sensor defect or other miscalibration problems.

TABLE VII  
PERFORMANCE ON REAL IMAGE.

Methods	Scene 1			Scene 2		
	RMSE ( $\times 10^{-2}$ )	RE ( $\times 10^{-3}$ )	SAM ( $\times 10^{-2}$ )	RMSE ( $\times 10^{-2}$ )	RE ( $\times 10^{-3}$ )	SAM ( $\times 10^{-2}$ )
VCA+FCLS	11.3	16.9	5.6	3.3	11	3.4
UsLMM	15.7	6.3	2.7	20	8	2.9
AEB	17.9	6.1	2.6	3.5	9	3.0
UsGNM	13.7	6.4	2.7	16	8	2.9
BCM-QP	19.7	12.4	3.5	2.9	13	3.3
BCM-MH	15.2	14.7	4.5	2.8	14	3.4

TABLE VIII  
ESTIMATED DIRICHLET PARAMETERS FOR THE MADONNA IMAGE (SCENE 1).

	Dirichlet parameters				number of pixels
	$\hat{c}_{1k}$	$\hat{c}_{2k}$	$\hat{c}_{3k}$	$\hat{c}_{4k}$	
$k = 1$	1.47	4.59	10.98	4.39	2144
$k = 2$	13.26	14.05	15.96	14.22	1064
$k = 3$	0.76	7.97	3.75	1.36	1502
$k = 4$	37.71	76.11	84.06	99.97	2483
$k = 5$	23.04	57.70	89.82	99.93	2807

## VII. CONCLUSIONS

This paper introduced a Bayesian model for unsupervised unmixing of hyperspectral images accounting for spectral variability. The proposed algorithm was based on a generalization of the normal compositional model and includes an additive Gaussian noise for modeling errors. This algorithm estimated the endmembers of the scene, their variabilities provided by their variances and the corresponding abundances. The observed image was also spatially segmented into regions sharing homogeneous abundance characteristics. The physical constraints of the abundances were ensured by choosing a Dirichlet distribution for each spatial class of the image. Due to the complexity of the resulting joint posterior distribution, a Markov chain Monte Carlo procedure based on a Gibbs algorithm was used to sample the posterior of interest and to approximate the

Bayesian estimators of the unknown parameters using the generated samples. The sampling was achieved using an Hamiltonian Monte Carlo method which is well suited for problems with a large number of parameters. The proposed algorithm showed good performance when processing data presenting endmember variability, spatial correlation between pixels and in absence of pure pixels in the observed scene. UsGNCM fully exploits both the spatial dimension (segmentation, abundance and noise estimation) and the spectral dimension (estimation of endmember means and variances). **Future work includes the study of endmember variability for nonlinear mixing models. Considering spectral correlation jointly with endmember variability is also an interesting issue which would deserve to be investigated.**

## APPENDIX A

### LIKELIHOOD COMPUTATION

Considering (5), we obtain

$$y_{\ell n} | \mathbf{a}_n, \mathbf{m}_\ell, \sigma_{r\ell}^2, \psi_n^2 \sim \mathcal{N} \left( \sum_{r=1}^R a_{rn} m_{\ell r}, \sum_{r=1}^R a_{rn}^2 \sigma_{r\ell}^2 + \psi_n^2 \right). \quad (30)$$

Assuming independence between the observations in different spectral bands, we obtain the following likelihood for the vector  $\mathbf{y}_n$

$$f(\mathbf{y}_n | \mathbf{A}, \mathbf{M}, \Sigma, \Psi) \propto \left[ \frac{1}{\prod_{\ell=1}^L \left( \sum_{r=1}^R a_{rn}^2 \sigma_{r\ell}^2 + \psi_n^2 \right)} \right]^{\frac{1}{2}} \times \exp \left\{ (\mathbf{y}_n - \mathbf{M} \mathbf{a}_n)^T \begin{bmatrix} \frac{1}{\sum_{r=1}^R a_{rn}^2 \sigma_{r1}^2 + \psi_n^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sum_{r=1}^R a_{rn}^2 \sigma_{rL}^2 + \psi_n^2} \end{bmatrix} (\mathbf{y}_n - \mathbf{M} \mathbf{a}_n) \right\}. \quad (31)$$

Using the property

$$\mathbf{y}^T \text{diag}(\boldsymbol{\sigma}^2) \mathbf{y} = (\boldsymbol{\sigma}^2)^T (\mathbf{y} \odot \mathbf{y}) \quad (32)$$

and denoting  $\Lambda_{\ell n} = \left( \sum_{r=1}^R a_{rn}^2 \sigma_{r\ell}^2 + \psi_n^2 \right)^{-1}$ , yield to (8).

## APPENDIX B

## SAMPLING THE CONDITIONAL DISTRIBUTIONS

A. Sampling the parameter matrix  $\mathbf{A}$ 

1) *Reparametrization*: Sampling the abundance matrix  $\mathbf{A}$  under the sum-to-one constraint introduced in (12) and (13) is not easy. Therefore, we propose to transform the sum-to-one constraint into an inequality constraint (which will be handled more easily in the algorithm as shown below), by considering the following reparametrization

$$a_{rn} = \begin{cases} (1 - t_{rn}), & \text{if } r = 1 \\ (\prod_{k=1}^{r-1} t_{kn}) (1 - t_{rn}), & \text{if } 1 < r < R \\ (\prod_{k=1}^{r-1} t_{kn}), & \text{if } r = R \end{cases} . \quad (33)$$

The transformation (33) has been introduced in [47] and has shown interesting properties for hyperspectral unmixing in [11]. Its main advantage is to express the positivity and the sum-to-one constraints for the abundances as follows

$$0 < t_{rn} < 1, \forall r \in 1, \dots, R-1. \quad (34)$$

Note that assigning a Dirichlet prior for  $\mathbf{a}_n$  corresponds to a beta distribution prior for the coefficient  $t_{rn}$  as shown in [11], [47]

$$t_{rn} | z_n = k, \mathbf{C}_{r:R,k} \sim \mathcal{B}e \left( \sum_{i=r+1}^R c_{ik}, c_{rk} \right), \text{ for } n \in \mathcal{I}_k \quad (35)$$

where  $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_K]$  is an  $(R \times K)$  matrix containing the Dirichlet parameters. The prior associated with the vector  $\mathbf{t}_n$  is finally obtained by assuming prior independence between its elements leading to

$$f(\mathbf{t}_n | z_n = k, \mathbf{c}_k) = \frac{\Gamma \left( \sum_{i=1}^R c_{ik} \right)}{\prod_{i=1}^R \Gamma(c_{ik})} \mathbf{1}_{[0,1]^{R-1}}(\mathbf{t}_n) \prod_{r=1}^{R-1} t_{rn}^{\sum_{i=r+1}^R c_{ik}-1} (1 - t_{rn})^{c_{rk}-1} \quad (36)$$

for  $n \in \mathcal{I}_k$ , where  $\mathbf{1}_{[0,1]^{R-1}}(\cdot)$  is the indicator of the set  $[0, 1]^{R-1}$  and  $\Gamma$  denotes the gamma function.

2) *Sampling the parameter matrix  $\mathbf{T}$* : It can be shown that the  $N$  vectors  $\mathbf{t}_n, n \in \{1, \dots, N\}$  are a posteriori independent leading to

$$f(\mathbf{T}|\mathbf{Y}, \mathbf{M}, \Sigma, \mathbf{C}) = \prod_{k=1}^K \prod_{n \in \mathcal{I}_k} f(\mathbf{t}_n | z_n = k, \mathbf{y}_n, \mathbf{M}, \Sigma, \mathbf{c}_k). \quad (37)$$

Moreover, using the likelihood (8) and the prior (36) leads to the following conditional distribution

$$\begin{aligned} f(\mathbf{t}_n | z_n = k, \mathbf{y}_n, \mathbf{M}, \Sigma, \mathbf{c}_k) &\propto \left( \prod_{\ell=1}^L \Lambda_{\ell n} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{\Lambda}_{:n}^T [(\mathbf{y}_n - \mathbf{M} \mathbf{a}_n) \odot (\mathbf{y}_n - \mathbf{M} \mathbf{a}_n)] \right\} \\ &\times \mathbf{1}_{[0,1]^{R-1}}(\mathbf{t}_n) \prod_{r=1}^{R-1} t_{rn}^{\sum_{i=r+1}^R c_{ik}-1} (1 - t_{rn})^{c_{rk}-1} \end{aligned} \quad (38)$$

for  $n \in \mathcal{I}_k$  and  $\mathbf{a}_n(\mathbf{t}_n)$  has been denoted as  $\mathbf{a}_n$  in (38) for brevity. The conditional distribution (38) is not easy to sample. However, the CHMC framework is well suited for sampling the independent vectors  $\mathbf{t}_n, n \in \{1, \dots, N\}$  in an effective parallel procedure that reduces the computational cost. Moreover, the small size of these vectors (of size  $(R-1) \times 1$ ) improves the convergence of the sampler. Note that the CHMC requires the definition of the potential energy  $U(\mathbf{t}_n) = -\log[f(\mathbf{t}_n | z_n = k, \mathbf{y}_n, \mathbf{M}, \Sigma, \mathbf{c}_k)]$  given by

$$U(\mathbf{t}_n) = U_1 + U_2 + U_3 \quad (39)$$

with

$$\begin{aligned} U_1 &= \frac{1}{2} \mathbf{\Lambda}_{:n}^T [(\mathbf{y}_n - \mathbf{M} \mathbf{a}_n) \odot (\mathbf{y}_n - \mathbf{M} \mathbf{a}_n)] \\ U_2 &= -\sum_{r=1}^R \left\{ \left( \sum_{i=r+1}^R c_{ik} - 1 \right) \log(t_{rn}) + (c_{rk} - 1) \log(1 - t_{rn}) \right\} \\ U_3 &= -\frac{1}{2} \sum_{\ell=1}^L \log(\Lambda_{\ell n}). \end{aligned} \quad (40)$$

Note finally that the derivatives of  $U$  with respect to the variable of interest  $\mathbf{t}_n$  (that are required for the CHMC steps) are provided in the Appendix C.

### B. Sampling the mean endmember matrix $\mathbf{M}$

Straightforward computations using the posterior distribution (19) yield

$$f(\mathbf{M}|\mathbf{Y}, \mathbf{T}, \Sigma) = \prod_{\ell=1}^L f(\mathbf{M}_{\ell:} | \mathbf{Y}_{\ell:}, \mathbf{T}, \Sigma_{:\ell}) \quad (41)$$

where

$$f(\mathbf{M}_{\ell} | \mathbf{Y}_{\ell}, \mathbf{T}, \boldsymbol{\Sigma}_{:\ell}) \propto \exp \left\{ -\frac{1}{2} [(\mathbf{Y}_{\ell} - \mathbf{M}_{\ell} \mathbf{A}) \odot (\mathbf{Y}_{\ell} - \mathbf{M}_{\ell} \mathbf{A})] \boldsymbol{\Lambda}_{\ell}^T \right\} \\ \times \exp \left( -\frac{\|\mathbf{M}_{\ell} - \widetilde{\mathbf{M}}_{\ell}\|^2}{2\epsilon^2} \right) \mathbf{1}_{[0,1]^R}(\mathbf{M}_{\ell}). \quad (42)$$

Equation (41) results from the independence between the columns of the matrix  $\mathbf{M}$  (vectors of small size  $R \times 1$ ). This interesting property promotes the use of a parallel CHMC algorithm for sampling  $\mathbf{T}$ . The potential energy  $V$  associated with the conditional distribution of  $\mathbf{M}_{\ell}$  is given by

$$V(\mathbf{M}_{\ell}) = \frac{1}{2} [(\mathbf{Y}_{\ell} - \mathbf{M}_{\ell} \mathbf{A}) \odot (\mathbf{Y}_{\ell} - \mathbf{M}_{\ell} \mathbf{A})] \boldsymbol{\Lambda}_{\ell}^T + \frac{\|\mathbf{M}_{\ell} - \widetilde{\mathbf{M}}_{\ell}\|^2}{2\epsilon^2}. \quad (43)$$

The derivatives of  $V$  with respect to  $\mathbf{M}_{\ell}$  are provided in the appendix C.

### C. Sampling the variance of the endmember matrix

Considering (19) yields the following conditional distribution for matrix  $\boldsymbol{\Sigma}$  containing the endmember variances

$$f(\boldsymbol{\Sigma} | \mathbf{Y}, \mathbf{T}, \mathbf{M}) = \prod_{\ell=1}^L f(\boldsymbol{\Sigma}_{:\ell} | \mathbf{Y}_{\ell}, \mathbf{T}, \mathbf{M}_{\ell}) \quad (44)$$

with

$$f(\boldsymbol{\Sigma}_{:\ell} | \mathbf{Y}_{\ell}, \mathbf{T}, \mathbf{M}_{\ell}) \propto \left( \prod_{\ell=1}^L \boldsymbol{\Lambda}_{\ell n} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} [(\mathbf{Y}_{\ell} - \mathbf{M}_{\ell} \mathbf{A}) \odot (\mathbf{Y}_{\ell} - \mathbf{M}_{\ell} \mathbf{A})] \boldsymbol{\Lambda}_{\ell}^T \right\} \\ \times \prod_{r=1}^R \frac{1}{\sigma_{r\ell}^2} \mathbf{1}_{\mathbb{R}^+}(\sigma_{r\ell}^2). \quad (45)$$

Sampling from (45) can again be performed using a CHMC algorithm (as in Sections B-A2 and B-B). The potential energy associated with the vector  $\boldsymbol{\Sigma}_{:\ell}$  is

$$W(\boldsymbol{\Sigma}_{:\ell}) = W_1 + W_2 + W_3 \quad (46)$$

with

$$W_1 = \frac{1}{2} [(\mathbf{Y}_{\ell} - \mathbf{M}_{\ell} \mathbf{A}) \odot (\mathbf{Y}_{\ell} - \mathbf{M}_{\ell} \mathbf{A})] \boldsymbol{\Lambda}_{\ell}^T \\ W_2 = \sum_{r=1}^R \log(\sigma_{r\ell}^2) \\ W_3 = -\frac{1}{2} \sum_{n=1}^N \log(\boldsymbol{\Lambda}_{\ell n}). \quad (47)$$

The derivatives of  $W$  with respect to  $\boldsymbol{\Sigma}_{:\ell}$  are provided in the appendix C.

#### D. Sampling the labels

The conditional distribution associated with the discrete random variable  $z_n$  is given by

$$f(z_n = k | \mathbf{t}_n, \mathbf{c}_k) \propto f(\mathbf{t}_n | z_n = k, \mathbf{c}_k) \exp \left[ 2\beta \sum_{n' \in \mathcal{V}(n)} \delta(k - z_{n'}) \right] \quad (48)$$

where  $f(\mathbf{t}_n | z_n = k, \mathbf{c}_k)$  has been defined in (36). Sampling from this conditional distribution is classically performed by drawing a discrete value in the finite set  $\{1, \dots, K\}$  with the probabilities (48).

#### E. Sampling the noise variance $\Psi$

Considering (19) yields the following conditional distribution for the noise variance matrix  $\Psi$

$$f(\Psi | \mathbf{z}, \mathbf{T}, \mathbf{Y}, \mathbf{M}, \Sigma, \mathbf{c}) = \prod_{n=1}^N f(\psi_n^2 | z_n = k, \mathbf{t}_n, \mathbf{y}_n, \mathbf{M}, \Sigma, \mathbf{c}_k) \quad (49)$$

with

$$\begin{aligned} f(\psi_n^2 | z_n = k, \mathbf{t}_n, \mathbf{y}_n, \mathbf{M}, \Sigma_{:,l}, \mathbf{c}_k) &\propto \left( \prod_{\ell=1}^L \Lambda_{\ell n} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \Lambda_{:,n}^T [(\mathbf{y}_n - \mathbf{M}\mathbf{a}_n) \odot (\mathbf{y}_n - \mathbf{M}\mathbf{a}_n)] \right\} \\ &\times \exp(-\lambda \psi_n^2) \mathbf{1}_{\mathbb{R}^+}(\psi_n^2) \end{aligned} \quad (50)$$

for  $n \in \mathcal{I}_k$ . This distribution is sampled using a parallel CHMC procedure with the following potential energy

$$H(\psi_n^2) = U_1 + U_3 + \lambda \psi_n^2. \quad (51)$$

#### F. Sampling the Dirichlet coefficients

Using (19) and (20), it can be easily shown that the conditional distribution of  $\mathbf{c}_k | \mathbf{T}, \mathbf{z}_{n \in \mathcal{I}_k}$  is given by

$$f(\mathbf{c}_k | \mathbf{T}, \mathbf{z}_{n \in \mathcal{I}_k}) \propto \prod_{n \in \mathcal{I}_k} \left\{ \left[ \frac{\Gamma(\sum_{r=1}^R c_{rk})}{\prod_{r=1}^R \Gamma(c_{rk})} \right]^{\gamma+1} \exp \left( -\alpha \sum_{r=1}^R c_{rk} + R\alpha \right) \prod_{r=1}^R a_{rn}^{c_{rk}-1} \right\} \quad (52)$$

for  $k \in \{1, \dots, K\}$ . This distribution is also sampled using a CHMC procedure. The corresponding potential energy is given by

$$P(\mathbf{c}_k) = P_1 + P_2 \quad (53)$$

with

$$\begin{aligned}
P_1 &= (\gamma + 1) \sum_{n \in \mathcal{I}_k} \left[ -\log \Gamma \left( \sum_{r=1}^R c_{rk} \right) + \sum_{r=1}^R \log \Gamma (c_{rk}) \right] \\
P_2 &= \sum_{n \in \mathcal{I}_k} \left[ \alpha \sum_{r=1}^R c_{rk} - R\alpha - \sum_{r=1}^R \log (a_{rn}^{c_{rk}-1}) \right].
\end{aligned} \tag{54}$$

## APPENDIX C

### DERIVATIVES OF THE POTENTIAL FUNCTIONS

The derivative of  $U$  with respect to  $\mathbf{t}_n$  is given by

$$\frac{\partial U}{\partial \mathbf{t}_n} = \frac{\partial U_1}{\partial \mathbf{a}_n} \frac{\partial \mathbf{a}_n}{\partial \mathbf{t}_n} + \frac{\partial U_2}{\partial \mathbf{t}_n} + \frac{\partial U_3}{\partial \mathbf{a}_n} \frac{\partial \mathbf{a}_n}{\partial \mathbf{t}_n} \tag{55}$$

with

$$\begin{aligned}
\frac{\partial U_1}{\partial \mathbf{a}_n} &= -[\boldsymbol{\Lambda}_{:n} \odot (\mathbf{y}_n - \mathbf{M}\mathbf{a}_n)]^T \mathbf{M} + \frac{1}{2} [(\mathbf{y}_n - \mathbf{M}\mathbf{a}_n) \odot (\mathbf{y}_n - \mathbf{M}\mathbf{a}_n)]^T \left( \frac{\partial \boldsymbol{\Lambda}_{:n}}{\partial \mathbf{a}_n} \right)^T \\
\left( \frac{\partial \boldsymbol{\Lambda}_{\ell n}}{\partial \mathbf{a}_n} \right)^T &= -2\boldsymbol{\Lambda}_{\ell n}^2 [\text{diag}(\mathbf{a}_n) \boldsymbol{\Sigma}_{:,l}] \\
\frac{\partial U_3}{\partial \mathbf{a}_n} &= \mathbf{a}_n^T \odot [\boldsymbol{\Sigma} \boldsymbol{\Lambda}_{:n}]^T \\
\frac{\partial U_2}{\partial t_{rn}} &= -\frac{\sum_{i=r+1}^R c_{ik} - 1}{t_{rn}} + \frac{c_{rk} - 1}{1 - t_{rn}}, \quad \forall r \in \{1, \dots, R-1\}
\end{aligned} \tag{56}$$

and

$$\frac{\partial a_{rn}}{\partial t_{in}} = \begin{cases} 0 & \text{if } i > r \\ \frac{a_{rn}}{t_{in}-1} & \text{if } i = r \\ \frac{a_{rn}}{t_{in}} & \text{if } i < r \end{cases} . \tag{57}$$

The derivative of  $V$  with respect to  $\mathbf{M}_\ell$  is given by

$$\frac{\partial V}{\partial \mathbf{M}_\ell} = -[\boldsymbol{\Lambda}_\ell \odot (\mathbf{Y}_\ell - \mathbf{M}_\ell \mathbf{A})] \mathbf{A}^T + \frac{1}{\epsilon^2} (\mathbf{M}_\ell - \widetilde{\mathbf{M}}_\ell) . \tag{58}$$

The derivatives of  $W$  with respect to  $\boldsymbol{\Sigma}_{:,l}$  are given by

$$\begin{aligned}
\frac{\partial W_1}{\partial \boldsymbol{\Sigma}_{:,l}} &= -\frac{1}{2} [(\mathbf{Y}_\ell - \mathbf{M}_\ell \mathbf{A}) \odot (\mathbf{Y}_\ell - \mathbf{M}_\ell \mathbf{A}) \odot \boldsymbol{\Lambda}_\ell \odot \boldsymbol{\Lambda}_\ell] (\mathbf{A} \odot \mathbf{A})^T \\
\frac{\partial W_2}{\partial \sigma_{r\ell}^2} &= \frac{\partial W_2}{\partial \sigma_{r\ell}^2} = \frac{1}{\sigma_{r\ell}^2}, \quad \forall r \in \{1, \dots, R\} \\
\frac{\partial W_3}{\partial \boldsymbol{\Sigma}_{:,l}} &= \frac{1}{2} [\boldsymbol{\Lambda}_\ell (\mathbf{A} \odot \mathbf{A})^T]
\end{aligned} \tag{59}$$

The derivatives of  $H$  with respect to  $\psi_n^2$  is given by

$$\frac{\partial T}{\partial \psi_n^2} = \frac{\partial U_1}{\partial \psi_n^2} + \frac{\partial U_3}{\partial \psi_n^2} + \lambda \quad (60)$$

with

$$\begin{aligned} \frac{\partial U_1}{\partial \psi_n^2} &= -\frac{1}{2} \sum_{\ell=1}^L [(\mathbf{y}_n - \mathbf{M}\mathbf{a}_n) \odot (\mathbf{y}_n - \mathbf{M}\mathbf{a}_n) \odot \mathbf{\Lambda}_{\cdot:n} \odot \mathbf{\Lambda}_{\cdot:n}] \\ \frac{\partial U_3}{\partial \psi_n^2} &= -\frac{1}{2} \sum_{\ell=1}^L \mathbf{\Lambda}_{\ell n} \end{aligned} \quad (61)$$

The derivative of  $P$  with respect to  $\mathbf{c}_{rk}$  is given by

$$\frac{\partial P}{\partial \mathbf{c}_{rk}} = \frac{\partial P_1}{\partial \mathbf{c}_{rk}} + \frac{\partial P_2}{\partial \mathbf{c}_{rk}} \quad (62)$$

with

$$\begin{aligned} \frac{\partial P_1}{\partial \mathbf{c}_{rk}} &= (\gamma + 1) \sum_{n \in \mathcal{I}_k} \left[ -\Upsilon \left( \sum_{r'=1}^R c_{r'k} \right) + \Upsilon(c_{rk}) \right] \\ \frac{\partial P_2}{\partial \mathbf{c}_{rk}} &= \sum_{n \in \mathcal{I}_k} [\alpha - \log(a_{rn})] \end{aligned} \quad (63)$$

where  $\Upsilon$  denotes the polygamma function, i.e., the derivative of the log-gamma function.

## APPENDIX D

### IDENTIFIABILITY PROBLEM

The noise corrupting hyperspectral images is generally of small variance which explains the prior used for this parameter. This choice of prior ensures problem identifiability as described in the following (we will adopt a similar approach as in [18]). Assume first that the endmember mean  $\mathbf{M}$  and the abundance  $\mathbf{A}$  are fixed parameters and that the only unknown variables are the noise variance  $\psi_n$  and the endmember covariance matrix  $\mathbf{\Sigma}$ . The log-likelihood associated with the  $n$ th pixel of the GNCM is given by

$$\log f(\mathbf{y}_n | \psi_n, \mathbf{\Sigma}) = \sum_{\ell=1}^L \left[ \frac{1}{2} \log(\mathbf{\Lambda}_{\ell n}) - \frac{p_{\ell n} \mathbf{\Lambda}_{\ell n}}{2} \right] \quad (64)$$

with  $\Lambda_{\ell n} = \left( \sum_{r=1}^R a_{rn}^2 \sigma_{r\ell}^2 + \psi_n^2 \right)^{-1}$  and  $p_{\ell n} = \left( y_{\ell n} - \sum_{r=1}^R a_{rn} m_{r\ell} \right)^2$ . To maximize this log-likelihood, we set its partial derivatives to zero. For the  $\ell$ th spectral band, the following equations are obtained

$$\begin{cases} \frac{\partial \log f(\mathbf{y}_n | \psi_n, \Sigma)}{\partial \sigma_{\ell 1}^2} = -\frac{a_{1n}^2 \Lambda_{\ell n}}{2} + \frac{a_{1n}^2 p_{\ell n} \Lambda_{\ell n}^2}{2} = 0 \\ \vdots \\ \frac{\partial \log f(\mathbf{y}_n | \psi_n, \Sigma)}{\partial \sigma_{\ell R}^2} = -\frac{a_{Rn}^2 \Lambda_{\ell n}}{2} + \frac{a_{Rn}^2 p_{\ell n} \Lambda_{\ell n}^2}{2} = 0 \\ \frac{\partial \log f(\mathbf{y}_n | \psi_n, \Sigma)}{\partial \psi_n^2} = -\frac{\Lambda_{\ell n}}{2} + \frac{p_{\ell n} \Lambda_{\ell n}^2}{2} = 0 \end{cases} \quad (65)$$

which leads to

$$\Lambda_{\ell n}^{-1} = \sum_{r=1}^R a_{rn}^2 \sigma_{r\ell}^2 + \psi_n^2 = p_{\ell n}. \quad (66)$$

When considering only one pixel, condition (66) shows that the system (65) has several maxima (with respect to  $\Sigma$  and  $\psi_n$ ) located on a hyperplane. However, when considering  $N$  pixels, we obtain the following linear system for the  $\ell$ th spectral band

$$\mathbf{H} \mathbf{v}_\ell = \mathbf{p}_\ell \quad (67)$$

with

$$\mathbf{H} = \begin{bmatrix} a_{11}^2 & \cdots & a_{R1}^2 & 1 & 0 & \cdots & 0 \\ a_{12}^2 & \cdots & a_{R2}^2 & 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \vdots & 0 & 0 & \ddots & 0 \\ a_{1N}^2 & \cdots & a_{RN}^2 & 0 & 0 & \cdots & 1 \end{bmatrix}, \text{ and } \mathbf{v}_\ell = \begin{bmatrix} \sigma_{1\ell}^2 \\ \vdots \\ \sigma_{R\ell}^2 \\ \psi_1^2 \\ \vdots \\ \psi_N^2 \end{bmatrix}. \quad (68)$$

This system has  $R + N$  unknowns (contained in  $\mathbf{v}_\ell$ ) and  $N$  independent equations (the rank of  $\mathbf{H}$  is obviously  $N$ ). Therefore, we have an infinite number of solutions and one has to add at least  $R$  equations or constraints to obtain a unique solution. If we consider all spectral bands, the system becomes

$$\mathbf{H} \mathbf{V} = \mathbf{P} \quad (69)$$

where  $\mathbf{V}$  is an  $(R + N) \times L$  matrix of unknowns and  $\mathbf{P}$  is an  $(N \times L)$  matrix. In this case, one has to include at least  $L \times R$  equations or constraints to ensure identifiability. In this paper, we have assumed that the noise variance  $\psi_n^2$  is smaller than the endmember variance which introduces

$N$  additional constraints. So, provided that  $N > RL$  (which is a realistic assumption that is generally satisfied), the obtained problem becomes identifiable since we have more independent equations than unknowns. Note for example that when considering  $L = 224$  bands and  $R = 4$  endmembers, the image size should be at least equal to 896 pixels which is a small image. This identifiability proof has been included in the technical report [45].

As shown above, the proposed unmixing problem is identifiable when considering  $N$  additional constraints on the noise variance. In this paper, these constraints have been achieved by considering prior enforcing small values of the additive noise variance. The considered exponential distribution is similar to the Laplace prior (with additional positivity constraint) that has been widely used in the Bayesian context to enforce sparsity [35], [36]. Note that this prior enforces an  $L_1$  norm constraint (as in LASSO problems [48]) for the noise variance.

## APPENDIX E

### JUSTIFICATION OF THE GAUSSIAN CHOICE

The endmember variability does not always follows a Gaussian distribution. However, the Gaussian distribution is certainly a good candidate to approximate the endmember variability. In this paper, we showed in Fig. 1 that the endmember variability was approximated using a non-isotropic ellipse around the true endmember. This can also be understood as an  $\ell_2$  regularization of endmember variability around the true endmember. In addition to these reasons, the Gaussian distribution has been successfully used in hyperspectral imagery to approximate endmember variability in [18], [20], [29]. Finally, to further justify our choice, we have performed some experiments using similar spectra from the USGS library<sup>6</sup> that present endmember variability. These spectra are displayed below in Fig. 15 (top). We have also computed the means of the spectra that have been subtracted from the observed spectra yielding spectral residuals. An histogram of the spectral residuals is displayed in Fig. 15 (bottom). This histogram is clearly close to a Gaussian distribution which justifies the Gaussian choice for endmember variability.

<sup>6</sup>[Online]. Available: <http://www.lx.it.pt/~bioucas/code.htm>

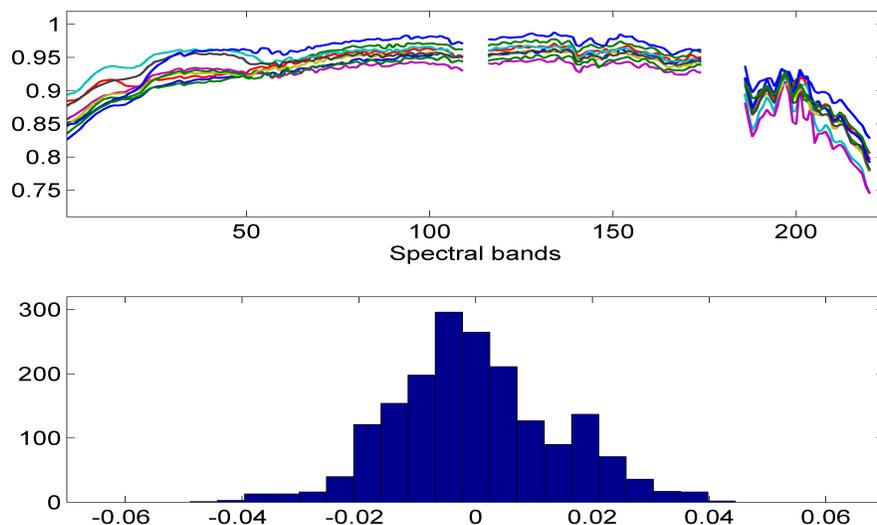


Fig. 15. (top) 9 USGS spectra showing the effect of endmember variability. (bottom) Histogram computed from the 9 USGS spectral residuals.

## REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, pp. 44–57, Jan. 2002.
- [2] 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. Remote Sens.*, vol. 29, no. 3, pp. 529–545, March 2001.
- [3] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 354–379, April 2012.
- [4] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. Bermudez, S. McLaughlin, and A. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan 2014.
- [5] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, 2011.
- [6] 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 Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.
- [7] 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, April 2005.
- [8] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries 4th Annu. JPL Airborne Geoscience Workshop*, vol. 1. Washington, D.C.: JPL Pub., 1993, pp. 11–14.
- [9] M. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. on Applied Geologic Remote Sens.*, vol. 2, Vancouver, April 1999, pp. 337–344.

- [10] J. Nascimento and J. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 863–878, March 2012.
- [11] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Unsupervised post-nonlinear unmixing of hyperspectral images using a Hamiltonian Monte Carlo algorithm," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2663–2675, June 2014.
- [12] B. Somers, G. P. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," *Remote Sens. Environ.*, vol. 115, no. 7, pp. 1603 – 1616, 2011.
- [13] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan 2014.
- [14] D. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. Green, "Mapping Chaparral in the Santa Monica Mountains Using Multiple Endmember Spectral Mixture Models," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 267–279, Sept. 1998.
- [15] C. Bateson, G. Asner, and C. Wessman, "Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1083–1094, Mar 2000.
- [16] M. Goenaga, M. Torres-Madronero, M. Velez-Reyes, S. Van Bloem, and J. Chinaea, "Unmixing analysis of a time series of hyperion images over the Gunica dry forest in Puerto Rico," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 6, no. 2, pp. 329–338, April 2013.
- [17] B. Somers, M. Zortea, A. Plaza, and G. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 396–408, April 2012.
- [18] 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, pp. 1403–1413, June 2010.
- [19] A. Zare, P. Gader, D. Drashnikov, and T. Glenn, "Beta compositional model for hyperspectral unmixing," in *Proc. IEEE GRSS WHISPERS*, Gainesville, USA, Jun. 2013.
- [20] A. Zare, P. Gader, and G. Casella, "Sampling piecewise convex unmixing and endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1655–1665, March 2013.
- [21] X. Du, A. Zare, P. Gader, and D. Dranishnikov, "Spatial and spectral unmixing using the beta compositional model," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 6, pp. 1994–2003, June 2014.
- [22] C. Song, "Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability?" *Remote Sens. Environ.*, vol. 95, no. 2, pp. 248 – 263, 2005.
- [23] N. Bali and A. Mohammad-Djafari, "Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 217–225, Feb. 2008.
- [24] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, Nov 2011.
- [25] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with  $\ell_1$ -norm spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2654–2665, May 2014.
- [26] R. Rand and D. Keenan, "Spatially smooth partitioning of hyperspectral imagery using spectral/spatial measures of disparity," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1479–1490, June 2003.
- [27] S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*. ser. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis, 2011.
- [28] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Residual component analysis of hyperspectral images:

- Application to joint nonlinear unmixing and nonlinearity detection,” *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2148–2158, May 2014.
- [29] D. Stein, “Application of the normal compositional model to the analysis of hyperspectral imagery,” in *Proc. IEEE Workshop Advances in Techniques for Analysis of Remotely Sensed Data*, Oct 2003, pp. 44–51.
- [30] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, “Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4133–4145, Nov. 2006.
- [31] F. Schmidt, A. Schmidt, E. Tréguier, M. Guiheneuf, S. Moussaoui, and N. Dobigeon, “Implementation strategies for hyperspectral unmixing using bayesian source separation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4003–4013, Nov 2010.
- [32] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, “Estimating the granularity coefficient of a Potts-Markov random field within a Markov Chain Monte Carlo algorithm,” *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2385–2397, June 2013.
- [33] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, “Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, July 2008.
- [34] 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.
- [35] N. Dobigeon, A. Hero, and J.-Y. Tourneret, “Hierarchical bayesian sparse image reconstruction with application to mrfm,” *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2059–2070, Sept 2009.
- [36] L. Chaari, J.-Y. Tourneret, and H. Batatia, “Sparse bayesian regularization using bernoulli-laplacian priors,” in *Proc. EUSIPCO*, Marrakech, Morocco, Sept 2013, pp. 1–5.
- [37] Z. Ma, “Bayesian estimation of the Dirichlet distribution with expectation propagation,” in *Proc. EUSIPCO*, Bucharest, Romania, Aug. 2012.
- [38] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.
- [39] M. Girolami and B. Calderhead, “Riemann manifold langevin and hamiltonian monte carlo methods,” *J. Roy. Stat. Soc. Ser. B*, vol. 73, no. 2, pp. 123–214, March 2011.
- [40] S. Lan, “Advanced bayesian computational methods through geometric techniques,” Ph.D. dissertation, University of California, Irvine, California, USA, 2013.
- [41] C. P. Robert and D. Cellier, “Convergence control of MCMC algorithms,” in *Discretization and MCMC Convergence Assessment*, C. P. Robert, Ed. New York: Springer Verlag, 1998, pp. 27–46.
- [42] A. Zare and D. Anderson, “Earth movers distance-based simultaneous comparison of hyperspectral endmembers and proportions,” *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 6, pp. 1910–1921, June 2014.
- [43] RSI (Research Systems Inc.), *ENVI User’s guide Version 4.0*, Boulder, CO 80301 USA, Sept. 2003.
- [44] D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon, “Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential of hyperspectral imagery,” in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, July 2011, pp. 3672–3675.
- [45] A. Halimi, N. Dobigeon, and J.-Y. Tourneret, “Technical report associated with the paper “Unsupervised unmixing of hyperspectral images accounting for endmember variability”,” University of Toulouse, France, Tech. Rep., May 2014. [Online]. Available: <https://sites.google.com/site/abderrahimhalimi/publications>

- [46] C. Févotte and N. Dobigeon, “Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization,” in *ArXiv e-prints*, March 2014.
- [47] M. J. Betancourt, “Cruising the simplex: Hamiltonian Monte Carlo and the Dirichlet distribution,” in *ArXiv e-prints*, Oct. 2013.
- [48] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, p. 267–288, 1996.