Toward a Sparse Bayesian Markov Random Field Approach to Hyperspectral Unmixing and Classification

Peng Chen, James Nelson, Jean-Yves Tourneret

Toward a Sparse Bayesian Markov Random Field Approach to Hyperspectral Unmixing and Classification

Peng Chen, James D. B. Nelson, and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*— Recent work has shown that existing powerful Bayesian hyperspectral unmixing algorithms can be significantly improved by incorporating the inherent local spatial correlations between pixel class labels via the use of Markov random fields. We here propose a new Bayesian approach to joint hyperspectral unmixing and image classification such that the previous assumption of stochastic abundance vectors is relaxed to a formulation whereby a common abundance vector is assumed for pixels in each class. This allows us to avoid stochastic reparameterizations and, instead, we propose a symmetric Dirichlet distribution model with adjustable parameters for the common abundance vector of each class. Inference over the proposed model is achieved via a hybrid Gibbs sampler, and in particular, simulated annealing is introduced for the label estimation in order to avoid the local-trap problem. Experiments on a synthetic image and a popular, publicly available real data set indicate the proposed model is faster than and outperforms the existing approach quantitatively and qualitatively. Moreover, for appropriate choices of the Dirichlet parameter, it is shown that the proposed approach has the capability to induce sparsity in the inferred abundance vectors. It is demonstrated that this offers increased robustness in cases where the preprocessing endmember extraction algorithms overestimate the number of active endmembers present in a given scene.**

***Index Terms*— Hyperspectral unmixing, image classification, spatial correlations, Markov random fields (MRFs), Markov chain Monte Carlo (MCMC), simulated annealing.**

I. Introduction

HYPERSPECTRAL imaging devices capture energy across hundreds of spectral bands over the electromagnetic spectrum and are currently of great interest to timely topics in the environmental sciences, space research, defence, and security. Unfortunately, owing to the complex image formation process and the limited spatial resolution of even the most sophisticated current sensors, an individual pixel will often contain a mixture of spectra from more than one material. Hyperspectral unmixing is the process that undoes this mixing and expresses the spectrum of each pixel as a superposed combination of constituent spectra, called the endmembers, weighted by a vector of fractional abundances which represents the proportions of each endmember.

Although some recent research considers nonlinear unmixing, such as [1] and those reviewed in [14], linear models continue to receive much attention. There are two main approaches to the treatment of endmember models. The first are database-led treatments. These presuppose that the user has access to a database of pure signatures that have been captured under the same, or similar, conditions as the data of interest. The second class of approaches are extraction-led and instead assume that the pure signatures are present, and can be accurately extracted, from the data itself. In the latter case, hyperspectral unmixing commonly follows two steps, namely endmember extraction and inversion. In the first step, the constituent spectra of pure materials present in the image are identified using an endmember extraction algorithm such as N-FINDR [46] and vertex component analysis (VCA) [35]. The inversion step is devoted to the abundance estimation. Since it represents proportions, it should be noted that the abundance vector at each pixel is often constrained such that is is nonnegative and sums to one. In other words, the abundance vectors belong to a standard simplex. Consequently many algorithms proposed in the literature are based on constrained optimization techniques such as, for example, the fully constrained least squares [20] and scaled gradient [44] algorithms.

Both endmember extraction and inversion have attracted approaches such as the minimum volume-based algorithms [4], [8], [26], [32], Bayesian approaches [2], [5], [11], [13], [34] and recent sparse regression based methods [22], [23]. In particular, Bayesian approaches provide a powerful framework to model the variability and uncertainty present in the data, endmembers, abundances, and/or their parameters. Furthermore, Bayesian formulations provide a statistically well-principled means to encapsulate and incorporate meaningful prior information, such as the constraints on the abundances, into the modelling process. Although sparse regularisation methods such as variants of generalised and fused Lasso [22], [23], can also be framed in terms of

This work was supported in part by the HYPANEMA ANR Project under Grant ANR-12-BS03-003 and in part by the Project ANR-11-LABX-0040-CIMI as part of the Program ANR-11-IDEX-0002-02. The work of J. Nelson was supported in part by the Engineering and Physical Sciences Research Council under Grant EP/N508470/1 and in part by the Defence Science and Technology Laboratory. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Peter Tay.

P. Chen is with the Department of Statistical Science, University College London, London, WC1E 6BT, U.K.

J. Nelson, deceased, was with the Department of Statistical Science, University College London, London, WC1E 6BT, U.K.

J.-Y. Tourneret is with the University of Toulouse, 31071 Toulouse, France (e-mail: jean-yves.tourneret@enseeiht.fr).

Digital Object Identifier 10.1109/TIP.2016.2622401

probabilistic priors, current algorithmic machinery is purely based on maximum a posteriori. As such, unlike fully Bayesian treatments, it therefore does not easily permit any measure of uncertainty to be expressed in the solution.

In recent years fully-Bayesian approaches to the extraction-based treatment of endmembers have proposed a variety of ways to model the abundances. Perhaps the most common Bayesian strategy has been to use the uniform distribution over the simplex as a prior distribution for the abundance vectors [2], [11], [13]. This is, at least in part, motivated by computational tractability. However, as in [5], [15], and [33], a more natural and "stronger" prior for the abundance vectors is the Dirichlet distribution as it automatically enforces the constraints on the abundances.

In parallel, Markov random fields (MRFs) have enjoyed extensive interest in hyperspectral image classification and segmentation [3], [27], [37], [42]. The theory of MRFs provides a well-founded and convenient means to model the spatial dependencies between important image features such as intensity, labels, and so on. The Markov property captures the local correlations inherent in spatial data. Furthermore, this prior can be naturally and readily incorporated into a wider Bayesian framework and is amenable to a plethora of powerful inferential machinery including Markov chain Monte Carlo from which the full posterior distribution of the underlying state can be recovered. As such, it is no surprise that MRFs are subject to continued investigation in this area [28].

In particular, contemporary approaches to hyperspectral unmixing have begun to investigate the utility of MRFs [38]. Perhaps the most accomplished first example of MRFs for Bayesian unmixing is the work by Eches et al. [16]. Image pixels are classified by assigning discrete labels to them. The Potts model (a special case of MRFs) is employed as a prior distribution for the labels. This takes into account the possible spatial correlations between the neighbouring pixels and favour smoother label configurations. However, for label estimation the Gibbs sampler, which is used for sampling labels from the posterior, suffers from the local-trap problem, i.e., it often becomes trapped in local minima. We will illustrate this problem with a simple experiment in this paper, and propose the use of a simulated annealing within Gibbs sampler to solve it.

The local trap problem is not an uncommon drawback of MCMC samplers, especially those that draw from MRF models. Fortunately, with some care, advanced MCMC methods can help mitigate such issues [29]. The simulated annealing (SA) scheme is a simple and elegant strategy to mitigate the deleterious effects of the trap. It allows the sampler to behave in a more exploratory fashion in the early iterations of the algorithm. It is perhaps no surprise that image analysis provided early examples of its use [19]. More up-to-date work, such as [25], [37], and [47] apply SA-within-Gibbs samplers or the Metropolis algorithm to perform image segmentation.

A. Contributions

In this work, the endmembers are assumed to have been extracted by an endmember extraction algorithm. We propose a new Bayesian model combined with MRFs to perform hyperspectral image unmixing and classification simultaneously. In the proposed model, we introduce labels to classify image pixels into different classes, according to the assumption that there exists a common abundance vector in each class. We then apply the Potts model for the label variables to exploit the spatial correlations between neighbouring pixels. Furthermore, the common abundance vector from each class is assumed to follow the same symmetric Dirichlet distribution with parameters that are constant and that can be tuned freely.

1) Dirichlet Model: The main thematic difference between the proposed work and that of [16] is that our approach assumes a common abundance model (CAM) whereas [16] assumes a stochastic abundance model (SAM). As will be seen in later sections, this fundamental difference significantly simplifies the model, allows us to model the abundances directly (instead of via a reparameterisation), and it leads to faster, and more accurate, estimates of abundances and pixel classes. In particular, unlike SAM, the proposed CAM model affords a closed form expression for the abundances distribution and, as described further in Section II-B and Section III, our model also avoids some other theoretical weaknesses suffered by [16].

2) Towards Sparsity: Bayesian approaches commonly rely on extraction preprocessing methods such as N-FINDR and VCA and others to construct the set of endmembers. Since the number of endmembers is not known a priori it is entirely possible that this could be overestimated. As discussed further in Section III-B and demonstrated by experiments, in contrast to SAM, the proposed model offers robustness to overestimation of the endmember number. This is a direct consequence of the key model difference in the way that the abundances are treated; our Dirichlet model admits the possibility of promoting sparsity among the abundances and therefore represents a significant behavioural difference between the two approaches. In effect, this contribution extends the MCMC frameworks to situations where the number of extracted endmembers is overestimated.

3) Synthetic Experiments: In Section V-B we apply the proposed SA to [16] to show that the convergence benefits are not simply confined to our proposed model. In Section V-C, we compare the proposed CAM approach to the SAM from [16] and show very clear computational performance and accuracy improvements. The experiments therein confirm that the new algorithm is at least one order of magnitude faster, and at least one order of magnitude more accurate, than [16] (cf. Tables II and III).

4) Real Experiments: Experiments on real data in Section VI illustrate that, in contrast to SAM, the proposed method demonstrates stability in the inferred label maps with respect to the number of endmembers (which is not usually known a priori) whereas the label maps of [16] change notably as R is varied. Moreover, unlike SAM, when the estimated number of endmembers is increased, the proposed method is able to perform some variable selection capability by effectively pushing the contributions of some of the redundant endmembers to values that are approximately zero. This key behaviour is simply not available when adopting the SAM approach.

5) Algorithm Convergence: Inference over the proposed model is performed via a Gibbs sampler combined with a Metropolis-Hastings algorithm and, unlike previous related approaches, we invoke an SA within-Gibbs sampler scheme to estimate the label variables whilst taking care to avoid the local-trap problem and incorporate a Gelman-Rubin convergence criterion. Quantitative and qualitative results from experiments on synthetic data and on a publicly available real dataset indicate the performance and speed advantages of the proposed model over [16].

B. Related Work and Scope

An extension of [16] is considered in [15] which (i) uses an adaptive neighbourhood structure in the MRF construction and (ii) uses a Dirichlet distribution to model the abundance means and variances. Our model differs from [15] in that the assumed symmetric Dirichlet distribution has deterministic parameters which do not depend on the classes. Not only does the assumption reflect our lack of knowledge about the proportions of endmembers in each class but also, in the case of its parameters with common value below 1, the Dirichlet distribution prefers a sparse distribution of the abundance vectors and consequently can be used to encourage sparse unmixing. This could benefit scenarios where the number of extracted endmembers is somewhat larger than the actual participating endmembers in a given pixel, i.e., where the image is large and complex or where the endmember extraction algorithms have overestimated the number of endmembers.

Although some of the advances in [15], such as the adaptive Markov random field, could also be incorporated into the proposed model, this is beyond the scope of this work. Instead we base our comparisons on the work in [16] so as to (i) contrast the relative accuracy and computational performance of the proposed CAM assumption against the SAM and (ii) assess the relative utility of the Dirichlet distribution to model abundances, not just (as in [15]) as an elegant means to avoid the reparameterisation of [16], but also to explore the additional robustness and potential sparsifying properties it offers when the number of extracted endmembers is overestimated earlier on in the processing chain.

C. Structure

The remainder of the paper is organized as follows. Section II formulates the problems of hyperspectral unmixing and image classification and then motivates, and provides, a brief introduction of Markov random fields and the Potts model. Sections III and IV describe the proposed model and corresponding inferential machinery. Sections V and VI describe the experiments with synthetic and real data. Finally, we draw conclusions in Section VII.

II. Preliminaries

A. Linear Mixing Model

The *linear mixing model* (LMM) assumes that the expected spectrum of each pixel in a hyperspectral image can be expressed as a linear combination of endmember spectra weighted by the corresponding fractional abundances. Thus given an L-band hyperspectral image consisting of P pixels, the linear mixing model is

$$\boldsymbol{y}_p = \boldsymbol{M}\boldsymbol{a}_p + \boldsymbol{e}_p \tag{1}$$

for $p = 1, ..., P$, where $\boldsymbol{y}_p$ denotes the spectral measurement at pixel p, $\boldsymbol{M} = [\boldsymbol{m}_1, ..., \boldsymbol{m}_R]$ is an $L \times R$ matrix comprising R endmembers, which can be known or estimated from the data, and $\boldsymbol{a}_p = [a_{1p}, ..., a_{Rp}]^T \in \boldsymbol{S}^{R-1}$ is the $R \times 1$ abundance vector that represents the proportions of each endmember and where the unit $(R-1)-$simplex

$$\mathbb{S}^{R-1} = \left\{ \boldsymbol{a}_p \mid a_{rp} \geq 0, r = 1, ..., R, \sum_{r=1}^{R} a_{rp} = 1 \right\}$$

encapsulates the abundance nonnegativity constraint (ANC) and the abundance sum-to-one constraint (ASC), namely

1) $a_{rp} \geq 0,\ r = 1, ..., R$ (nonnegative)
2) $\sum_{r=1}^{R} a_{rp} = 1$ (sum-to-one).

We assume that the errors $\boldsymbol{e}_1, ..., \boldsymbol{e}_P$ are independent and identically distributed according to an L-dimensional normal distribution with zero mean and unknown variance, i.e.,

$$\boldsymbol{e}_p \sim \mathcal{N}(\boldsymbol{0}, s^2 \boldsymbol{I}_L)$$

for $p = 1, ..., P$, where s^2 is an unknown scalar, and $\boldsymbol{I}_L$ is the identity matrix of size L. In addition, we denote the observed hyperspectral image or data as $\boldsymbol{Y} = [\boldsymbol{y}_1, ..., \boldsymbol{y}_P]$ and the corresponding abundance matrix as $\boldsymbol{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_P]$.

B. Image Classification

In this paper the problem of image classification or labelling will be solved jointly with hyperspectral unmixing. Let $\mathcal{S} = \{1, ..., P\}$ index the image pixels or sites and $\mathcal{L} = \{1, ..., K\}$ denotes the set of class labels. In image classification, one attempts to assign a label $z_p \in \mathcal{L}$ to each pixel or site $p \in \mathcal{S}$ of the image, resulting in a set of label variables $\boldsymbol{Z} = \{z_1, ..., z_P\}$, according to some classification rule or assumption. In general, it is required that pixels in each class have some similar characteristics, which are significantly different across classes. Here we present two specific classification rules as follows

Rule A: stochastic abundance model
: Classify pixels into the same class if their abundance vectors share a certain distribution with the same parameters, i.e., set $z_p = k$ if $\boldsymbol{a}_p \sim \text{Distribution}(\boldsymbol{\Lambda}_k)$, where $\boldsymbol{\Lambda}_k \in \{\boldsymbol{\Lambda}_1, ..., \boldsymbol{\Lambda}_K\}$ denotes the set of common parameters for class k.

Rule B: common abundance model
: Classify pixels into the same class if they share the same abundance vector, i.e., set $z_p = k$ if $\boldsymbol{a}_p = \boldsymbol{a}_k$, where $\boldsymbol{a}_k \in \{\boldsymbol{a}_1, ..., \boldsymbol{a}_K\}$ denotes the common abundance for class k.

Note that the above rules are both made in terms of the abundance vectors, but the difference is that the abundance vectors in each class are assumed to be randomly distributed or the same according to Rule A and Rule B respectively.

C. Markov Random Fields

In nature most materials such as, for example water, woods, or grass, tend to exist in clusters or clumps. Consequently, neighbouring pixels in hyperspectral images are fairly likely to have some similar properties such as spectrum, abundance vector, and label.

1) Properties: This spatial correlation can be exploited by various means. In well principled probabilistic frameworks, it is common to capture this notion by the incorporation of a Markov random field (MRF) prior on the image labels $\boldsymbol{Z}$ [28]. For then, we have that $\mathbb{P}(\boldsymbol{Z}) > 0$ and $\mathbb{P}(\boldsymbol{Z}|\boldsymbol{Z}_{-p}) = \mathbb{P}(\boldsymbol{Z}|\boldsymbol{Z}_{\boldsymbol{\mathcal{S}}_p})$ where $\boldsymbol{\mathcal{S}}_p$ is a neighbourhood of pixel, or site, p and $\boldsymbol{Z}_{-p} = \boldsymbol{Z}\backslash\{p\}$ is the label field with site p removed. If $\boldsymbol{Z}$ satisfies these MRF conditions then, as a consequence of the well-known Hammersley-Clifford theorem, that establishes the equivalence between MRFs and Gibbs distributions, we can write $\mathbb{P}(\boldsymbol{Z}) \propto \exp(-T^{-1}U(\boldsymbol{Z}))$ where the constant $T \in \mathbb{R}_+$ is known as the temperature and $U(\boldsymbol{Z})$ is called the prior energy function.

2) Potts Model: The Potts model is a simple and natural way to describe multiple discrete labels in an MRF framework. It can be defined via the full-conditional distribution

$$\mathbb{P}(z_p \mid z_{-p}) \propto \exp\left[\beta \sum_{p' \in \boldsymbol{\mathcal{S}}_p} \delta(z_p - z_{p'})\right] \tag{2}$$

where $\beta > 0$ is called *granularity coefficient*, $C(\beta)$ is the normalizing constant or *partition function*, and $\delta(\cdot)$ is the Kronecker delta function such that $\delta(z) = 1$ if $z = 0$, and $\delta(z) = 0$ otherwise. The granularity coefficient β accounts for the strength of local interactions over the label field $\boldsymbol{Z}$ and plays an important role in the application to image classification. It tunes the degree of homogeneity in the estimated label map so that a small value of β induces a noisy label map with a large number of small regions, while a large value of β leads to a homogeneous label map with few large regions.

Note that sampling from (2) can be easily achieved using a Gibbs sampler [16]. This mainly involves drawing the variables $z_1, ..., z_P$ one by one from the set of labels $\{1, ..., K\}$ with their full-conditional distributions. However, as it will be discussed in Section IV-B, this sampling scheme can be problematic especially when β has a large value.

III. Hierarchical Bayesian Model

Eches et al. [16] considered an MRF-based Bayesian approach to the problems of hyperspectral unmixing and image classification. We start by briefly introducing Eches' approach, and then propose a new MRF-based hierarchical Bayesian model, which is arguably simpler and promotes sparsity of the abundance vectors.

A. Eches' Model: CLRSAM

The core of Eches' approach lies in a reparametrization of the abundance vectors. In particular, each component of the abundance vector $\boldsymbol{a}_p$ is rewritten as a function of another set of random variables $\boldsymbol{t}_p = [t_{1,p}, ..., t_{R,p}]^T$ as follows

$$a_{r,p} = \frac{\exp(t_{r,p})}{\sum_{i=1}^{R} \exp(t_{i,p})} \tag{3}$$

for $r = 1, ..., R$, where $t_{r,p}$ are named the logistic coefficients. This reparametrization can also be viewed as a transformation from $\boldsymbol{t}_p$ to $\boldsymbol{a}_p$, and, although not introduced in this way by the authors, it is the inverse of the so-called *centred logratio* (CLR) transformation which originates from the area of compositional data analysis [39], i.e.,

$$\boldsymbol{t}_p = \mathrm{clr}(\boldsymbol{a}_p) = \left[\ln \frac{a_{1,p}}{g(\boldsymbol{a}_p)}, ..., \ln \frac{a_{R,p}}{g(\boldsymbol{a}_p)}\right] \tag{4}$$

where $g(\boldsymbol{a}_p) = \sqrt[R]{a_{1,p} \cdots a_{R,p}}$. Since the CLR is a one-to-one transformation from the unit $(R-1)$-simplex to a real space, it opens up the possibility of using standard unconstrained multivariate method. Indeed, the components of $\boldsymbol{t}_p$, which depend on the class to which the corresponding pixel belongs, are assumed to be independent and Gaussian distributed

$$t_{r,p} \mid z_p = k, \mu_{r,k}, \sigma^2_{r,k} \sim \mathcal{N}(\mu_{r,k}, \sigma^2_{r,k}).$$

Consequently, the distribution of $a_{r,p}$, as a function of $t_{r,p}$, also depends on the class to which the corresponding pixel belongs. As such it is an example of what we refer to as a SAM: it complies with Rule A presented in Section II-B. Henceforth, putting CLR together with SAM, we shall therefore refer to it as the CLRSAM for convenience. However the distribution of $a_{r,p}$ does not have a closed form, and the assumption of independence between $t_{r,p}$ does not hold theoretically, since the CLR transformation defined in (4) implies that the $t_{r,p}$ should sum to zero. The directed acyclic graph (DAG) of Eches' CLRSAM is depicted in Fig. 1.

B. Proposed Model

We here propose a new Bayesian model based on classification rule B presented in Section II-B, namely the common abundance model. Recall that Rule B indicates that the abundance vectors belonging to the kth class are all equal (to $\widetilde{\boldsymbol{a}}_k$), i.e., that $\boldsymbol{a}_p = \widetilde{\boldsymbol{a}}_k$ for all $p \in \boldsymbol{\mathcal{S}}$ such that $z_p = k$, where $k \in \boldsymbol{\mathcal{L}}$, and thus the linear mixing model becomes

$$\boldsymbol{y}_p \mid z_p = k, \widetilde{\boldsymbol{a}}_k, s^2 \sim \mathcal{N}(\boldsymbol{M}\widetilde{\boldsymbol{a}}_k, s^2\boldsymbol{I}_L)$$

for $p \in \boldsymbol{\mathcal{S}}$. We then avoid the need to reparameterize by assuming the abundance vector $\widetilde{\boldsymbol{a}}_k$ follows a symmetric Dirichlet distribution with constant parameters $[\alpha, ..., \alpha] \in \mathbb{R}^R$ that are equal to each other, namely

$$\begin{aligned} \widetilde{\boldsymbol{a}}_k &\sim \mathrm{Dir}(\alpha, ..., \alpha) \\ f(\widetilde{\boldsymbol{a}}_k) &\propto \prod_{r=1}^{R} \widetilde{a}_{r,k}^{\alpha-1} \end{aligned}$$

with $\widetilde{a}_{r,k} > 0, \forall r, k$, and $\sum_{r=1}^{R} \widetilde{a}_{r,k} = 1, \forall k$, and where α is called the concentration parameter. When α is fixed to 1, the Dirichlet distribution is equivalent to a uniform distribution over the unit $(R-1)$-simplex, and can be regarded as a non-informative prior for $\widetilde{\boldsymbol{a}}_k$. In contrast, values of α below 1 prefer sparse distributions of $\widetilde{\boldsymbol{a}}_k$, i.e., most values within a single sample of $\widetilde{\boldsymbol{a}}_k$ will be close to 0, and the majority of probability mass will be concentrated on a few of the values. This property is potentially advantageous when the set of active endmembers for a given pixel is small compared to those active in the entire image. In practice, α can be tuned or learned either by hand

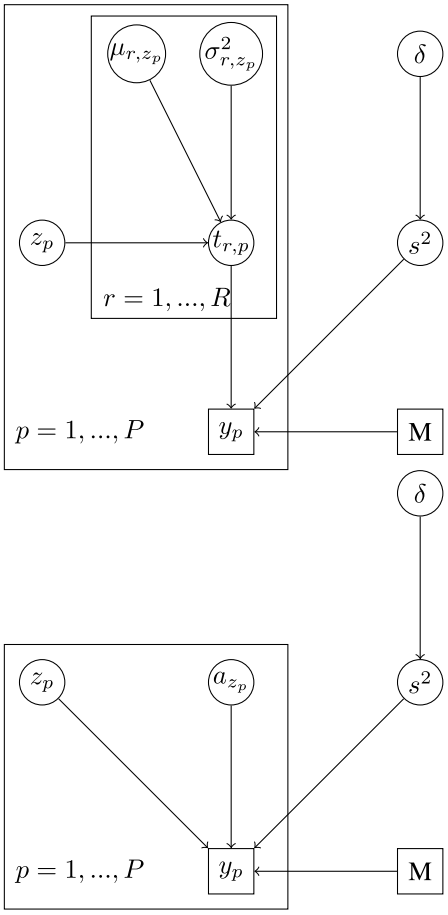

Fig. 1. DAGs of Eches' CLRSAM model (top) and the proposed model (bottom). Circle nodes and square nodes represent unknown and known random variables respectively, and rectangular boxes represent repetitive structures.

or via a cross-validation scheme. Automatic tuning is left as further work.

An inverse gamma prior Inv-Gamma$(1, \delta)$ is assigned to the error variance s^2, and the Jeffreys prior $1/\delta$ is chosen for the hyperparameter δ. An alternative would be to place a Jeffrey's prior on s^2 but a prior on δ is done here largely for convenience as it avoids fixing δ to a specific value. The label variable z_p is defined as in the Potts model, but instead of fixing the value of the granularity coefficient β as in Eches' CLRSAM, we will choose a cooling schedule for inverse β and apply SA at the model inference stage. As discussed further in Section IV-B, this provides further robustness to the sampler and avoids the local trap problem. Consequently, a new hierarchical Bayesian model is built as follows, with its DAG shown in Fig. 1(bottom).

1st level—likelihood

$$\boldsymbol{y}_p \mid z_p = k, \widetilde{\boldsymbol{a}}_k, s^2 \sim \mathcal{N}(\boldsymbol{M}\widetilde{\boldsymbol{a}}_k, s^2 \boldsymbol{I}_L)$$

2nd level—prior

$$\begin{aligned} \widetilde{\boldsymbol{a}}_k &\sim \text{Dir}(\alpha, ..., \alpha) \\ \mathbb{P}(z_p \mid z_{-p}) &\propto \exp\left[\beta \sum_{p' \in \boldsymbol{\mathcal{S}}_p} \delta(z_p - z_{p'})\right] \\ s^2 \mid \delta &\sim \text{Inv-Gamma}(1, \delta) \end{aligned} \tag{5}$$

Algorithm 1 Gibbs sampler for the proposed model

Initialise
  Choose starting values $\widetilde{\boldsymbol{a}}_k^{(0)}, z_p^{(0)}, s^{2(0)}, \delta^{(0)}, T^{(0)}$
  Set $\beta = 1/T^{(0)}$ **and** $i = 0$
repeat
  Sample $\widetilde{\boldsymbol{a}}_k^{(i+1)}$ **from** $f(\widetilde{\boldsymbol{a}}_k \mid z_p^{(i)}, s^{2(i)}, \delta^{(i)}, \boldsymbol{Y}, \boldsymbol{M})$
  Sample $z_p{}^{(i+1)}$ **from** $\mathbb{P}(z_p \mid \widetilde{\boldsymbol{a}}_k^{(i+1)}, s^{2(i)}, \delta^{(i)}, \beta, \boldsymbol{Y}, \boldsymbol{M})$
  Sample $s^{2(i+1)}$ **from** $f(s^2 \mid \widetilde{\boldsymbol{a}}_k^{(i+1)}, z_p^{(i+1)}, \delta^{(i)}, \boldsymbol{Y}, \boldsymbol{M})$
  Sample $\delta^{(i+1)}$ **from** $f(\delta \mid \widetilde{\boldsymbol{a}}_k^{(i+1)}, z_p^{(i+1)}, s^{2(i+1)}, \boldsymbol{Y}, \boldsymbol{M})$
  Decrease T **according to** $T^{(i+1)} = T^{(i)} r^{i+1} + T_e$
  Set $\beta = 1/T^{(i+1)}$ **and** $i = i + 1$
until $T \to 0$ **& convergence**

3rd level—hyperprior

$$f(\delta) \propto 1/\delta.$$

Putting this all together, it follows that the joint posterior is therefore decomposed as

$$\begin{aligned} f(\widetilde{\boldsymbol{A}}, s^2, \delta, Z | \boldsymbol{y}_{1:P}) \propto\ & f(\widetilde{\boldsymbol{A}}) \mathbb{P}(Z) f(s^2|\delta) \\ & \times \prod_{p=1}^{K} f(\boldsymbol{y}_p | z_p, \widetilde{\boldsymbol{a}}_{z_p}, s^2) \end{aligned} \tag{6}$$

where $\widetilde{\boldsymbol{A}} := (\widetilde{\boldsymbol{a}}_1, ..., \widetilde{\boldsymbol{a}}_K)$ denotes the abundances matrix with density $f(\widetilde{\boldsymbol{A}}) = \prod_{k=1}^{K} f(\widetilde{\boldsymbol{a}}_k)$.

IV. Model Inference

Once the proposed model has been constructed one is then left with the task of formulating the posterior distribution of the unknown parameters so that inference can be performed. However, the posterior distribution is too complex to obtain closed-form expressions of Bayesian estimators, such as the posterior mean and mode, of each unknown parameter. We therefore invoke MCMC methods to generate samples from the posterior distribution (6), and use the samples to approximate Bayesian estimators.

A. Hybrid Gibbs Sampler

As summarised in Algorithm (1), a Gibbs sampler is used for the proposed model. Let $\boldsymbol{\Upsilon} = (\widetilde{\boldsymbol{a}}_k, z_p, s^2, \delta, \boldsymbol{Y}, \boldsymbol{M})$ denotes the set of all random variables in the proposed model, and $\boldsymbol{\Upsilon}_{-\boldsymbol{v}}$ the set of random variables except $\boldsymbol{v}$. By applying the Bayes' theorem, we obtain the following results.

1) Full-Conditional Distribution of $\widetilde{\boldsymbol{a}}_k$:

$$\begin{aligned} f(\widetilde{\boldsymbol{a}}_k \mid \boldsymbol{\Upsilon}_{-\widetilde{\boldsymbol{a}}_k}) &\propto \prod_{p \in \boldsymbol{P}_k} f(\boldsymbol{y}_p \mid \widetilde{\boldsymbol{a}}_k, z_p, s^2) f(\widetilde{\boldsymbol{a}}_k) \\ &\propto \prod_{p \in \boldsymbol{P}_k} \exp\left[-\frac{1}{2s^2} \|\boldsymbol{y}_p - \boldsymbol{M}\widetilde{\boldsymbol{a}}_k\|^2\right] \prod_{r=1}^{R} \widetilde{\boldsymbol{a}}_{r,k}^{\alpha-1}. \end{aligned} \tag{7}$$

where $\boldsymbol{P}_k = \{p \in \{1, ..., P\} \mid z_p = k\}$. Unfortunately this is an unrecognised form of distribution that cannot be directly sampled. We therefore use a Metropolis-Hastings algorithm

to draw samples from it. We have chosen a proposal that simplifies the acceptance ratio and takes the form

$$q(\widetilde{\boldsymbol{a}}_k^{(i+1)}) \propto \prod_{p\in \boldsymbol{P}_k} \exp\left[-\frac{1}{2s^2}\|\boldsymbol{y}_p - \boldsymbol{M}\widetilde{\boldsymbol{a}}_k^{(i)}\|^2\right] \propto \exp\left[-\frac{1}{2}(\widetilde{\boldsymbol{a}}_k^{(i)} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1}(\widetilde{\boldsymbol{a}}_k^{(i)} - \boldsymbol{\mu})\right] \quad (8)$$

where by denoting $\boldsymbol{M}' = [\boldsymbol{m}_1, ..., \boldsymbol{m}_{R-1}]$ and $\boldsymbol{u} = [1, ..., 1]^T \in \mathbb{R}^{R-1}$

$$\boldsymbol{\mu} = \boldsymbol{\Lambda}\left[\frac{n_k}{s^2}(\boldsymbol{M}' - \boldsymbol{m}_R\boldsymbol{u}^T)^T\left(\frac{1}{n_k}\sum_{p\in P_k}\boldsymbol{y}_p - \boldsymbol{m}_R\boldsymbol{u}^T\right)\right]$$

$$\boldsymbol{\Lambda} = \left[\frac{n_k}{s^2}(\boldsymbol{M}' - \boldsymbol{m}_R\boldsymbol{u}^T)^T(\boldsymbol{M}' - \boldsymbol{m}_R\boldsymbol{u}^T)\right]^{-1}$$

and n_k is the number of sites in $\boldsymbol{P}_k$. Thus, the proposal distribution defined in (8) shares the form of a multivariate Gaussian distribution truncated on the $(R-1)$-simplex, namely $\mathcal{N}_{\mathbb{S}^{R-1}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}) := \mathbf{1}_{\mathbb{S}^{R-1}}(\widetilde{\boldsymbol{a}}_k^{(i+1)})\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, where $\mathbf{1}_{\mathbb{S}^{R-1}}$ is the indicator function on the $(R-1)$-simplex, and we have

$$\widetilde{\boldsymbol{a}}_k^{(i+1)} \mid \boldsymbol{\Upsilon}_{-\widetilde{\boldsymbol{a}}_k} \sim \mathcal{N}_{\mathbb{S}^{R-1}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}).$$

Sampling from such a distribution can be achieved by using a standard rejection sampling. However a more efficient Gibbs sampling method is available in [12]. Adapting this into our treatment is straightforward and results in the M-H algorithm acceptance probability

$$\min\left\{1, \frac{p(\widetilde{\boldsymbol{a}}_k)^{(i+1)}}{p(\widetilde{\boldsymbol{a}}_k)^{(i)}}\right\} = \min\left\{1, \prod_{r=1}^{R}\left(\frac{\widetilde{a}_{r,k}^{(i+1)}}{\widetilde{a}_{r,k}^{(i)}}\right)^{\alpha}\right\}.$$

Although the proposal in (8) is specific to this work, it is derived from the quite classical strategy of choosing a convenient proposal close to the distribution from which one would like to sample.

2) Full-Conditional Distribution of z_p:

$$\mathbb{P}(z_p = k \mid \boldsymbol{\Upsilon}_{-z_p}) \propto \mathbb{P}(z_p \mid z_{-p}) f(\boldsymbol{y}_p \mid \boldsymbol{a}_{z_p}, z_p, s^2) \propto \exp\left[\beta \sum_{p'\in\boldsymbol{\mathcal{S}}_p} \delta(z_p - z_{p'})\right] \times \exp\left[-\frac{1}{2s^2}\|\boldsymbol{y}_p - \boldsymbol{M}\widetilde{\boldsymbol{a}}_k\|^2\right]. \quad (9)$$

Note that, since (9) defines an MRF $Z \mid \boldsymbol{\Upsilon}_{-Z}$, sampling can be achieved by using SA in order to avoid the local-trap problem. This will be discussed in more detail in Section IV-B but we briefly note here that this involves setting $\beta = 1/T$, and sampling z_p from (9) while decreasing T according to $T^{(i+1)} = T^{(i)}r^{i+1} + T_e$. Also note that updating the label locations randomly or according to the checkerboard pattern (i.e., updating the even and then odd sites) generally improves the convergence properties of the Gibbs sampler. However, in our experiments, the convergence of the proposed sampler was sufficiently fast without using these more elaborated moves.

3) Full-Conditional Distribution of s^2:

$$f(s^2 \mid \boldsymbol{\Upsilon}_{-s^2}) \propto f(s^2 \mid \delta)\prod_{p=1}^{P} f(\boldsymbol{y}_p \mid \boldsymbol{a}_{z_p}, z_p, s^2).$$

As s^2 has a conjugate inverse-gamma prior, it follows

$$s^2 \mid \boldsymbol{\Upsilon}_{-s^2} \sim \text{Inv-Gamma}\left(1+\frac{LP}{2},\ \delta+\sum_{p=1}^{P}\frac{\|\boldsymbol{y}_p - \boldsymbol{M}\boldsymbol{a}_{(z_p)}\|^2}{2}\right).$$

4) Full-Conditional Distribution of δ:

$$f(\delta \mid \boldsymbol{\Upsilon}_{-\delta}) \propto f(s^2 \mid \delta) f(\delta)$$

and it can be shown that $\delta \mid \boldsymbol{\Upsilon}_{-\delta} \sim \text{Inv-Gamma}(1,\ s^2)$.

B. Local-Trap Problem

The so-termed local trap problem occurs when an MCMC sampler gets trapped within a local mode of a multi-modal distribution separated by low-density barriers. It is one of several well-known drawbacks of MCMC samplers [29]. The local trap problem may be observed when a new sample is always generated in a neighbourhood of the current sample. In the case of Gibbs sampling, for example, the update of one component of the random variables depends on the previous values of the other components, which results in local moves in the joint sample space. When there is high correlation between components the moves are expected to be more "local" and slow, and thus may fail to escape a local mode separated by low-density barriers [17].

For the Potts model defined in (2), there is a strong dependence between the neighbouring z_ps. This is especially apparent when the granularity coefficient β takes a large value. As a result when a Gibbs sampler is used to sample from a Potts model it is quite likely to suffer from slow and small moves. In the Gibbs sampler introduced for the proposed model, the same issue arises when we sample the label variable $\boldsymbol{Z}$ from the MRF defined by (9). If β is fixed too large initially, the Gibbs sampler used for sampling $\boldsymbol{Z}$ may get trapped within some local mode of the MRF and fail to find the global one.

To alleviate or overcome the local-trap problem, many advanced MCMC methods [29] have been proposed in the literature, including parallel tempering, slice sampler, adaptive Metropolis algorithm, stochastic approximation Monte Carlo [29], as well as the collapsed Gibbs sampler of van Dyk and Park [45]. In particular, [40] considers estimation of β in the Potts model within an MCMC algorithm. In our case, however, we propose to consider the method of simulated annealing [28], which provides a perfectly reasonable solution.

Simulated annealing (SA) is a stochastic algorithm for finding the global minimum of a given function that may possess several local minima. It simulates the physical annealing procedure, i.e., a process whereby a physical substance is heated and then allowed to cool down slowly to form a "low-energy configuration" [28]. For example, consider a system whose energy configuration has the following Gibbs distribution

$$f_T(x) \propto \exp[-E(x)/T]$$

where $T > 0$ is the temperature parameter, and $E(x)$ denotes the energy of the system. The SA method simply applies a sampling algorithm, be it Metropolis, Gibbs, or otherwise, such that the value of the temperature T is decreased gradually at each iteration step. Initially, T is set very high and x is set to a random configuration. At a fixed T, the samples are drawn according to the distribution $f_T(x)$. After the sampling converges to the equilibrium at the current temperature, T is then decreased in accordance with a given cooling strategy. The process is repeated until T is near to zero (at which, to carry on the annealing analogy, the system is "frozen" near its low-energy configuration).

Geman and Geman [19] established an SA convergence result which states that if the decreasing sequence of temperatures satisfy

$$\lim_{i\to\infty} T^{(i)} = 0$$

and

$$T^{(i)} \geq \frac{\triangle}{\ln(i)} \tag{10}$$

where $\triangle$ is a problem-dependent constant, then the algorithm is guaranteed to converge to the global minimum irrespective of the initial configuration. However, the cooling schedule implied in (10) is too slow to be of practical use. Instead, heuristic, faster schedules have to be used in practice.

In the proposed model, β in (5) can be considered as the inverse temperature $T = 1/\beta$. Smaller β will present a weaker spatial prior and will therefore tend to result in "noisier" maps, whilst larger values give stronger priors and lead to "smoother" maps. By allowing β to start off relatively small, the sampler is able to explore a large class of different labelling configurations early on. This behaviour may help the sampler avoid the local trap problem. As the "temperature" decreases (as β increases), the prior becomes stronger and helps stabilize the resulting sampled chain of configurations towards spatially homogeneous solutions. Other variables could have been chosen for similar SA treatment. The choice of β was inspired by the observation that, in our simulations, the CLRSAM sometimes resulted in either "noisy" maps for small fixed β or, sometimes (for example, see Section V-B) would suffer from the local-trap problem for larger β. By varying β in this manner one is able to derive the benefits of both exploration and stability.

Since the SA is embedded within the Gibbs sampler, this aspect of our proposed scheme has some similarities with the simulated tempering idea discussed in [36] and first proposed in [31]. However, the proposed SA-within-Gibbs sampler and simulated tempering have also differences that deserve to be mentioned here. Simulated tempering consists of sampling a sequence of distributions that vary randomly over time according to an appropriate temperature cooling schedule. In particular, simulated tempering considers an enlarged space containing the parameter vector to be sampled and the temperature which is, itself, considered as a dynamical parameter. Conversely, the algorithm proposed in this paper updates the temperature at each iteration (or equivalently, the value of β) in order to stabilize the generated values of β. The main motivation for using this strategy is its simplicity. Note that other optimization-within-MCMC algorithms that are described in [41] could also be investigated to sample the value of β.

For practicality, we choose an exponential cooling schedule [21]

$$T^{(i)} = T^{(0)} r^i + T_e \tag{11}$$

where i denotes the ith iteration, $T^{(0)}$ is an initial value of the temperature which should be set large enough to encourage sufficient exploration of the state space early on in the algorithm, T_e is the limit value of the temperature that should be close enough to zero to ensure that the algorithm reaches a reasonable low-energy equilibrium configuration, and r $(0 < r < 1)$ is the decreasing rate which should be given a value close enough to 1 to ensure that the temperature decreases slowly enough but far away enough from 1 to ensure that the algorithm is tractable. The choices of r and $T^{(0)}$ are not critical parameters in that, if set large enough, they do not have a big impact on the results. In other words, they do not have to be carefully tuned to each new data set. The trade-off is merely quality versus computation time. As both parameters are raised the computation time increases. However, there will be a limit to the accuracy of the solution (it does not increase indefinitely with respect to these parameters).

C. Convergence Diagnostics

Owing to the complex nature of the proposed model and sampler, convergence diagnostics is far from straightforward. For one thing, there is a large number of random variables involved. In practice, however, it is fortunately unnecessary to check the convergence of each variable. We will therefore concentrate on the main (multivariate) variables of interest, namely the abundance variable $\boldsymbol{A}$ and the label variable $\boldsymbol{Z}$. Visual inspection provides a natural means to formulate a choice of convergence diagnostic because the label variable Z can be conveniently shown as label map, and the abundance variable $\boldsymbol{A}$ can also be presented as abundance maps of each endmember in $\boldsymbol{M}$. Convergence is assumed to hold when one observes that the label map and the abundance maps become relatively stable without any considerable changes as the number of iterations increases. However, although seemingly common (especially in hyperspectral unmixing tasks), visual inspection can be inaccurate and somewhat arbitrary. Hence, we will mostly rely on the Gelman-Rubin diagnostic [18] in this work.

1) Gelman-Rubin Diagnostic: The main idea of the Gelman-Rubin diagnostic is to run multiple chains with widely differing starting values and then to compare the variance within the chains with the variance across the chains. After convergence, the two variances should be similar. We run $m \geq 2$ chains of length $2n$ with over-dispersed starting values, and keep the last n samples in each chain. We first calculate the within-chain variance $W = m^{-1} \sum_{j=1}^{m} s_j^2$, where s_j^2 is the variance of the jth chain, i.e.,

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\theta_{ij} - \bar{\theta_j})^2$$

and where θ_{ij} denotes the value of ith sample in jth chain, and $\bar{\theta}_j$ is the mean value of jth chain. The sample estimate W likely underestimates the true variance of the stationary distribution since the chains may have not converged to the stationary distribution. We then calculate the between-chain variance

$$B = \frac{n}{m-1}\sum_{j=1}^{m}(\bar{\theta}_j - \bar{\bar{\theta}})^2\,, \quad \bar{\bar{\theta}} = \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}_j.$$

The between-chain variance B is the variance of the chain means multiplied by n because each chain is based on n draws. We can then estimate the variance of the stationary distribution as a weighted average of W and B as follows

$$V = \left(1 - \frac{1}{n}\right)W + \frac{1}{n}B\,.$$

Because of over-dispersion of the starting values, V overestimates the true variance. Finally we calculate the potential scale reduction factor $R = \sqrt{V/W}$. When R is close enough to 1 (e.g., smaller than 1.05, say), we then assume the convergence has been reached. If there is more than one parameter, we can calculate the potential scale reduction factor for each one. An alternative is to use the multivariate potential scale reduction factor [7].

After convergence has been determined, another question is how long the chain should be run. In this work, we simply follow the typical routine and thin the chain by only taking every kth value (e.g., every 10th value), in order to decrease autocorrelations in the chain.

V. Experiments on Synthetic Images

Unmixing experiments were performed on synthetic data constructed in accordance with rule B. Although this naturally favours the proposed model it is, nonetheless, instructive to contrast some of the key behavioural differences with that of the CLRSAM (Eches et al. [15]). A direct comparison can be made here since both the CLRSAM and our proposed approach incorporates the same MRF model. Indeed, in Section V-B, we examine the benefits of the SA method for CLRSAM. In Section V-C we then compare the computational performance and accuracy of the abundance and label estimation obtained by the proposed method and the CLRSAM (both using SA). Later on, in Section VI, we consider the relative benefits of sparsity of the proposed approach on a popular real hyperspectral data set.

A. Synthetic Data Generation

The generated synthetic image is endowed with $L = 224$ spectral bands, $P = 25 \times 25$ pixels, and $K = 3$ different classes. The label map is illustrated in Fig. 3 (left). The procedure of synthetic image generation is as follows:

1) Choose $R = 3$ significantly different endmembers from the USGS spectral library splib06,[1] and hence the endmember matrix $\boldsymbol{M}$;

[1]The splib06 library, released by USGS in September 2007, comprises spectral signatures of 240 different mineral types, with reflectance values given in 224 spectral bands covering wavelength range of 400-2500 nm. Available online: http://speclab.cr.usgs.gov/spectral.lib06

TABLE I

The Ground Truth Abundance Vectors for Each Class in the Synthetic Image

a_1^T	a_2^T	a_3^T
[0.6 0.3 0.1]	[0.3 0.5 0.2]	[0.3 0.2 0.5]

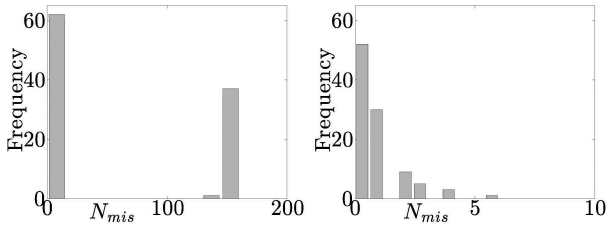

Fig. 2. Distributions of the number of mislabelled pixels N_{mis} before using SA (left) and after using SA (right).

2) Generate a label map Z with 3 classes by simulating the Potts model with $\beta = 1.1$;
3) Choose the common abundance vector for each class as in Table I, and hence the actual abundance matrix $\boldsymbol{A}$ by setting $\widetilde{\boldsymbol{a}}_p = \widetilde{\boldsymbol{a}}_k$ if $z_p = k$;
4) Generate the data Y according to (1) (LMM) with variance $s^2 = 0.001$.

B. SA Performance

In our first experiment, we illustrate the benefits of SA by applying it directly to Eches' CLRSAM. We note that Eches' approach does not apply SA when sampling the label variable $\boldsymbol{Z}$. We therefore embed SA into the CLRSAM to produce an extended version and compare its performance with that of the original model. In the original CLRSAM, the granularity coefficient β is fixed to 1.1. When applying SA, β is the inverse temperature T. So β will increase from 0 to 1.1 if we decrease the temperature T from infinity to 0.91. The exponential cooling schedule in (11) then becomes

$$T^{(i)} = T^{(0)}r^i + 0.91$$

where we will set $T^{(0)} = 100$ and $r = 0.95$, and run a large enough number of iterations to allow T to decrease close enough to 0.91.

We measure the performances of Eches' CLRSAM before and after using SA with respect to the number of mislabelled pixels, denoted by N_{mis}, which is computed by comparing the estimated label map with the actual label map. Here the label map is estimated by the converged mode of the label variable. We run 100 experiments and record the numbers of misclassified pixels over all experiments.

Given the synthetic image we implement the two models (over 100 experiments each) and obtain the distributions of N_{mis} shown in Fig. 2. It is clear that the original CLRSAM results in a distribution of N_{mis} with two significant modes, one is around 0 with probability of about 0.6, and the other one is around 150 with probability of about 0.4. It indicates that the model suffers from getting trapped in a local mode of the posterior of $\boldsymbol{Z}$ whose corresponding label map is shown

Fig. 3. Left: the correct label map. Right: a false label map resulted from Eches' CLRSAM.

TABLE II
RESULTS OF THE PERFORMANCE MEASURES OF ECHES' CLRSAM MODEL AND THE PROPOSED MODEL

Model	N_{mis}	MSE($\widehat{\boldsymbol{A}}$)	t_{cost} (s.)
Eches' model	1.2	8.14×10^{-4}	74.6
Proposed model	0	1.39×10^{-5}	5.5

in Fig. 3 (right). On the other hand, after using SA the distribution of N_{mis} is concentrated around 0 with maximum N_{mis} of only 6. In other words, the modified model avoids the local mode of the posterior of $\boldsymbol{Z}$. We can conclude that the introduction of the SA technique helps Eches' CLRSAM successfully avoid the local-trap problem for this example and thus delivers a more accurate estimate of the label map. In subsequent experiments herein this modified, enhanced, version of Eches' CLRSAM will be used instead of the original one.

C. MSE and Computational Performance

In this experiment, we compare Eches' CLRSAM model and the proposed model. For simplicity, the concentration parameter α is set to unity. The next section explores the sparsity effects of choosing $\alpha < 1$. Three performance measures are used, namely, the number of mislabelled pixels (N_{mis}), the computational time (t_{cost}), and the mean squared error of the estimated abundance defined by

$$\text{MSE}(\widehat{\boldsymbol{A}}) = \frac{1}{RP} \sum_{r=1}^{P} \sum_{p=1}^{P} (\widehat{a}_{r,p} - a_{r,p})^2.$$

Table II summarises the mean performance measures over 10 experiments for the two models on the synthetic image. We found that the proposed model outperforms Eches' CLRSAM, with no mislabelled pixels, and an MSE($\widehat{\boldsymbol{A}}$) over 10 times smaller than that of Eches' CLRSAM with an average computational cost of only 5.5 seconds.

Of course, it should be noted that the results corroborate to some extent the fact that the assumption made in the synthetic image generation, namely Rule B, is well-aligned to the properties of the proposed model. In the next section, however, we find that the sparsifying property of the proposed approach provides potential further benefits compared to that of Eches' CLRSAM method when considering unmixing of real hyperspectral data.

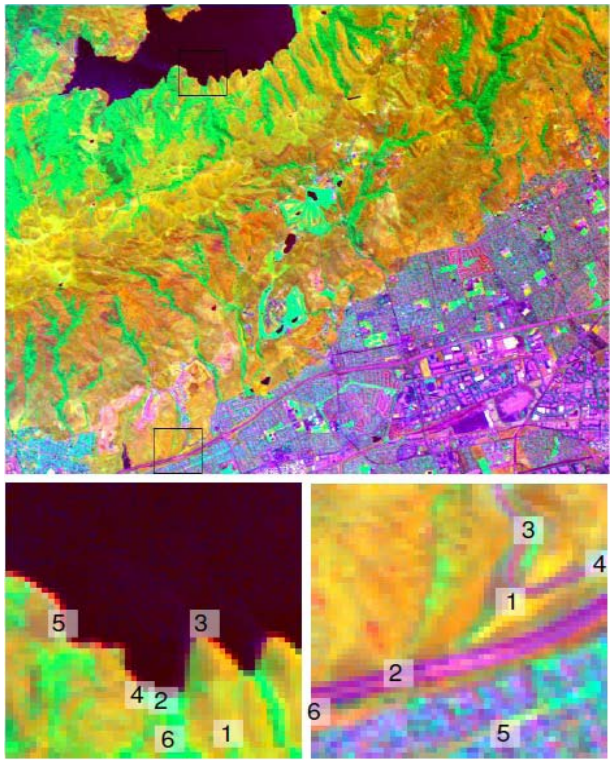

Fig. 4. False color image of the Moffett Field dataset (top) and the two regions of interest (bottom-left: ROI 1, bottom-right: ROI 2). The numbers 1-6 indicate the endmembers extracted from each region by the VCA algorithm.

VI. EXPERIMENTS WITH REAL IMAGES

The popular Moffett Field hyperspectral image was used to qualitatively evaluate the performance of the proposed model on real data. This dataset, widely used to test spectral unmixing, and other hyperspectral imaging techniques [10], [13], [24], was collected by the AVIRIS sensor over Moffett Field, California around the southern part of the San Francisco Bay in 1997. The top panel of Fig. 4 shows a false color image of the Moffett Field scene with two regions of interest (ROI). This scene consists of a large lake in the top-left of the image, a coastal area composed of vegetation and soil and an urban area in the bottom-right. The dataset contains 224 spectral bands covering a wavelength range from 400 to 2500 nanometers, at a nominal spectral resolution of 10 nm, and it has been reduced from the original 224 bands to $L = 188$ bands by removing noisy and water absorption bands.

A. ROI 1

ROI 1 has $P = 50 \times 50$ pixels, and covers $K = 4$ general classes of areas: lake, lake shore, vegetation and bare land. Thus there are at least three endmembers present in the region associated with water, vegetation and soil.

1) Stability With Respect to Number of Endmembers: Despite the existence of methods that estimate the number of endmembers present such as ELM [30] and HySIME [6], they can clearly never be infallible and are often applied as a preprocessing step independently from the unmixing methodology itself. It is prudent and instructive, therefore, to consider the stability of the two unmixing approaches as

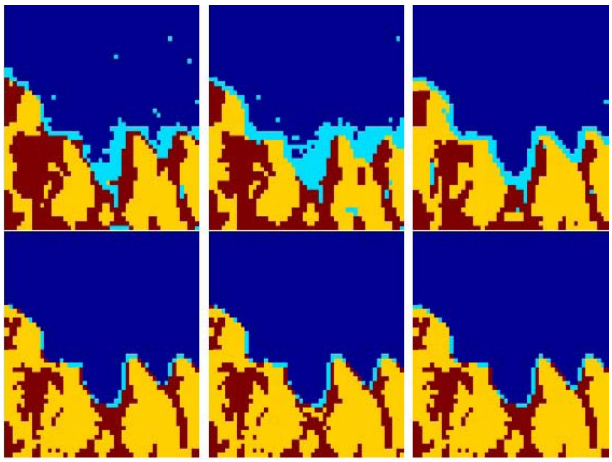

Fig. 5. The label maps of ROI 1 estimated by Eches' CLRSAM model (top) and the proposed model (bottom) for $R = 3, 4$, and 5 respectively. Compare with the false colour image in the lower-left panel of Fig. 4.

TABLE III

COMPUTATIONAL TIMES (SECONDS) OF ECHES' CLRSAM MODEL AND THE PROPOSED MODEL ON ROI 1 FOR $R = 3$, 4 AND 5 RESPECTIVELY

Model	$R = 3$	$R = 4$	$R = 5$
Eches' model	266.2	333.2	357.8
Proposed model	30.5	30.9	31.8

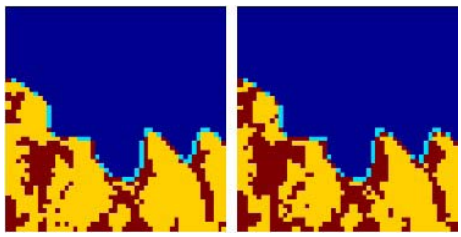

Fig. 6. The label maps estimated by the proposed model with $\alpha = 1$ and $\alpha = 0.01$.

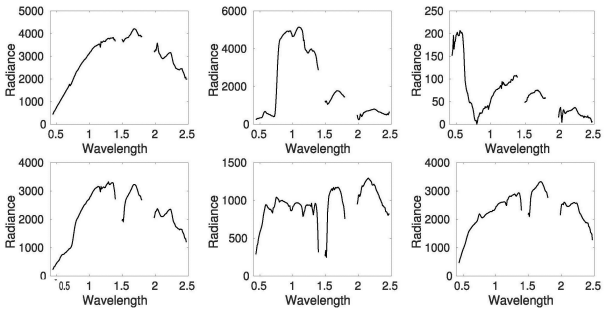

Fig. 7. The six endmembers extracted from ROI 1. Top row, left-to-right: soil, vegetation, water taken from location 1-3 resp. in the bottom-left panel of Fig. 4; bottom row: possible mixes between vegetation/soil, soil/water, soil/vegetation taken from locations 4-6 resp. in the bottom-left panel of Fig. 4.

the number of endmembers is varied. To this end, we proceed as follows. We set $K = 4, \alpha = 1$, and, as is the case in all experiments on the real data, used $T_e = 0.91$ (which is equivalent to $\beta = 1.1$ in Eches' CLRSAM model). To study the effects of the numbers of endmembers, we let R take on the values 3, 4, and 5.

As can be seen from Fig. 5, the proposed model gives more accurate and stable label maps than those of Eches' CLRSAM as the number of endmembers increases. The colours dark-blue, light-blue, orange, and red represent estimates of the water, littoral, soil, and vegetation classes respectively. In contrast to the proposed method, CLRSAM contains some clearly suspect estimates such as single shore pixels that either appear on land or are surrounded by the water class. Moreover, as the number of assumed endmembers R is varied, the CLRSAM label maps change significantly compared with the proposed method. This result confirms that CLRSAM is more sensitive to knowledge about the number of endmembers present. Table III indicates that there is also a significant speed advantage too. The proposed method is not only quicker by an order of magnitude it also appears to scale better as R is increased.

2) Sparsity for Endmember Shrinkage and Selection: In the next experiment we fix $K = 4$ and set the numbers of endmembers to be some way larger than expected, say $R = 6$ to test the ability of the proposed approach to mitigate endmember redundancy by the use of sparsity (by employing $\alpha < 1$). We thus compare the behaviour of the system when $\alpha = 1$ to that of $\alpha = 0.01$.

Six endmembers are extracted from ROI 1 via VCA, as illustrated in Fig. 7. These correspond to the pixels marked in the bottom-left panel of Fig. 4 (bottom-left). VCA finds distinct pixels from the image data itself and assigns them as endmember signatures under the assumption that the pure pixels are present in an image scene. Indeed, the first three signatures represent soil, vegetation, and water and they are similar to the actual signatures of these materials in the database. The other three signatures, as indicated by their location, might be mixed signatures of soil, vegetation and water. For example, the fourth endmember m_4, located in the border between bare soil area and vegetation area, is probably a spectral mixture of soil and vegetation.

When α is decreased from 1 to 0.01, as can be seen from Table IV, the proportions of endmembers m_2 and m_6 become almost zero. In effect, the redundancy of the endmember set (deliberately introduced by choosing $R = 6$ endmembers), is mitigated somewhat by choosing a small value of α. For small α, the Dirichlet prior induces concentration of mass, or sparsity, in the manner in which the abundance values are distributed over the simplex. As a further consequence, the distributions of the endmembers themselves become more concentrated. In the "shore" area, for instance, the number of constituent endmembers decreases from 6 to 3. For this example, in addition to Table IV, we also see from Fig. 6 that, although the choice of α has a significant impact on the estimation of the abundances, it has little effect on the inferred label map.

Notwithstanding, the sparse method yields an abundance solution which is much closer to something that one could reasonably expect. This fact alone should motivate the preference for the sparse solution in this case. Although the inferred label map in both ROI experiments were somewhat similar, it is also important to note that the inferred endmembers offer more refined information than class. The solutions obtained

TABLE IV

THE ESTIMATED ABUNDANCE VECTORS IN EACH CLASS OF ROI 1 FOR $\alpha = 1$ AND 0.01 RESPECTIVELY

	$\alpha = 1$				$\alpha = 0.01$			
Area	shore	veg.	land	lake	land	lake	veg.	shore
m1	0.108	0.100	0.657	0.004	0.768	0.000	0.201	0.000
m2	0.117	0.276	0.000	0.000	0	0.000	0.014	0
m3	0.363	0.165	0.109	0.942	0.000	0.924	0.168	0.043
m4	0.116	0.377	0.197	0.000	0.044	0	0.617	0.168
m5	0.295	0.000	0.000	0.053	0.178	0.076	0	0.789
m6	0.001	0.082	0.038	0.000	0.009	0.000	0.000	0.000

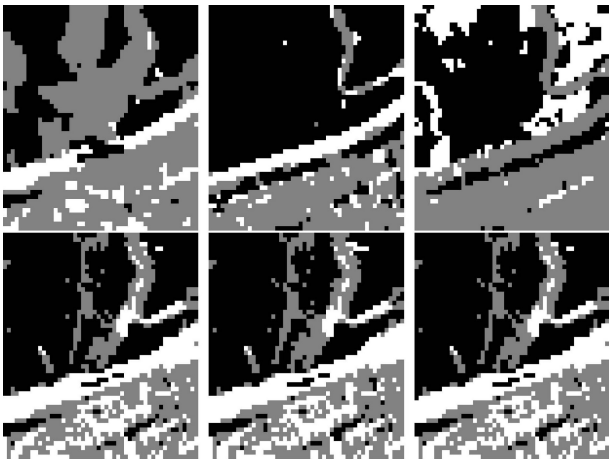

Fig. 8. The label maps of ROI 2 estimated by Eches' CLRSAM model (top) and the proposed model (bottom) for $R = 3, 4$ and 5 respectively. Compare with the false colour image in the lower-right panel of Fig. 4.

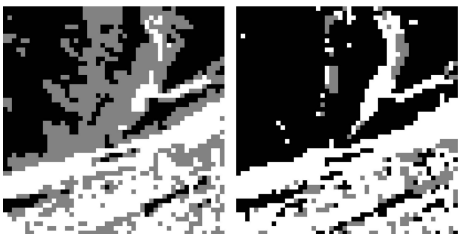

Fig. 9. The label maps of ROI 2 estimated by the proposed model with $\alpha = 1$ and $\alpha = 0.01$ (for $K = 3$, $R = 6$).

from realistic and reasonable endmember proportions lead to better interpretability.

B. ROI 2

The other region of interest studied, ROI 2, consists of three main classes: urban/roads, vegetation, and bare land. It is therefore likely that there are at least three materials present in this region associated with vegetation, soil, and concrete.

1) Stability With Respect to Number of Endmembers: Similar to the ROI 1 experiment, we consider the robustness of the inferred label maps to variations in the value of R. We set $K = 3, \alpha = 1$ and let $R = 3, 4$, and 5. It is shown in Fig. 8 that the proposed model identifies the two vegetation strips in the top-middle of the region, the road running from the bottom-left to the middle-right, and the road loop in the top-right. The false colour map in the bottom-right of Fig. 4 reveals that the lower-right part area comprises a fine-grained "speckled" mixture of

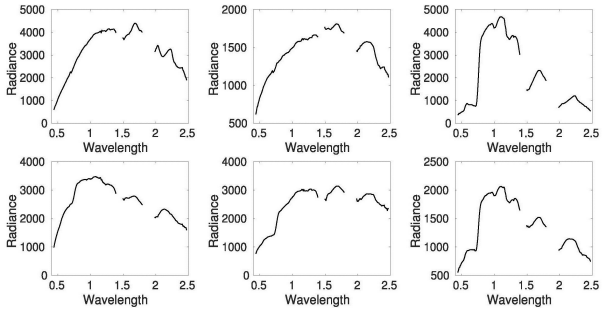

Fig. 10. The six endmembers extracted from ROI 2. Top row, left-to-right: soil, soil/construction, vegetation taken from location 1-3 resp. in the bottom-right panel of Fig. 4; bottom row: possible mixes between soil/ construction, construction/soil/vegetation, vegetation/construction taken from locations 4-6 resp. in the bottom-right panel of Fig. 4.

TABLE V

THE ESTIMATED ABUNDANCE VECTORS IN EACH CLASS OF ROI 2 FOR $\alpha = 1$ AND 0.01 RESPECTIVELY

	$\alpha = 1$			$\alpha = 0.01$		
Area	Construction	land	veg.	Construction	land	veg.
m1	0.041	0.628	0.378	0.000	0.295	0.000
m2	0.507	0.341	0.449	0	0	0
m3	0.171	0.031	0.173	0.000	0.000	0.407
m4	0.120	0.000	0.000	0.000	0.000	0.002
m5	0.069	0.000	0.000	0.151	0.260	0.210
m6	0.092	0.000	0.000	0.849	0.445	0.381

urban/roads/concrete (purple), vegetation (green), and soil/bare land (orange). These seem to be also present in the lower-right part of the proposed estimated maps (although it is difficult to determine exactly how accurate the estimation is). However, this behaviour does not seem to be reflected as much by CLRSAM where the majority of the lower-right is swathed in what appears to be the vegetation class. Furthermore, it is also apparent that the proposed model gives more accurate and stable label maps than those of Eches CLRSAM model as the number of endmembers is varied.

2) Sparsity for Endmember Shrinkage and Selection: Fig. 9 shows that, for the case where $K = 3$ and $R = 6$, the proposed method does finally yield a significantly different label map for two different values of α. We see that, in particular, the interface between the vegetation and soil is markedly different. It appears that large regions of soil are misclassified as vegetation. On the other hand, when α is reduced from 1 to 0.01, these areas are classified more accurately. Although some of the vegetation area is now missing the road loop in the top-right is more apparent than it is in any of the other experiments. It is also noteworthy that the choice of $\alpha = 0.01$ gives a result closer to that of, say, the $R = 4$ experiment illustrated in Fig. 8: there is a 77.5% agreement between the $R = 4$ experiment and $(R = 6; \alpha = 0.01)$ experiment whereas there is a 74.2% agreement between the $R = 4$ and $(R = 6; \alpha = 1)$ experiments. Furthermore, it can be seen from Table V that the proportions of endmembers m_2 and m_4 become almost zeros. Again, the Dirichlet prior has counterbalanced the redundancy in the endmember selection somewhat (cf. Fig. 10) and has also produced more concentrated endmember distributions.

VII. Conclusion

Our experiments with synthetic and real data indicated that the proposed model outperforms Eches' CLRSAM model with respect to the accuracy of unmixing and classification as well as computational time. In particular, when we set the concentration parameter to a value below 1, the proposed model showed some ability to perform sparse unmixing. Image classification was jointly solved with hyperspectral unmixing in this work. Therefore, the accuracy of image classification depended on the accuracy of unmixing, and vice versa. In this sense, the two tasks can be seen as mutually beneficial; this arguably helps produce better results. Our work distinguishes between two different classification rules in this paper, namely, Rule A of stochastic abundance vectors in each class and Rule B of common abundance vectors in each class. Since the proposed model is built based on Rule B, we might be tempted to conclude that the assumption of common abundance vector in each class is more compatible with reality but follow-on work is required to study this further. We have shown that the proposed method offers stability to the assumed number of endmembers. Furthermore, when the number of endmembers is set too large, the Dirichlet concentration parameter embedded in our model can help mitigate the endmember redundancy.

Markov random fields (MRFs) have been widely used to model spatial correlations between image pixels. However, as demonstrated in this work, MCMC inference over MRFs tends to suffer from the local-trap problem. This is because of, on the one hand, the local nature of the inferential MCMC update scheme and, on the other hand, the local spatial interactions inherently defined by the Markov property on the lattice. However, we have shown that simulated annealing, as well as other advanced MCMC methods, can be used to alleviate or overcome this problem.

Our future work will focus on improving the stability of sparse unmixing with the proposed model. To construct a fully-practicable sparse Bayesian method, one would need to establish a means by which the Dirichlet concentration parameter α can be set or tuned, either a priori or learned from training data. Inspiration can be drawn here from works such as [6] and [30] who have proposed ways to estimate the number of endmembers a priori. In addition, we can also exploit the spectral correlations together with spatial correlations in hyperspectral images. This work could then potentially benefit situations where there is a relatively large total number of endmembers present in the image but where the number of endmembers that participate in the formation of a given pixel is relatively small. This could ultimately pave the way to the construction of a sparse Bayesian approach to unmixing using a spectral library rather than endmembers extracted from the image itself.

Acknowledgment

The authors would like to thank Dr. Olivier Eches for sharing code and implementation advice. Peng Chen and Jean-Yves Tourneret would like to dedicate this work to their co-author James Nelson who sadly and unexpectedly passed away in September 2016. They will never forget his drive that motivated the writing this paper and his personal kindness.

References

[1] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised post-nonlinear unmixing of hyperspectral images using a Hamiltonian Monte Carlo algorithm," *IEEE Trans. Signal Process.*, vol. 23, no. 6, pp. 2663–2675, Jun. 2014.

[2] M. Arngren, M. N. Schmidt, and J. Larsen, "Unmixing of hyperspectral images using Bayesian non-negative matrix factorization with volume prior," *J. Signal Process. Syst. Signal Image Video Technol.*, vol. 65, no. 3, pp. 479–496, Dec. 2011.

[3] 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.

[4] J. M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. 1st Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Grenoble, France, Aug. 2009, pp. 1–4.

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

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

[7] S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," *J. Comput. Graph. Statist.*, vol. 7, no. 4, pp. 434–455, 1998.

[8] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[9] N. Chopin, "Fast simulation of truncated Gaussian distributions," *Statist. Comput.*, vol. 21, no. 2, pp. 275–288, Apr. 2011.

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

[11] 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.

[12] N. Dobigeon and J.-Y. Tourneret, "Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex," Univ. Toulouse, Toulouse, France, Tech. Rep., 2007. [Online]. Available: http://dobigeon.perso.enseeiht.fr/

[13] 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, Jul. 2008.

[14] N. Dobigeon, J.-Y. Tourneret, C. Richard, J.-C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.

[15] O. Eches, J. A. Benediktsson, N. Dobigeon, and J.-Y. Tourneret, "Adaptive Markov random fields for joint unmixing and segmentation of hyperspectral images," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 5–16, Jan. 2013.

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

[17] D. Gamerman and H. F. Lopes *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, London, U.K.: Chapman & Hall, 2006.

[18] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statist. Sci.*, vol. 7, no. 4, pp. 457–472, 1992.

[19] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[20] 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. 39, no. 3, pp. 529–545, Mar. 2001.

[21] L. Ingber, "Simulated annealing: Practice versus theory," *Math. Comput. Model.*, vol. 18, no. 11, pp. 29–57, Dec. 1993.

[22] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
[23] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
[24] X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, Jan. 1999.
[25] C.-T. Li, "Multiresolution image segmentation integrating Gibbs sampler and region merging algorithm," *Signal Process.*, vol. 83, no. 1, pp. 67–78, Jan. 2003.
[26] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.
[27] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
[28] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, London, U.K.: Springer, 2009.
[29] F. Liang, C. Liu, and R. Carroll, *Advanced Markov Chain Monte Carlo Methods: Learning From Past Samples*, Hoboken, NJ, USA: Wiley, 2011.
[30] B. Luo, J. Chanussot, S. Douté, and L. Zhang, "Empirical automatic estimation of the number of endmembers in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 24–28, Jan. 2013.
[31] E. Marinari and G. Parisi, "Simulated tempering: A new Monte Carlo scheme," *Statist. Comput.*, vol. 19, no. 6, pp. 451–458, 1992.
[32] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.
[33] R. Mittelman, N. Dobigeon, and A. O. Hero, "Hyperspectral image unmixing using a multiresolution sticky HDP," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1656–1671, Apr. 2012.
[34] S. Moussaoui, C. Carteret, D. Brie, and A. Mohammad-Djafari, "Bayesian analysis of spectral mixture data using Markov chain Monte Carlo methods," *Chemometrics Intell. Lab. Syst.*, vol. 81, no. 2, pp. 137–148, Apr. 2006.
[35] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[36] R. M. Neal, "Sampling from multimodal distributions using tempered transitions," *Statist. Comput.*, vol. 6, no. 4, pp. 353–366, Dec. 1992.
[37] R. Neher and A. Srivastava, "A Bayesian MRF framework for labeling terrain using hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1363–1374, Jun. 2005.
[38] R. Nishii and T. Ozaki, "Contextual unmixing of geospatial data based on Markov random fields and conditional random fields," in *Proc. 1st Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Grenoble, France, Aug. 2009, pp. 478–481.
[39] V. Pawlowsky-Glahn and A. Buccianti, *Compositional Data Analysis: Theory Applications*, Hoboken, NJ, USA: Wiley, 2011.
[40] 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. Signal Process.*, vol. 22, no. 6, pp. 2385–2397, Jun. 2013.
[41] M. Pereyra *et al.*, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 224–241, Mar. 2016.
[42] G. Rellier, X. Descombes, F. Falzon, and J. A. Zerubia, "Texture feature analysis using a gauss-Markov model in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1543–1551, Jul. 2004.
[43] C. P. Robert, "Simulation of truncated normal variables," *Statist. Comput.*, vol. 5, no. 2, pp. 121–125, Jun. 1995.
[44] C. Theys, N. Dobigeon, J.-Y. Tourneret, and H. Lanteri, "Linear unmixing of hyperspectral images using a scaled gradient method," in *Proc. IEEE 15th Workshop Statist. Signal Process. (SSP)*, Aug. 2009, pp. 729–732.
[45] D. A. van Dyk and T. Park, "Partially collapsed Gibbs samplers: Theory and methods," *J. Amer. Statist. Assoc.*, vol. 103, no. 482, pp. 790–796, 2008.
[46] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE.*, vol. 3753, pp. 266–275, Oct. 1999.
[47] Y. Zhu, Y. Tan, Y. Hua, G. Zhang, and J. Zhang, "Automatic segmentation of ground-glass opacities in lung CT images by using Markov random field-based algorithms," *J. Digit. Imag.*, vol. 25, no. 3, pp. 409–422, Jun. 2012.

Peng Chen was born in Shandong, China, in 1986. He received the B.Sc. degree in mathematics and applied mathematics from China University of Geosciences, Beijing, in 2009, the M.Sc. degree in systems theory from Northeastern University in 2011, and the M.Sc. degree in statistics from University College London in 2014. He is currently a Model Analyst with Finup Group, Beijing, China, where he is involved on credit risk modeling and statistical learning applications in financial field.

James Nelson received the Ph.D. degree in applied harmonic analysis from Mathematics Department, Anglia Polytechnic University, Cambridge, U.K., in 2001. He held post-doctoral positions with the University of Cranfield from 2001 to 2004, the University of Southampton from 2004 to 2006, and the University of Cambridge from 2006 to 2010. In 2010, he joined University College London as a Lecturer with the Department of Statistical Science. His research interests include wavelet analysis with a focus on detection, enhancement, and classification for signal and image processing and machine learning.

Jean-Yves Tourneret (SM'08) received the Ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT) of Toulouse in 1989 and the Ph.D. degree from the National Polytechnic Institute of Toulouse in 1992. He is currently a Professor with ENSEEIHT and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activities are centered on statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo methods.