Adaptive Markov Random Fields for Joint Unmixing and Segmentation of Hyperspectral Images

Olivier Eches, Jon Atli Benediktsson, Nicolas Dobigeon, and Jean-Yves Tourneret

Abstract—Linear spectral unmixing is a challenging problem in hyperspectral imaging that consists of decomposing an observed pixel into a linear combination of pure spectra (or *endmembers*) with their corresponding proportions (or *abundances*). Endmember extraction algorithms can be employed for recovering the spectral signatures while abundances are estimated using an inversion step. Recent works have shown that exploiting spatial dependencies between image pixels can improve spectral unmixing. Markov random fields (MRF) are classically used to model these spatial correlations and partition the image into multiple classes with homogeneous abundances. This paper proposes to define the MRF sites using similarity regions. These regions are built using a self-complementary area filter that stems from the morphological theory. This kind of filter divides the original image into *flat zones* where the underlying pixels have the same spectral values. Once the MRF has been clearly established, a hierarchical Bayesian algorithm is proposed to estimate the abundances, the class labels, the noise variance and the corresponding hyperparameters. A hybrid Gibbs sampler is then constructed to generate samples according to the corresponding posterior distribution of the unknown parameters and hyperparameters. Simulations conducted on synthetic and real AVIRIS hyperspectral data demonstrate the good performance of the algorithm.

Index Terms—Morphological filter, Markov random field, spectral unmixing, hyperspectral images, segmentation.

I. INTRODUCTION

Hyperspectral images are very high resolution remote sensing images that have been acquired in a hundred of spectral bands simultaneously. Since the growing availability of such images within the last years, many studies have been conducted by the image processing community for the analysis of these images. A particular attention has been devoted to the spectral unmixing problem. Classical unmixing algorithms assume that the image pixels are linear combinations of a given number of pure materials spectra or *endmembers*

with corresponding fractions referred to as *abundances* [1] (the most recent techniques have been reported in [2]). The mathematical formulation of this linear mixing model (LMM) for an observed pixel p in L bands is

$$\boldsymbol{y}_p = \boldsymbol{M}\boldsymbol{a}_p + \boldsymbol{n}_p \qquad (1)$$

where $\boldsymbol{M} = [\boldsymbol{m}_1,...,\boldsymbol{m}_R]$ is the $L \times R$ spectral signature matrix, $\boldsymbol{a}_p$ is the $R \times 1$ abundance vector and $\boldsymbol{n}_p$ is the $L \times 1$ additive noise vector. This paper assumes that the additive noise vector is white Gaussian with the same variance in each band as in [3], [4]. For a hyperspectral image with P pixels, by denoting $\boldsymbol{Y} = [\boldsymbol{y}_1,\ldots,\boldsymbol{y}_P]$, $\boldsymbol{A} = [\boldsymbol{a}_1,\ldots,\boldsymbol{a}_p]$ and $\boldsymbol{N} = [\boldsymbol{n}_1,\ldots,\boldsymbol{n}_P]$, the LMM for the whole image is

$$\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{A} + \boldsymbol{N}. \qquad (2)$$

The unmixing problem consists of estimating the endmember spectra contained in $\boldsymbol{M}$ and the corresponding abundance matrix $\boldsymbol{A}$. Endmember extraction algorithms (EEA) are classically used to recover the spectral signatures. These algorithms include the minimum volume simplex analysis (MVSA) [5] and the well-known N-FINDR algorithm [6]. After the EEA step, the abundances are estimated under the sum-to-one and positivity constraints. Several methods have been proposed for the *inversion* step. They are based on constrained optimization techniques such as the fully constrained least squares (FCLS) algorithm [7] or on Bayesian techniques [8], [9]. The Bayesian paradigm consists of assigning appropriate prior distributions to the abundances and to solve the unmixing problem using the joint posterior distribution of the unknown model parameters.

Another approach based on a fuzzy membership process introduced in [10] inspired a Bayesian technique where spatial correlation between pixels are taken into account [11]. This approach that used Markov random fields (MRFs) to model pixel dependencies resulted in a joint segmentation and unmixing algorithm. MRFs have been introduced by Besag in [12] with their pseudo-likelihood approximation. The Gibbs distribution inherent to MRFs was exploited in [13]. Since this pioneer work, MRFs have been actively used in the image processing community for modeling spatial correlations. Examples of applications include the segmentation of SAR or brain magnetic resonance images [14] [15]. Other interesting works involving MRFs for segmentation and classification include [16]–[18]. A major drawback of MRFs is their computational cost, which is proportional to the image size. In [19], the authors proposed to partition the image in two independent set

of pixels, allowing the sampling algorithm to be parallelized. However, this method is only valid for a 4-pixel neighborhood.

This paper studies a novel approach for introducing spatial correlation between adjacent pixels of an hyperspectral image allowing computational cost of MRFs to be reduced significantly. The neighborhood relations are usually defined between spatially close pixels or *sites*. This contribution proposes to define a new neighborhood relation between sites regrouping spectrally consistent pixels. These *similarity regions* are built using a filter stemming from mathematical morphology. Mathematical morphology is a nonlinear image processing methodology based upon lattice theory [20], [21] that has been widely used for image analysis (see [22] and references therein), with a focus on hyperspectral images in [23]. Based on mathematical morphology, Soille developed a self-complementary area filter in [24] that allows one to properly define structures while removing meaningless objects. The self-complementary area filter has also been used in [25] for classifying hyperspectral images. This paper defines similarity regions using the same self-complementary area filter. After image partitioning, image neighborhoods are defined between similarity regions ensuring a distance criterion between their spectral medians. The resulting MRF sites are less numerous than the number of pixels which reduces computational complexity.

This new way of defining MRFs is applied to the joint unmixing and segmentation algorithm of [11]. After a pre-processing step defining the similarity regions, an implicit classification is carried out by assigning hidden discrete variables or class *labels* to image regions. Then, a Potts-Markov field [26] is chosen as a prior for the labels, using the proposed neighborhood relation. Therefore, a pixel belonging to a given similarity region must belong to the class that shares not only the same abundance mean vector and covariance matrix but also the same spectral characteristics. In addition to the label prior, the Bayesian method used in this work requires to define an abundance prior distribution. Instead of reparameterizing the abundances as in [11], we choose a Dirichlet distribution whose parameters can be selected to adjust the abundance means and variances for each class. The Dirichlet distribution is classically used as prior for parameters subjected to positivity and sum-to-one constraints [27]. The associated hyperparameters are assigned non-informative prior distributions according to a hierarchical Bayesian model.

The resulting joint posterior distribution of the unknown model parameters and hyperparameters can be computed from the likelihood and the priors. Deriving the Bayesian estimators such as the minimum mean square error (MMSE) and maximum a posteriori (MAP) estimators is too difficult from this posterior distribution. One might think to handle this problem by using the well-known expectation maximization (EM) algorithm. However, this algorithm can have serious shortcomings including the convergence to a local maximum of the posterior [28, p. 259]. Moreover, using the EM algorithm to jointly solve the unmixing and classification problem is not straightforward. Therefore, we study as in [11] a Markov chain Monte Carlo (MCMC) method that bypasses these shortcomings and allow samples asymptotically distributed according to the posterior of interest to be generated. Note that this method has some analogy with previous works proposed for the analysis of hyperspectral images [9], [16]. The samples generated by the MCMC method are then used to compute the Bayesian estimators of the image labels and class parameters. Therefore, the proposed Bayesian framework jointly solves the classification and abundance estimation problems.

The paper is organized as follows. Section II describes the morphological area filter and its associated MRF. Section III presents the hierarchical Bayesian model used for the joint unmixing and segmentation of hyperspectral images. The MCMC algorithm used to generate samples according to the joint posterior distribution of this model is described in Section IV. Simulation results on synthetic and real hyperspectral data are presented in Sections V and VI. Conclusions and future works are finally reported in Section VII.

II. TECHNICAL BACKGROUND

This section presents in more details the morphological self-complementary area filter and introduces the MRF that is used for describing the dependence between the regions.

A. Adaptive neighborhood

In order to build the adaptive neighborhood on hyperspectral data, a *flattening* procedure stemming from the self-complementarity property [24] was employed in [25]. Self-complementarity is an important property in morphological theory and allows the structure of interest to be preserved independently of their contrasts while removing small meaningless structures (e.g., cars, trees,...) in very high resolution remote sensing images. The algorithm developed by Soille in [24] exploits this property in a two step procedure that divides the image into flat zones, i.e., regions whose neighboring pixels have the same values satisfying any area criterion λ. This procedure is repeated until the desired minimal flat zone size λ is obtained. Note that this self-complementary area filter cannot be directly used on hyperspectral images since the complete ordering property that any morphological operator needs is absent from these data. The strategy studied in [25] uses principal component analysis (PCA) to reduce data dimensionality. The area filtering is then computed on the data projected on the first principal component defined by the largest covariance matrix eigenvalue. The resulting flat zones contain pixels that are spectrally consistent and are therefore considered in the same similarity region.

As stated in the introduction, the main contribution of this paper consists of using the similarity region building method developed in [25] as a pre-processing step for a spatial unmixing algorithm. The regions resulting from the method derived in [25] are considered for each band of the data. Spatial information is then extracted from each of these regions by computing the corresponding median vector. More precisely, if we denote the number of similarity regions by S and the sth region by $\boldsymbol{\Omega}_s$ ($s = 1, \ldots, S$), then the vector median value for this region is defined as

$$\boldsymbol{\Upsilon}_s = \mathrm{med}(\boldsymbol{Y}_{\boldsymbol{\Omega}_s}), \tag{3}$$

where $\boldsymbol{Y}_{\boldsymbol{\Omega}_s}$ is the matrix of observed pixels belonging to the region $\boldsymbol{\Omega}_s$ and $\dim(\boldsymbol{\Upsilon}_s) = L$ is the number of spectral bands. As explained in [25], the median vector ensures spectral consistency as opposed to the mean vector.

As in [11], this paper assumes that the classes contain neighboring pixels that have *a priori* close abundances. This spatial dependency is modeled using the resulting similarity regions that contain spectrally consistent pixels. In other words, if we denote as $\mathcal{C}_1, \ldots, \mathcal{C}_K$ the image classes, a label vector of size $S \times 1$ (with $S \geq K$) denoted as $\boldsymbol{z} = [z_1, \ldots, z_S]^T$ with $z_s \in \{1, \ldots, K\}$ is introduced to identify the class of each region $\boldsymbol{\Omega}_s$, i.e., $z_s = k$ if and only if all pixels of $\boldsymbol{\Omega}_s$ belong to $\mathcal{C}_k$. Note that, in each class, the abundance vectors to be estimated are assumed to share the same first and second order statistical moments, i.e., $\forall k \in \{1, \ldots, K\}$, $\forall \boldsymbol{\Omega}_s \in \mathcal{C}_k$, $\forall p \in \boldsymbol{\Omega}_s$

$$\mathrm{E}\left[\boldsymbol{a}_p\right] = \boldsymbol{\mu}_k$$
$$\mathrm{E}\left[\left(\boldsymbol{a}_p - \boldsymbol{\mu}_k\right)\left(\boldsymbol{a}_p - \boldsymbol{\mu}_k\right)^T\right] = \boldsymbol{\Lambda}_k. \tag{4}$$

Therefore, the kth class of the hyperspectral image to be unmixed is fully characterized by its abundance mean vector $\boldsymbol{\mu}_k$ and its abundance covariance matrix $\boldsymbol{\Lambda}_k$.

B. Adaptive Markov random fields

Since the work of Geman and Geman [13], MRFs have been widely used in the image processing community (for examples, see [29], [30]). The advantages of MRFs have also been outlined in [16], [17], [31], [32] for hyperspectral image analysis and in [11] for spectral unmixing. Considering two *sites* of a given lattice (e.g., two image pixels) with coordinates i and j, the neighborhood relation between these two sites must be symmetric: if i is a neighbor of j then j is a neighbor of i. In image analysis, this neighborhood relation is applied to the nearest pixels depending on the neighborhood structure, for example the fourth, eighth or twelfth nearest pixels. Once the neighborhood structure has been established, we can define the MRF. Let z_p denote a random variable associated with the pth site of a lattice (having P sites). The variables $z_1, \ldots, z_P$ (indicating site classes) take their values in a finite set $\{1, \ldots, K\}$ where K is the number of possible classes. The whole set of random variables $\{z_1, \ldots, z_P\}$ forms a random field. An MRF is then defined when the conditional distribution of z_i given the other sites is positive for every z_i and if it only depends on its neighbors $\boldsymbol{z}_{\mathcal{V}(i)}$, i.e.,

$$f\left(z_i | \boldsymbol{z}_{\text{-}i}\right) = f\left(z_i | \boldsymbol{z}_{\mathcal{V}(i)}\right) \tag{5}$$

where $\mathcal{V}(i)$ represents the set of neighbors and $\boldsymbol{z}_{\text{-}i} = \{z_j; j \neq i\}$. In the case of a Potts-Markov model, given a discrete random field $\boldsymbol{z}$ attached to an image with P pixels, the Hammersley-Clifford theorem yields the joint probability density function of $\boldsymbol{z}$

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

where β is the *granularity* coefficient, $G(\beta)$ is the normalizing constant or *partition function* and $\delta(\cdot)$ is the Kronecker function ($\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ otherwise). Note that

drawing a label vector $\boldsymbol{z} = [z_1, \ldots, z_P]$ from the distribution (6) can be easily achieved without knowing $G(\beta)$ by using a Gibbs sampler [11]. The hyperparameter β tunes the degree of homogeneity of each region in the image. As illustrated in [11], the value of β has an influence on the number and the size of the regions. Moreover, its value clearly depends on the neighborhood structure [33]. Note that it is often unnecessary to consider values of $\beta \geq 2$ for the 1st-order neighborhood structure, as mentioned in [34, p. 237].

In this paper, we propose an MRF depending on new lattice and neighborhood structures. More precisely, our set of sites is composed with the similarity regions built by the area filter. These regions are successively indexed in the preprocessing step. We introduce the following binary relation $\leq$ to define the partially ordered set (poset) composed with the similarity regions $\{\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_S\}$: if $s \leq t$ then we assume $\boldsymbol{\Omega}_s \leq \boldsymbol{\Omega}_t$. For obvious reason, this binary relation has the reflectivity, antisymmetry and transitivity properties necessary for the definition of the poset. It is also straightforward to see that for any subset of $\{\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_S\}$, a supremum (join) and an infimum (meet) exist. For this reason, the poset $\{\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_S\}$ is a lattice allowing the similarity regions to be used as sites for a neighborhood structure. This neighborhood structure is based upon the square distance between the corresponding median vector which is compared to a given threshold. In other terms, $\boldsymbol{\Omega}_s$ and $\boldsymbol{\Omega}_t$ are neighbors if the relation $D_{s,t} = \|\boldsymbol{\Upsilon}_s - \boldsymbol{\Upsilon}_t\|^2 \leq \tau$ is fulfilled[1], where τ is a fixed value. By denoting $\mathcal{V}_\tau(s)$ the set of regions that are neighbors of $\boldsymbol{\Omega}_s$ and by associating a random discrete hidden variable z_s to every similarity region $\boldsymbol{\Omega}_s$, the following relation can be easily established $f(z_s | \boldsymbol{z}_{\text{-}s}) = f(z_s | \mathcal{V}_\tau(s))$, thus implying that the set of labels z_s is an MRF with

$$P(z_s = k | \boldsymbol{z}_{\text{-}s}) \propto \exp\left[\beta \sum_{t \in \mathcal{V}_\tau(t)} \delta(z_s - z_t)\right] \tag{7}$$

where $\propto$ means "proportional to".

III. HIERARCHICAL BAYESIAN MODEL

This section studies a Bayesian model based on the adaptive MRF introduced in the previous Section. The unknown parameter vector of this model is denoted as $\boldsymbol{\Upsilon} = \{\boldsymbol{A}, \boldsymbol{z}, \sigma^2\}$, where σ^2 is the noise variance, $\boldsymbol{z}$ contains the labels associated with the similarity regions and $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_P]$ is the abundance matrix with $p = 1, \ldots, P$ and $\boldsymbol{a}_p = [a_{1,p}, \ldots, a_{R,p}]^T$.

A. Likelihood

Since the additive noise in (1) is white, the likelihood function of the pth pixel $\boldsymbol{y}_p$ is

$$f\left(\boldsymbol{y}_p | \boldsymbol{a}_p, \sigma^2\right) \propto \frac{1}{\sigma^L} \exp\left[-\frac{\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_p\|^2}{2\sigma^2}\right]. \tag{8}$$

By assuming independence between the noise vectors $\boldsymbol{n}_p$, the image likelihood is

$$f\left(\boldsymbol{Y} | \boldsymbol{A}, \sigma^2\right) = \prod_{p=1}^{P} f\left(\boldsymbol{y}_p | \boldsymbol{a}_p, \sigma^2\right). \tag{9}$$

[1] $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T \boldsymbol{x}}$ is the standard ℓ_2 norm.

B. Parameter priors

This section defines the prior distributions of the unknown parameters and their associated hyperparameters that will be used for the LMM.

1) Label prior: The prior distribution for the label z_s is the Potts-Markov random field whose distribution is given in (7). Using the Hammersley-Clifford theorem, we can show that the joint prior distribution associated with the label vector $\boldsymbol{z} = [z_1, \ldots, z_S]^T$ is also a Potts-Markov random field (see Appendix), i.e.,

$$P(\boldsymbol{z}) \propto \exp \left[\beta \sum_{s=1}^{S} \sum_{t \in \mathcal{V}_\tau(t)} \delta(z_s - z_t) \right] \qquad (10)$$

with a known granularity coefficient β (fixed *a priori*).

2) Abundance prior distribution: The abundance vectors have to satisfy the positivity and sum-to-one constraints. This paper proposes to use Dirichlet prior distributions for these vectors. More precisely, the prior distribution for the abundance $\boldsymbol{a}_p$ is defined conditionally upon its class

$$\boldsymbol{a}_p | z_s = k, \boldsymbol{u}_k \sim \mathcal{D}_R(\boldsymbol{u}_k) \qquad (11)$$

where $\mathcal{D}_R(\boldsymbol{u}_k)$ is the Dirichlet distribution with parameter vector $\boldsymbol{u}_k = (u_{1,k} \ldots, u_{R,k})^T$. Note that the vector $\boldsymbol{u}_k$ depends on the region defined by pixels belonging to class k. Assuming independence between the abundance vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_P$, the joint abundance prior is

$$f(\boldsymbol{A}|\boldsymbol{z}, \boldsymbol{U}) = \prod_{k=1}^{K} \prod_{\boldsymbol{\Omega}_s \in \mathcal{C}_k} \prod_{p \in \boldsymbol{\Omega}_s} f(\boldsymbol{a}_p | z_s = k, \boldsymbol{u}_k) \qquad (12)$$

with $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K]$.

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

$$\sigma^2 | \nu, \delta \sim \mathcal{IG}(\nu, \delta) \qquad (13)$$

where ν and δ are adjustable hyperparameters. This paper assumes $\nu = 1$ (as in [8]) and estimates δ jointly with the other unknown parameters and hyperparameters.

C. Hyperparameter priors

Hierarchical Bayesian algorithms can be used to estimate the hyperparameters defining the parameter priors. These algorithms require to define prior distributions for the hyperparameters (sometimes referred to as hyperpriors). The values of the vectors $\boldsymbol{u}_k$ are important for a correct description of the classes, since the mean vector $\boldsymbol{\mu}_k$ and the covariance matrix $\boldsymbol{\Lambda}_k$ defined in (4) explicitly depend on these vectors. The lack of prior information for these hyperparameters leads us to choose an improper uniform distribution on the interval $\mathbb{R}^+$. Since these parameters are independent, the joint prior distribution is

$$f(\boldsymbol{U}) = \mathbf{1}_{\mathbb{R}_+^{RK}}(\boldsymbol{U}) \qquad (14)$$

where $\mathbf{1}_{\mathbb{R}_+}(\cdot)$ denotes the indicator function defined on $\mathbb{R}^+$. The noise hyperparameter δ has been assigned a non-informative Jeffreys' prior (see [35, p. 131] for motivations)

$$f(\delta) \propto \frac{1}{\delta} \mathbf{1}_{\mathbb{R}_+}(\delta). \qquad (15)$$

At this last hierarchy level within the Bayesian inference, the hyperparameter vector can be defined as $\boldsymbol{\Gamma} = \{\boldsymbol{U}, \delta\}$.

D. Joint distribution

The joint posterior of the unknown parameter and hyperparameter vector $(\boldsymbol{\Theta}, \boldsymbol{\Gamma})$ can be obtained from the hierarchical Bayesian model associated with the directed acyclic graph (DAG) depicted in Fig. 1

$$f(\boldsymbol{\Theta}, \boldsymbol{\Gamma}|\boldsymbol{Y}) = f(\boldsymbol{Y}|\boldsymbol{\Theta})f(\boldsymbol{\Theta}|\boldsymbol{\Gamma})f(\boldsymbol{\Gamma}). \qquad (16)$$

Straightforward computations lead to

$$f(\boldsymbol{\Theta}, \boldsymbol{\Gamma}|\boldsymbol{Y}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{LP}{2}} \prod_{p=1}^{P} \exp\left[-\frac{\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_p\|^2}{2\sigma^2}\right]$$

$$\times \exp\left[\beta \sum_{t=1}^{S} \sum_{t \in \mathcal{V}_\tau(t)} \delta(z_s - z_t)\right]$$

$$\times \frac{\delta^{\nu-1}}{(\sigma^2)^{\nu+1}} \prod_{k=1}^{K} \prod_{\boldsymbol{\Omega}_s \in \mathcal{C}_k} \prod_{p \in \boldsymbol{\Omega}_s} \left[\frac{\Gamma(u_{0,k})}{\prod_{r=1}^{R} \Gamma(u_{r,k})}\right.$$

$$\left. \times \prod_{r=1}^{R} a_{r,p}^{u_{r,k}-1} \mathbf{1}_{\mathcal{S}}(\boldsymbol{a}_p)\right] \mathbf{1}_{\mathbb{R}_+^{RK}}(\boldsymbol{U}) \qquad (17)$$

where $u_{0,k} = \sum_{r=1}^{R} u_{r,k}$, $\Gamma(.)$ is the gamma function and $\mathcal{S}$ is the simplex defined by the sum-to-one and positivity constraints. This distribution is far too complex to obtain closed-form expressions for the MMSE or MAP estimators of $(\boldsymbol{\Theta}, \boldsymbol{\Gamma})$. Thus, we propose to use MCMC methods for generating samples asymptotically distributed according to (17). By excluding the first N_{bi} generated samples (belonging to the so-called burn in period), it is then possible to approximate the MMSE and MAP estimators from the remaining samples.
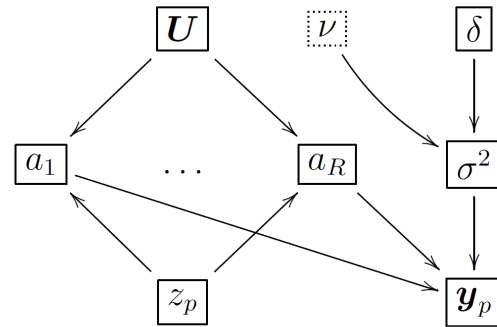

Fig. 1. DAG for the parameter priors and hyperpriors (the fixed parameters appear in dashed boxes).

IV. HYBRID GIBBS SAMPLER

This section studies a hybrid Metropolis-within-Gibbs sampler that iteratively generates samples according to the full conditional distributions of $f(\boldsymbol{\Theta}, \boldsymbol{\Gamma}|\boldsymbol{Y})$. The algorithm is summarized in Algo. 1 and its main steps will now be detailed.

A. Generating samples according to $P[z_s = k|\boldsymbol{z}_{\text{-}s}, \boldsymbol{A}_s, \boldsymbol{u}_k]$

For a given similarity region $\boldsymbol{\Omega}_s$, Bayes' theorem yields the conditional distribution of z_s

$$P[z_s = k|\boldsymbol{z}_{\text{-}s}, \boldsymbol{A}_s, \boldsymbol{u}_k] \propto f(z_s|\boldsymbol{z}_{\text{-}s}) \prod_{p \in \boldsymbol{\Omega}_s} f(\boldsymbol{A}_p|z_s, \boldsymbol{u}_k)$$

where $\boldsymbol{A}_s$ is the abundance matrix associated with the pixels belonging to the neighborhood $\boldsymbol{\Omega}_s$. Since the label of a given neighborhood is the same for all pixels, it makes sense that the abundance vectors of $\boldsymbol{\Omega}_s$ contribute to the conditional distribution of z_s. The complete expression of the conditional distribution is

$$P[z_s = k|\boldsymbol{z}_{\text{-}s}, \boldsymbol{A}_s, \boldsymbol{u}_k] \propto \exp\left[\beta \sum_{t \in \mathcal{V}_\tau(t)} \delta(z_s - z_t)\right]$$
$$\times \prod_{p \in \boldsymbol{\Omega}_s} \frac{\Gamma(u_{0,k})}{\prod_{r=1}^{R} \Gamma(u_{r,k})} \prod_{r=1}^{R} a_{r,p}^{u_{r,k}-1} \mathbf{1}_{\mathcal{S}}(\boldsymbol{a}_p). \quad (18)$$

Note that sampling from this conditional distribution can be achieved by drawing a discrete value in the finite set $\{1, \ldots, K\}$ with the normalized probabilities (18).

B. Generating samples according to $f(\boldsymbol{a}_p|z_s = k, \boldsymbol{y}_p, \sigma^2)$

The Bayes' theorem leads to

$$f(\boldsymbol{a}_p|z_s = k, \boldsymbol{y}_p, \sigma^2) \propto f(\boldsymbol{a}_p|z_s = k, \boldsymbol{u}_k) f(\boldsymbol{y}_p|\boldsymbol{a}_p, \sigma^2)$$

or equivalently to

$$f(\boldsymbol{a}_p|z_s = k, \boldsymbol{y}_p, \sigma^2) \propto \exp\left[-\frac{\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_p\|^2}{2\sigma^2}\right] \mathbf{1}_{\mathcal{S}}(\boldsymbol{a}_p)$$
$$\times \prod_{r=1}^{R} a_{r,p}^{u_{r,k}-1}. \quad (19)$$

Since it is not easy to sample according to (19), we propose to use a Metropolis-Hastings step for generating the $R-1$ first abundance samples and to compute the Rth abundance using $a_{R,p} = 1 - \sum_{r=1}^{R-1} a_{r,p}$. The proposal distribution for this move is a Gaussian distribution with the following mean and covariance matrix (from [8])

$$\begin{cases} \boldsymbol{\Delta} = \left[\frac{1}{\sigma^2}\left(\boldsymbol{M}^* - \boldsymbol{m}_R \boldsymbol{u}^T\right)^T \left(\boldsymbol{M}^* - \boldsymbol{m}_R \boldsymbol{u}^T\right)\right]^{-1}, \\ \boldsymbol{\mu} = \boldsymbol{\Delta}\left[\frac{1}{\sigma^2}\left(\boldsymbol{M}^* - \boldsymbol{m}_R \boldsymbol{u}^T\right)^T \left(\boldsymbol{y}_p - \boldsymbol{m}_R\right)\right], \end{cases} \quad (20)$$

where $\boldsymbol{M}^* = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{R-1}]$ and $\boldsymbol{u} = [1, \ldots, 1]^T \in \mathbb{R}^{R-1}$. This distribution is truncated on the set defined by the abundance constraints (see [36] and [8] for more details).

C. Generating samples according to $f(\sigma^2|\boldsymbol{Y}, \boldsymbol{A}, \delta)$

The conditional distribution of σ^2 is

$$f(\sigma^2|\boldsymbol{Y}, \boldsymbol{A}, \delta) \propto f(\sigma^2|\delta) \prod_{p=1}^{P} f(\boldsymbol{y}_p|\boldsymbol{a}_p, \sigma^2).$$

As a consequence, $\sigma^2|\boldsymbol{Y}, \boldsymbol{A}, \delta$ is distributed according to the following inverse-gamma distribution

$$\sigma^2|\boldsymbol{Y}, \boldsymbol{A}, \delta \sim \mathcal{IG}\left(\frac{LP}{2} + 1, \delta + \sum_{p=1}^{P} \frac{\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_p\|^2}{2}\right). \quad (21)$$

D. Generating samples according to $f(u_{r,k}|\boldsymbol{z}, \boldsymbol{a}_r)$

The Dirichlet parameters are generated for each endmember r $(r = 1, \ldots, R)$ and each class $\mathcal{C}_k$ $(k = 1, \ldots, K)$

$$f(u_{r,k}|\boldsymbol{z}, \boldsymbol{a}_r) \propto f(u_{r,k}) \prod_{\boldsymbol{\Omega}_s \in \mathcal{C}_k} \prod_{p \in \boldsymbol{\Omega}_s} f(\boldsymbol{a}_p|z_s = k, \boldsymbol{u}_k)$$

which leads to

$$f(u_{r,k}|\boldsymbol{z}, \boldsymbol{a}_r) \propto \prod_{\boldsymbol{\Omega}_s \in \mathcal{C}_k} \prod_{p \in \boldsymbol{\Omega}_s} \left[\frac{\Gamma(u_{0,k})}{\Gamma(u_{r,k})} a_{r,p}^{u_{r,k}-1}\right] \mathbf{1}_{\mathbb{R}^+}(u_{r,k}). \quad (22)$$

Since it is not easy to sample from (22), we propose to use a Metropolis-Hastings move. More precisely, samples are generated using a random-walk defined by the Gaussian distribution $\mathcal{N}(0, w^2)$, where the variance w^2 has been adjusted to obtain an acceptance rate between 0.15 and 0.50 as recommended in [37, p. 55].

E. Generating samples according to $f(\delta|\sigma^2)$

The conditional distribution of δ is the following gamma distribution

$$\delta|\sigma^2 \sim \mathcal{G}\left(1, \frac{1}{\sigma^2}\right) \quad (23)$$

where $\mathcal{G}(a, b)$ is the gamma distribution with shape parameter a and scale parameter b [38, p. 581].

Algorithm 1 Hybrid Gibbs sampler for joint unmixing and segmentation

---

- % *Initialization:*
  1: Generate $\boldsymbol{z}^{(0)}$ by randomly assign a discrete value from $(1, \ldots, K)$ to each region $\boldsymbol{\Omega}_s$,
  2: Generate $\boldsymbol{U}^{(0)}$ and $\delta^{(0)}$ from the probability density functions (pdfs) in Eqs. (14) and (15),
  3: Generate $\boldsymbol{A}^{(0)}$ and $\sigma^{2(0)}$ from the pdfs in Eqs. (12) and (13),

- % *Iterations:*
  1: **for** $t = 1, 2, \ldots$ **do**
  2:   **for** each pixel $p = 1, \ldots, P$ **do**
  3:     Sample $\boldsymbol{a}_p^{(t)}$ from the pdf in Eq. (19),
  4:   **end for**
  5:   Sample $\sigma^{2(t)}$ from the pdf in Eq. (21),
  6:   **for** each region $\boldsymbol{\Omega}_s$ $s = 1, \ldots, S$ **do**
  7:     Sample $z_s^{(t)}$ from the pdf in Eq. (18),
  8:   **end for**
  9:   **for** each class $\mathcal{C}_k$ $k = 1, \ldots, K$ **do**
  10:     Sample $u_{r,k}$ from the pdf in Eq. (22),
  11:   **end for**
  12:   Sample δ from the pdf in Eq. (23),
  13: **end for**

TABLE I
ACTUAL AND ESTIMATED ABUNDANCE MEAN AND VARIANCE ($\times 10^{-3}$).

		Actual values	Adaptive-MRF		FCLS	
			M known	VCA	M known	VCA
Class 1	$\mathrm{E}[\boldsymbol{a}_{p,\ p\in\mathcal{I}_1}]$	$[0.6, 0.3, 0.1]^T$	$[0.58, 0.29, 0.13]^T$	$[0.64, 0.21, 0.14]^T$	$[0.58, 0.29, 0.13]^T$	$[0.65, 0.21, 0.13]^T$
	$\mathrm{Var}[a_{p,r,\ p\in\mathcal{I}_1}]$	$[5, 5, 5]^T$	$[4.3, 4.1, 6.3]^T$	$[14.3, 12.2, 11.5]^T$	$[4.5, 4.2, 6.7]^T$	$[15.1, 13, 12.1]^T$
Class 2	$\mathrm{E}[\boldsymbol{a}_{p,\ p\in\mathcal{I}_2}]$	$[0.3, 0.5, 0.2]^T$	$[0.29, 0.49, 0.2]^T$	$[0.13, 0.67, 0.2]^T$	$[0.29, 0.49, 0.2]^T$	$[0.13, 0.67, 0.2]^T$
	$\mathrm{Var}[a_{p,r,\ p\in\mathcal{I}_2}]$	$[5, 5, 5]^T$	$[4.7, 4.8, 8.7]^T$	$[8.7, 14.3, 13.6]^T$	$[5, 4.8, 9.1]^T$	$[10.3, 14.1, 14.6]^T$
Class 3	$\mathrm{E}[\boldsymbol{a}_{p,\ p\in\mathcal{I}_3}]$	$[0.3, 0.2, 0.5]^T$	$[0.3, 0.2, 0.49]^T$	$[0.19, 0.18, 0.63]^T$	$[0.29, 0.2, 0.49]^T$	$[0.17, 0.19, 0.64]^T$
	$\mathrm{Var}[a_{p,r,\ p\in\mathcal{I}_3}]$	$[5, 5, 5]^T$	$[5.1, 4.9, 9.8]$	$[11.2, 11.7, 18.8]^T$	$[5.3, 4.9, 10]^T$	$[12.8, 12.6, 19.8]^T$

V. SIMULATION RESULTS ON SYNTHETIC DATA

The first experiments evaluate the performance of the proposed algorithm for unmixing a 25×25 synthetic image with $K = 3$ different classes. The image contains $R = 3$ mixed components (construction concrete, green grass and micaceous loam) whose spectra have been extracted from the spectral libraries distributed with the ENVI package [39] (these spectra have $L = 413$ spectral bands ranging from wavelength $0.4\mu m$ to $2.5\mu m$, from the visible to the near infrared and are plotted in [40]). The synthetic label map shown in Fig. 2 (left) has been generated using a Potts-Markov random field with a granularity coefficient $\beta = 2$, allowing large and distinct regions to be constructed. The abundance means and variances in each class have been chosen to ensure a single endmember is prominent in a given class. The actual values of these parameters reported in Table I show that the 1st endmember is more present in class 1 (with average concentration of 60%), the 2nd endmember is more present in class 2 (with average concentration of 50%) and the 3rd endmember is more present in class 3 (with average concentration of 50%). All the abundance variances have been fixed to 5×10^{-3}. The abundance maps used to mix the endmembers are depicted in Fig. 3 (top). Note that a white (resp. black) pixel in the fraction map indicates a large (resp. small) value of the abundance coefficient. The noise variance has been chosen in order to have an average signal-to-noise ratio SNR = 20dB, i.e., $\sigma^2 = 0.001$. The similarity regions have been built using the self-complementary area filter with an area criterion $\lambda = 5$. The neighborhoods have been established using a threshold $\tau = 5 \times 10^{-3}$. The proposed sampler has been run with $N_{\mathrm{MC}} = 5000$ iterations including $N_{\mathrm{bi}} = 500$ burn-in iterations. The estimates of the class labels are obtained using the MAP estimator approximated by retaining the samples that maximizes the posterior conditional probabilities of $\boldsymbol{z}$. These estimates depicted in Fig. 2 (right) agree with the actual class labels. After computing the estimated class labels, the abundances have been estimated conditionally to these estimates using the MMSE estimator. The estimated abundances represented in Fig. 3 (bottom) are also in good agreement with the actual abundances. Moreover, the mean vectors $\boldsymbol{\mu}_k$ are of great interest since they characterize each class $\mathcal{C}_k$. For illustration purposes, the histograms of the abundance MMSE estimates belonging to the 2nd class (i.e., for $k = 2$) have

been depicted in Fig. 4. This figure shows that the abundance estimates are concentrated around the actual values of $\boldsymbol{\mu}_2$. Similar results would be obtained for the other classes. They are omitted here for brevity.
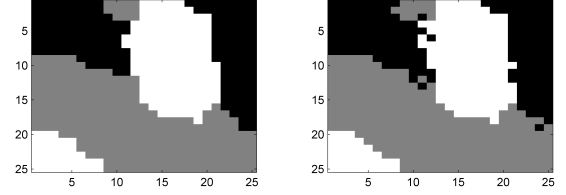

Fig. 2. Left: the actual label map. Right: the label map estimated by the proposed Gibbs sampler.
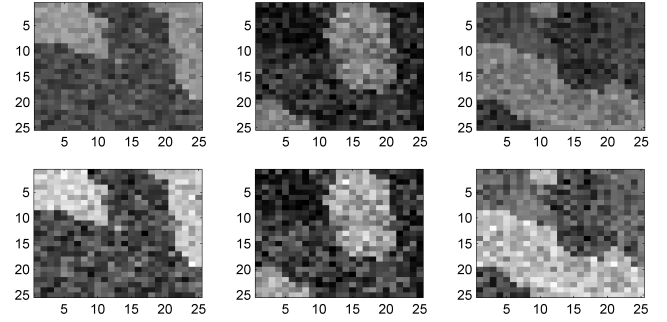

Fig. 3. Top: abundance maps of the 3 pure materials. Bottom: abundance maps of the 3 pure materials estimated by the hybrid Gibbs sampler (from left to right: construction concrete, green grass, micaceous loam).
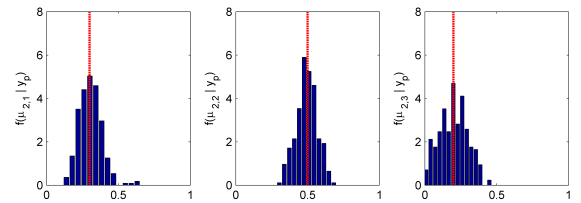

Fig. 4. Histograms of the abundance MMSE estimates for the 2nd class. The dotted lines correspond to the actual abundance means contained in $\boldsymbol{\mu}_2 = [\mu_{2,1}, \mu_{2,2}, \mu_{2,3}]^T$.

The proposed method has been tested on a larger dataset

composed of 50 different images of size 25×25 pixels. For each image, one classification map and one abundance map have been generated as explained above. The algorithm has been run on each image by assuming first that the endmember spectra are known. Table I shows the estimated abundance means and variances for each class obtained by averaging the results from the 50 Monte Carlo runs. The estimated abundances are clearly in accordance with the actual ones. Then, to evaluate the sensibility of the proposed unmixing algorithm with respect to the endmember matrix, we have proposed to use endmembers identified by the vertex component analysis (VCA) [41] instead of the endmembers actually used to generate the data. The estimated abundance means and variances (averaged over the 50 Monte Carlo runs) are reported in Table I. The estimations are slightly less accurate when the endmember spectra have been estimated using VCA. In particular, variances of the estimated abundances are greater when endmembers provided by VCA are used in the unmixing process. This behavior of the proposed unmixing technique is compared with the behavior of the FCLS algorithm when used in the same two scenarios (M known in one scenario and M estimated by VCA in the other). Similarly, one can notice that the abundance mean variances increase when using FCLS with estimated endmembers.

The proposed spatial hybrid Gibbs sampler has been compared with its "local" MRF counterpart developed in [11] and with the non-spatial Bayesian algorithm developed in [8]. As a performance criterion, the global mean square errors (MSEs) of the estimated abundances have been computed. The global MSE for the rth abundance is defined as

$$\text{MSE}_r^2 = \frac{1}{P} \sum_{p=1}^{P} (\hat{a}_{r,p} - a_{r,p})^2 \qquad (24)$$

where $\hat{a}_{r,p}$ denotes the MMSE estimate of the abundance $a_{r,p}$. The obtained results are reported in Table II with the corresponding computation times. The algorithm developed in this paper (referred to as "Adapt.-MRF") performs similarly or better than the two other algorithms (referred to as "Local-MRF" and "Bayesian") in terms of global MSE. However, the proposed algorithm shows the lowest computational time which is a very interesting property[2]. By assuming $P \gg SK$, the computational complexity of the proposed algorithm is of $\mathcal{O}(N_{\text{MC}}P)$. As a comparison, the previously developed spatial algorithm in [11] has a larger computational complexity of $\mathcal{O}(N_{\text{MC}}PK)$. Note that the pre-processing step required by the "Adapt.-MRF" method has been included in the time evaluation.

VI. REAL AVIRIS HYPERSPECTRAL IMAGE

A. Moffett Field

This section considers a real hyperspectral image of size 50×50 (available in [40]) to evaluate the performance of the different algorithms. This image has been extracted from a larger image acquired in 1997 by the Airborne Visible Infrared

---

[2]These simulations have been run on an unoptimized MATLAB 7.1 64bit implementation on a Core(TM)2Duo 2.93GHz computer.

TABLE II
GLOBAL MSEs OF EACH ABUNDANCE COMPONENT AND EXECUTION
TIMES FOR THE THREE UNMIXING ALGORITHMS.

	Bayesian	Local-MRF	Adapt.-MRF
MSE_1^2	5.3×10^{-3}	3.4×10^{-4}	3.2×10^{-4}
MSE_2^2	5.4×10^{-3}	9.5×10^{-5}	9.5×10^{-5}
MSE_3^2	2.3×10^{-4}	2.4×10^{-4}	2.3×10^{-4}
Time (sec.)	4.6×10^3	2×10^3	1.6×10^3

Imaging Spectrometer (AVIRIS) over Moffett Field, CA and has been used intensively in the geoscience community [8], [11], [42], [43]. The data set has been reduced from the original 224 bands to $L = 189$ bands by removing water absorption bands. First, the image has been pre-processed by PCA to determine the number of endmembers present in the scene, as explained in [1] and applied in [11]. Note that several other techniques could be used to perform such preprocessing step. For example, the number of endmembers could be estimated by using the minimum noise fraction (MNF) method [44], the HySime algorithm [45] or other strategies exploiting virtual dimensionality, as in [46] or [47]. Then, the N-FINDR algorithm, proposed by Winter in [6] has been used to estimate the endmember spectra. The $R = 3$ extracted endmembers shown in [11] correspond to vegetation, water and soil, and have been used as the mean vectors m_1, m_2 and m_3. The proposed algorithm has been applied to this image with a number of classes being $K = 4$ and $N_{\text{MC}} = 5000$ iterations (with 500 burn-in iterations). The number of classes has been fixed to $K = 4$ since prior knowledge on the scene allows one to identify 4 areas in the image: water point, lake shore, vegetation and soil. The minimum flat zone size and the threshold of the neighborhood distance have been respectively fixed to $\lambda = 10$ and $\tau = 0.005$.

The estimated classification and abundance maps for the proposed hybrid Gibbs algorithm are depicted in Figs. 5 (left) and 6 (top). The results provided by the algorithm are very similar to those obtained with its "local" MRF counterpart [11], as shown in Figs. 5 (right) and 6 (bottom).
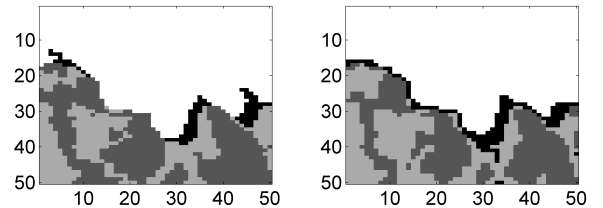

Fig. 5. Label map estimated by the proposed algorithm (left) and by the "local" MRF algorithm (right).

We have also compared the reconstruction error (RE) and the spectral angle mapper (SAM) [1] for the proposed algorithm and two classical unmixing algorithms: the FCLS algorithm and the"local" MRF algorithm [11]. The reconstruction
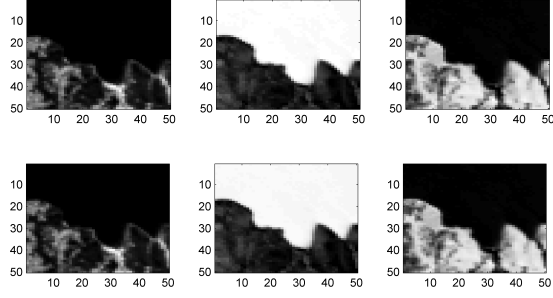
Fig. 6. Abundance maps estimated by the proposed algorithm (top) and the "local" MRF algorithm (bottom). From left to right: vegetation, water and soil.

error has been computed using the following relation

$$\text{RE} = \sqrt{\frac{1}{PL} \sum_{p=1}^{P} \|\boldsymbol{y}_p - \hat{\boldsymbol{y}}_p\|^2} \qquad (25)$$

where the number of pixels is $P = 2500$ for the region considered in our experiments, L is the number of spectral bands (note that the water absorption bands have been removed) and $\boldsymbol{y}_p$ and $\hat{\boldsymbol{y}}_p$ are the observed and estimated spectra for the pth pixel. The spectral angle mapper used in [1] as an unmixing performance criterion is given below

$$\text{SAM} = \frac{1}{P} \sum_{p=1}^{P} \theta\left(\boldsymbol{y}_p, \hat{\boldsymbol{y}}_p\right) \qquad (26)$$

where

$$\theta\left(\boldsymbol{y}_p, \hat{\boldsymbol{y}}_p\right) = \arccos\left(\frac{\langle \boldsymbol{y}_p, \hat{\boldsymbol{y}}_p \rangle}{\|\boldsymbol{y}_p\|\|\hat{\boldsymbol{y}}_p\|}\right)$$

and $\arccos(\cdot)$ is the inverse cosine function. The estimated mixing matrix used in all the experiments has been computed using the N-FINDR procedure. The results reported in Table III show that the proposed hybrid Gibbs sampler is less computationally intensive than its "local" counterpart with similar performance. Note that the FCLS algorithm shows the lowest execution time with similar performances. However, the proposed algorithm provides a classification map in addition to the abundance estimates contrary to FCLS algorithm. Moreover, the samples generated by the MCMC method can also provide confidence intervals for the estimates or probability of presence in some parts of the image [9].

TABLE III
PERFORMANCE COMPARISON BETWEEN NON-SPATIAL AND SPATIAL BASED UNMIXING ALGORITHMS.

	RE ($\times 10^{-2}$)	SAM ($\times 10^{-3}$)	Time (s.)
FCLS	1.63	15.07	0.388
local-MRF	1.66	15.07	8.4×10^3
adaptive-MRF	1.63	15.16	6.3×10^3

B. Influence of τ and λ

The influence of the intrinsic MRF parameters τ and λ on the estimation performance deserves to be studied. For each

simulation scenario, we have changed one of the mentioned parameters while keeping the others fixed to the values given previously. Firstly, the influence of the neighborhood distance parameter τ has been tested with the two extremum values $\tau = 5$ and $\tau = 5 \times 10^{-5}$. With $\tau = 5$, the neighborhood system is very tolerant since the similarity regions are accepted very easily as neighbors. Conversely, the threshold value $\tau = 5 \times 10^{-5}$ ensures a stricter neighborhood system. The resulting label and abundance maps for $\tau = 5$ and 5×10^{-5} are respectively given in Figs. 7 and 8. Reducing the parameter τ does not change the classification results. Conversely, when $\tau = 5$, the resulting label map suffers from under-segmentation, as the continental part of the image globally belongs to a single class. Note that the abundance maps do not change significantly for the two values of τ, which means that the abundance estimation is relatively unaffected by the precision chosen for the classification. However, this parameter must be chosen carefully to avoid under-segmentation effects.
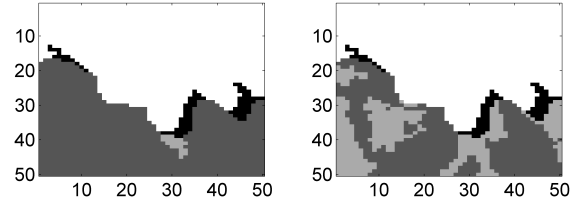

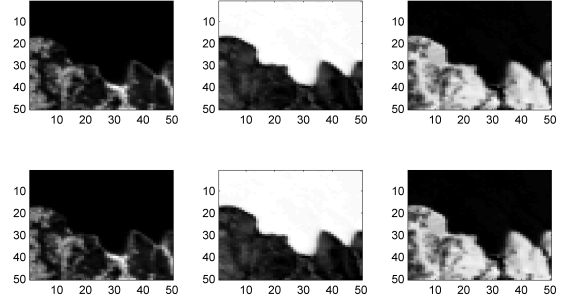Fig. 7. Classification maps obtained with $\tau = 5$ (left) and $\tau = 5 \times 10^{-5}$ (right).

Fig. 8. Abundance maps obtained with $\tau = 5$ (top) and $\tau = 5 \times 10^{-5}$ (bottom).

The influence of the area filtering parameter λ has also been studied by applying the proposed method with $\lambda = 5$ and $\lambda = 20$, i.e., allowing smaller and larger similarity regions to be built. Label maps are shown in Fig. 9 (for brevity the abundance maps have not been depicted since they do not change for the two values of λ). For $\lambda = 5$, the classification results are slightly different from those in Fig. 5, especially for the "shore" class (black) that can be found also in the continental part of some regions. Indeed, the area size being smaller, the self-complementary area filter tends to build much precise flat zones. In the classification map obtained for $\lambda = 20$, the "shore" class is essentially present in the "vegetation band" in the middle of the continental part. This

can be explained by the area filter behavior (resulting from this threshold value) building large flat zones that could remove smaller structures of interests, such as the "shore" regions.
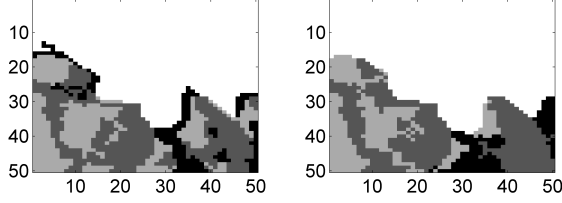
Fig. 9. Classification maps obtained with $\lambda = 5$ (left) and $\lambda = 20$ (right).

C. Influence of R and K

The performance of the proposed algorithm has finally been evaluated for different values of the number of endmembers R and the number of classes K. First, the proposed algorithm has been applied on the Moffett field image with larger numbers of endmembers ($R = 4$ and $R = 5$) and a number of classes fixed to $K = 4$ (as in the previous paragraph). The estimated classification and abundance maps obtained with these two different values of R are depicted in Fig. 10 (middle and right) and 11, respectively. When increasing the number of endmembers, the corresponding label maps do not change significantly. With $R = 5$, the "soil" abundance map has been split into different materials. However, this fact is not represented in the corresponding label map which is similar to the label map in Fig. 5 (left).
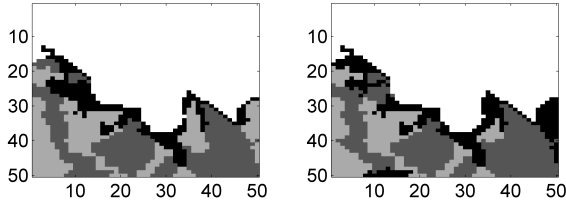
Fig. 10. Classification maps obtained with $R = 4$ (left) and $R = 5$ (right).

The influence of the number of classes K has been investigated by applying the proposed algorithm on the Moffett image with larger numbers of classes ($K = 5$ and $K = 10$) and a number of endmembers equal to $R = 3$ (as in the previous paragraph). The estimated classification and abundance maps obtained with these two different values of K are given in Fig. 12 (middle and right) and 13, respectively. Increasing the number of classes has no influence on the abundance maps results. However, the estimated label maps suffer from over-segmentation, especially for $K = 10$, where nine classes are describing the continental part of the image.

D. Cuprite Image

This section evaluates the performance of the proposed method on a larger real hyperspectral image. The image has been extracted from the AVIRIS Cuprite scene, acquired over a
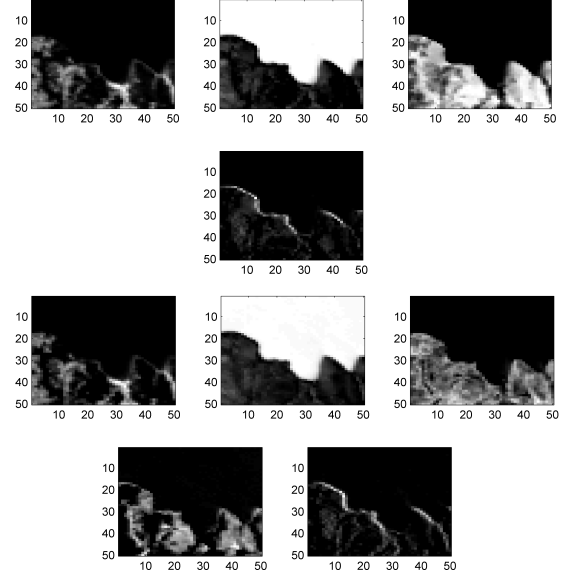
Fig. 11. Abundance maps obtained with $R = 4$ (4 top maps) and $R = 5$ (5 bottom maps).
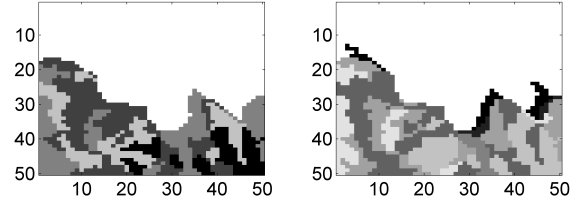
Fig. 12. Classification maps obtained for $R = 3$ with $K = 5$ (left) and $K = 10$ (right).

mining site in Nevada, in 1997. The geological characteristics of the complete data have been mapped in [48], [49]. The area of interest of size 190×250 has been previously studied in [41] for testing the VCA algorithm with $R = 14$ and in [11] using the "local" counterpart of our proposed method with $K = 14$. Therefore, these values have been respectively chosen for R and K. The endmembers used in this experiment have been extracted using the VCA algorithm (see [11] for
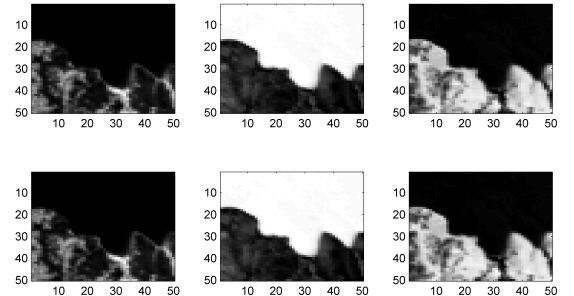
Fig. 13. Relevant abundance maps obtained with $K = 5$ (top) and $K = 10$ (bottom).

more details). Due to the scale of this image, the area filtering parameter λ has been set to 50 while τ has not been changed, i.e., $\tau = 5 \times 10^{-3}$. The proposed algorithm has been used to estimate the abundance and label maps related to the analyzed scene. These maps are depicted in Fig. 14 and 15, respectively. The extracted abundance maps agree with the ones previously obtained in [11]. Moreover, when compared with the classification map obtained in [11], the estimated classification map recovers the general shape of the area and limits the number of "isolated" regions.
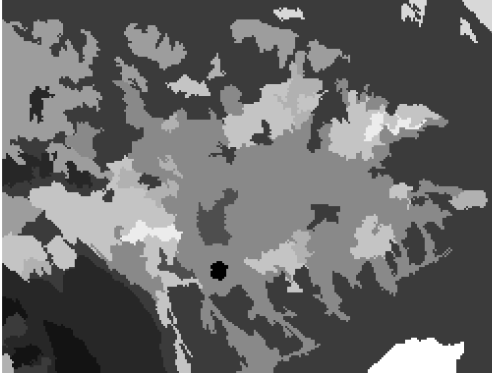

Fig. 14. Classification map obtained on the 190×250 Cuprite area with $K = 14$.

VII. Conclusions

A joint unmixing and segmentation algorithm based on a new Markov random field (MRF) has been introduced. The sites of this MRF have been built using a morphological self-complementary area filter. The resulting similarity regions partition the image into multiple classes that are characterized by close abundances sharing the same means and covariances. A Bayesian model based on this new MRF and on ideas presented in a previous study was derived. The complexity of this Bayesian model was alleviated by implementing a hybrid Gibbs sampler generating data asymptotically distributed according to the posterior distribution of interest. Simulations conducted on synthetic and real hyperspectral data showed that the proposed algorithm achieved similar classification and unmixing performance than its "local" MRF counterpart at the price of a reduced computational time. The focus of future works will be to investigate the generalization ability of the self-complementary area filter to multiple bands by using a method inspired by [50]. Another way of defining the neighbors of each site (using other criteria than the median of the corresponding similarity region) will also be explored. Finally, a fully Bayesian approach jointly estimating the endmember matrix, the similarity regions and the other parameters of interest would deserve to be investigated.

VIII. Acknowledgements

The authors would like to thank M. Fauvel from the University of Toulouse for providing us with the adaptive neighborhood building codes.

IX. Appendix: label prior distribution

The Hammersley-Clifford theorem [34, p. 231] yields

$$P(\boldsymbol{z}) \propto \prod_{s=1}^{S} \frac{P(z_s | z_1^*, \ldots, z_{s-1}^*, z_{s+1}, \ldots, z_S)}{P(z_s^* | z_1^*, \ldots, z_{s-1}^*, z_{s+1}, \ldots, z_S)}$$

where the labels marked with a star z_s^* are arbitrary auxiliary variables. As a consequence

$$
\begin{aligned}
P(\boldsymbol{z}) \quad \propto \quad & \exp \Bigg\{ \beta \sum_{s=1}^{S} \Bigg[\Bigg(\sum_{t \in \mathcal{V}_\tau(t), t < s} \delta(z_s - z_t^*) \\
& + \sum_{t \in \mathcal{V}_\tau(t), t > s} \delta(z_s - z_t) \Bigg) \\
& - \Bigg(\sum_{t \in \mathcal{V}_\tau(t), t < s} \delta(z_s^* - z_t^*) + \sum_{t \in \mathcal{V}_\tau(t), t > s} \delta(z_s^* - z_t) \Bigg) \Bigg] \Bigg\}.
\end{aligned}
$$

By interverting the indexes s and t, we have

$$\sum_{s=1}^{S} \sum_{t \in \mathcal{V}_\tau(t), t < s} \delta(z_s - z_t^*) = \sum_{s=1}^{S} \sum_{t \in \mathcal{V}_\tau(t), t > s} \delta(z_s^* - z_t). \quad (27)$$

This allows us to write the final expression

$$P(\boldsymbol{z}) \propto \exp \left[\beta \sum_{s=1}^{S} \sum_{t \in \mathcal{V}_\tau(t)} \delta(z_s - z_t) \right]. \quad (28)$$

References

[1] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, pp. 44–56, Jan. 2002.

[2] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, and P. Gader, "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, 2012, to appear.

[3] C.-I Chang, X.-L Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. and Remote Sensing*, vol. 36, no. 3, pp. 898–912, May 1998.

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

[5] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, vol. 3, Boston, MA, Jul. 2008, pp. 250–253.

[6] M. E. Winter, "Fast autonomous spectral endmember determination in hyperspectral data," in *Proc. 13th Int. Conf. on Applied Geologic Remote Sensing*, vol. 2, Vancouver, April 1999, pp. 337–344.

[7] D. C. Heinz and C.-I Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 39, no. 3, pp. 529–545, March 2001.

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

[9] 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 Processing*, vol. 19, no. 6, pp. 1403–1413, Jun 2010.

[10] J. T. Kent and K. V. Mardia, "Spatial classification using fuzzy membership models," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 10, no. 5, pp. 659–671, 1988.

[11] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. and Remote Sensing*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.

[12] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc. Ser. B*, vol. 36, no. 2, pp. 192–236, 1974.

[13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.
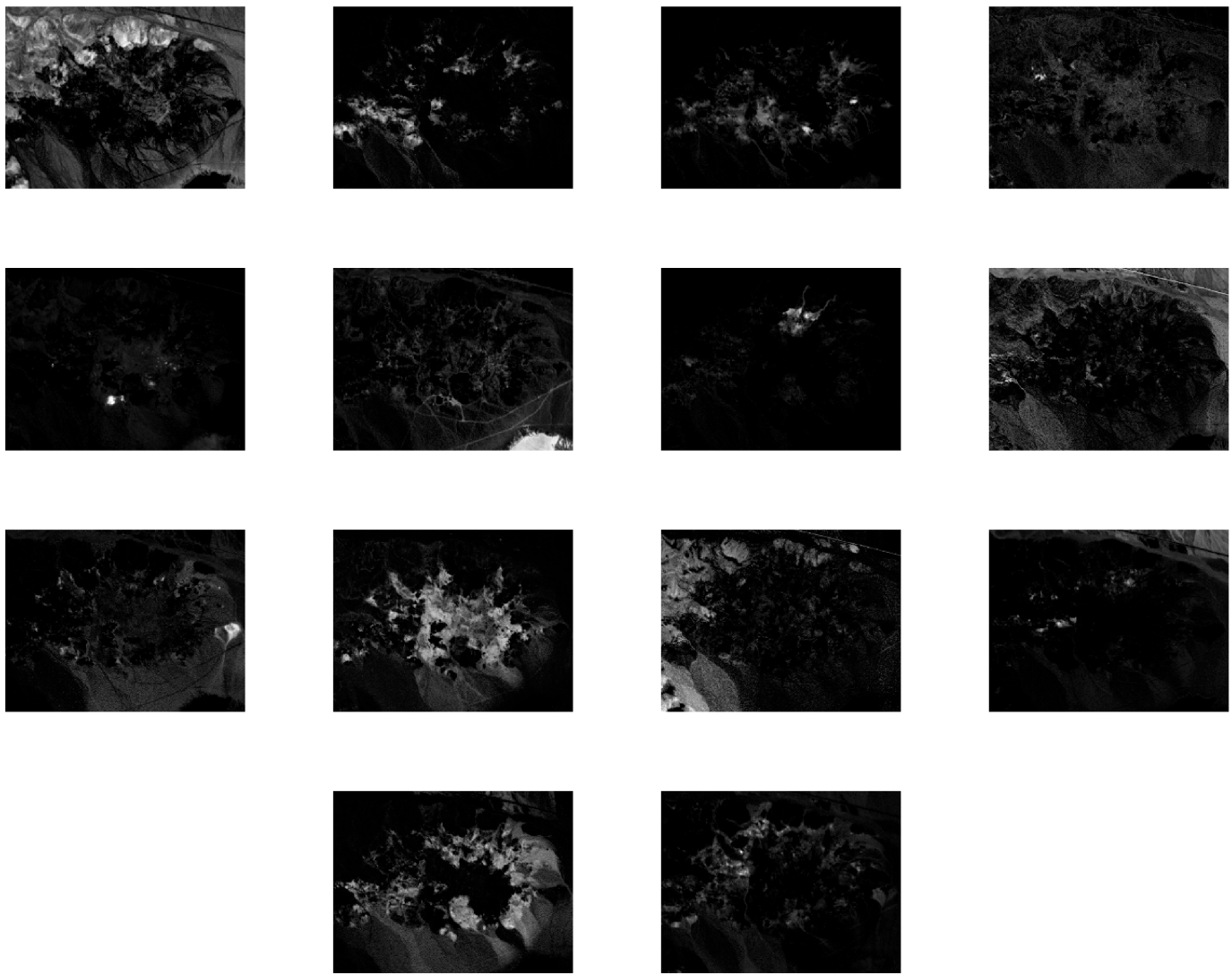
Fig. 15. Fraction maps of the 190×250 Cuprite area.

[14] F. Arduini, C. Dambra, and C. S. Regazzoni, "A coupled MRF model for SAR image restoration and edge-extraction," in *Proc. IEEE Int. Conf. Geoscience and Remote Sensing*, vol. 2, 1992, pp. 1120–1122.

[15] K. Held, E. Kops, J. Krause, W. Wells, R. Kikinis, and H. Muller-Gartner, "Markov random field segmentation of brain mr images," *IEEE Trans. Med. Imag.*, vol. 16, no. 6, pp. 878–886, Dec. 1997.

[16] N. Bali and A. Mohammad-Djafari, "Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis," *IEEE Trans. Image Processing*, vol. 17, no. 2, pp. 217–225, Feb. 2008.

[17] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia, "Texture feature analysis using a Gauss-Markov model in hyperspectral image classification," *IEEE Trans. Geosci. and Remote Sensing*, vol. 42, no. 7, pp. 1543–1551, July 2004.

[18] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. and Remote Sensing Letters*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[19] O. Féron and A. Mohammad-Djafari, "Image fusion and unsupervised joint segmentation using a HMM and MCMC algorithms," *J. Electron. Imaging*, vol. 14, no. 2, May 2005.

[20] J. Serra, *Image Analysis and Mathematical Morphology vol. I*. New-York: Academic Press, 1982.

[21] ——, *Image Analysis and Mathematical Morphology: Theoretical Advances vol. II*. New-York: Academic Press, 1988.

[22] P. Soille, *Morphological image analysis: principles and applications*. New-York: Springer, 1999.

[23] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, no. 3, pp. 480–491, March 2005.

[24] P. Soille, "Beyond self-duality in morphological image analysis," *Image and Vision Computing*, vol. 23, no. 2, pp. 249–257, June 2005.

[25] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial-spectral kernel based approach for the classification of remote sensing images," *Pattern Recognition*, vol. 45, no. 1, pp. 281–392, 2012.

[26] F. Y. Wu, "The Potts model," *Rev. Modern Phys.*, vol. 54, no. 1, pp. 235–268, Jan. 1982.

[27] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. and Remote Sensing*, vol. 50, no. 3, pp. 863–878, March 2012.

[28] J. Diebolt and E. H. S. Ip., "Stochastic EM: method and application," in *Markov chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London: Chapman & Hall, 1996.

[29] C. Kevrann and F. Heitz, "A Markov random field model-based approach to unsupervised texture segmentation using local and global statistics," *IEEE Trans. Image Processing*, vol. 4, no. 6, pp. 856–862, June 1995.

[30] A. Tonazzini, L. Bedini, and E. Salerno, "A Markov model for blind image separation by a mean-field EM algorithm," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 473–481, Feb. 2006.

[31] R. S. Rand and D. M. Keenan, "Spatially smooth partitioning of hyperspectral imagery using spectral/spatial measures of disparity," *IEEE Trans. Geosci. and Remote Sensing*, vol. 41, no. 6, pp. 1479–1490, June 2003.

[32] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. and Remote Sensing*, vol. 48, no. 11, pp. 4085 –4098, Nov. 2010.

[33] B. D. Ripley, *Statistical inference for spatial processes*. Cambridge: Cambridge University Press, 1988.

[34] J.-M. Marin and C. P. Robert, *Bayesian core: a practical approach to computational Bayesian statistics*. New-York: Springer, 2007.

[35] 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.

[36] N. Dobigeon and J.-Y. Tourneret, "Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex," IRIT/ENSEEIHT/TeSA, France, Tech. Rep., march 2007. [Online]. Available: http://www.enseeiht.fr/~dobigeon/papers/ Dobigeon_TechReport_2007b.pdf

[37] G. O. Roberts, "Markov chain concepts related to sampling algorithms," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London: Chapman & Hall, 1996, pp. 45–57.

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

[39] RSI (Research Systems Inc.), *ENVI User's guide Version 4.0*, Boulder, CO 80301 USA, Sept. 2003.

[40] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical Bayesian algorithm," *IEEE J. Sel. Topics Signal Processing*, vol. 3, no. 3, pp. 582–591, Jun. 2010.

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

[42] E. Christophe, D. Léger, and C. Mailhes, "Quality criteria benchmark for hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, no. 9, pp. 2103–2114, Sept. 2005.

[43] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.

[44] A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. and Remote Sensing*, vol. 26, no. 1, pp. 65–74, Jan. 1994.

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

[46] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 42, no. 3, pp. 608–619, March 2004.

[47] B. Luo, J. Chanussot, S. Douté, and X. Ceamanos, "Empirical automatic estimation of the number of endmembers in hyperspectral images," *IEEE Geosci. and Remote Sensing Letters*, no. 99, 2012, to appear.

[48] R. N. Clark, G. A. Swayze, and A. Gallagher, "Mapping minerals with imaging spectroscopy, U.S. Geological Survey," *Office of Mineral Resources Bulletin*, vol. 2039, pp. 141–150, 1993.

[49] R. N. Clark *et al.*, "Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems," *J. Geophys. Res.*, vol. 108, no. E12, pp. 5–1–5–44, Dec. 2003.

[50] S. Velasco-Forero and J. Angulo, "Supervised ordering in R^p: Application to morphological processing of hyperspectral images," *IEEE Trans. Image Processing*, vol. 20, no. 11, pp. 3301–3308, Nov. 2011.