A Bayesian model for joint unmixing and robust classification of hyperspectral image

Adrien Lagrange, Mathieu Fauvel, Stéphane May, Nicolas Dobigeon

A BAYESIAN MODEL FOR JOINT UNMIXING AND ROBUST CLASSIFICATION OF HYPERSPECTRAL IMAGES

Adrien Lagrange[⋆] *Mathieu Fauvel*[†] *Stéphane May*[◇] *Nicolas Dobigeon*[⋆]

[⋆] University of Toulouse, IRIT/INP-ENSEEIHT, Toulouse, France
[†] University of Toulouse, INP-ENSAT, UMR 1201 DYNAFOR, France,
[◇] Centre National d'Études Spatiales (CNES), Toulouse, France
firsname.name@{enseeiht,ensat,cnes,enseeiht}.fr

ABSTRACT

Supervised classification and spectral unmixing are two methods to extract information from hyperspectral images. However, despite their complementarity, they have been scarcely considered jointly. This paper presents a new hierarchical Bayesian model to perform simultaneously both analysis in order to ensure that they benefit from each other. A linear mixture model is proposed to described the pixel measurements. Then a clustering is performed to identify groups of statistically similar abundance vectors. A Markov random field (MRF) is used as prior for the corresponding cluster labels. It promotes a spatial regularization through a Potts-Markov potential and also includes a local potential induced by the classification. Finally, the classification exploits a set of possibly corrupted labeled data provided by the end-user. Model parameters are estimated thanks to a Markov chain Monte Carlo (MCMC) algorithm. The interest of the proposed model is illustrated on synthetic and real data.

Index Terms— Bayesian model, Markov random Field, supervised learning, image interpretation.

1. INTRODUCTION

Hyperspectral images are mainly interpreted via two widely used techniques, namely spectral unmixing (SU) and classification. SU aims at retrieving elementary components (referred to as endmembers) present in the image and the corresponding proportions within each pixel [1]. Conversely, classification assigns a unique label to each pixel using a predetermined nomenclature [2]. Both analysis own distinct advantages making them complementary. In particular, unmixing is an unsupervised subpixel analysis relying on physical descriptions of the observations [1, 3, 4]. To the contrary, supervised classification provides a semantic description of the hyperspectral image relying on external labeled data. Classification methods are extensively used to interpret remote sensing images and in particular hyperspectral images because of the multitude of available methods and the quality of their results [5–8]. Despite its potential interest in hyperspectral image analysis, the joint exploitation of the high-level (classification) and low-level (unmixing) approaches has been barely proposed [9, 10]. This paper proposes to fill this gap.

In [11], the authors proposed a Bayesian model as well as a corresponding algorithm to perform unmixing and spatial clustering according to the homogeneity of abundance vectors, which is a property also exploited recently in [12]. This paper extends this approach to handle the availability of a set of labeled data, akin to any conventional supervised framework which allows end-user to provide a set of labeled pixels. More precisely, the spectral-spatial clustering proposed in [11] is enriched to be also informed by the classification step. Moreover, since the classification step is reciprocally informed by the clustering, this model allows errors in the ground-truth labels to be identified and corrected. Indeed, label mistakes in user-provided labeled data is a well known issue when conducting supervised classification since they can impair the training process [13, 14]. By exploiting both labeled data and clustering, the robustness to labeling errors of the obtained classifier is improved, as illustrated in [15]. The resulting Bayesian model allows abundance vectors, clustering label and classification labels to be estimated simultaneously. Consequently, the proposed algorithm produces a hierarchical description of the hyperspectral image in terms of unmixing, spectral-spatial segmentation and thematic classification.

The paper is organized as follows. Section 2 presents the model and more particularly develops the strategy to handle the unmixing, clustering and classification tasks jointly. A Markov chain Monte Carlo (MCMC) method is derived in Section 3 to sample according to the posterior of interest. Section 4 shows the results obtained on synthetic and real data. Conclusion is reported in Section 5.

2. FROM UNMIXING TO CLASSIFICATION

2.1. Problem statement

This work aims at performing the unmixing and classification of an hyperspectral image composed of P pixel spectra $\mathbf{y}_p \in \mathbb{R}^D$ ($p \in \mathcal{P} \triangleq \{1, \ldots, P\}$) which are measured in D spectral bands. The R endmembers $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ associated to elementary components of the mixing are assumed to be known. Two main quantities will be estimated: an abundance map $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_P]$ and a classification map $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_P]$, where $\mathbf{a}_p$ is the abundance vector associated to the pth pixel and $\omega_p \in \mathcal{J} \triangleq \{1, \ldots, J\}$ is the classification label relating this pixel to a particular semantic class, with J the number of classes. Each pixel is also characterized by a cluster label $z_p \in \mathcal{K} \triangleq \{1, \ldots, K\}$ assigning this pixel to a group of homogeneous pixels. To reflect possible heterogeneity of the semantic class, each class is assumed to contain one or several clusters. Within a traditional supervised classification context, a partial ground-truth map of the image is provided as a training set. Formally, a subset of the P pixels is assigned class labels $c_p \in \mathcal{J}$. These labels are assumed to be potentially corrupted, e.g., due to some misclassification by the end-user. In the following, $\mathcal{L} \subset \mathcal{P}$ stands for the set of indexes of this subset of labeled pixels and, conversely, $\mathcal{U} = \mathcal{P} \backslash \mathcal{L}$ denotes the set of indexes of the remaining (i.e., non-labeled) pixels. The hierarchical Bayesian model described hereafter is derived to perform the low-level (e.g., unmixing) and high-level (e.g., classification) tasks jointly, while simultaneously exploiting the high-level external information $\mathbf{c}_{\mathcal{L}} \triangleq \{c_p, p \in \mathcal{L}\}$. This model, represented in Fig. 1, is detailed in the following paragraphs.

Par of this work has been supported Centre National d'Études Spatiales (CNES) and Occitanie region.

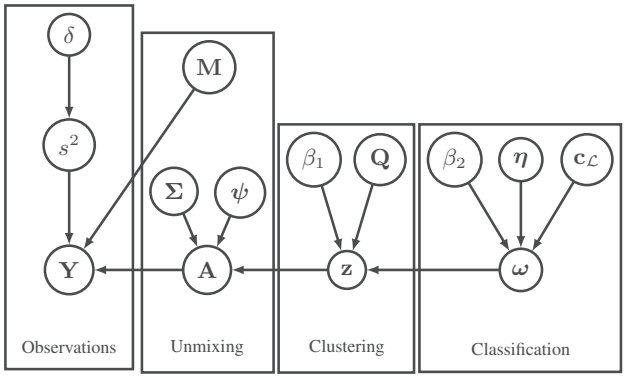

Fig. 1: Directed acyclic graph of the proposed model.

2.2. Bayesian hierarchical model

Mixing model – Following the conventional linear mixing model [1], each pixel of the observed hyperspectral image is described as a linear combination of R endmembers corrupted by an additive noise

$$\mathbf{y}_p = \mathbf{M}\mathbf{a}_p + \mathbf{n}_p \tag{1}$$

where $\mathbf{n}_p$ is the noise associated to the pth pixel, assumed to be white and Gaussian, i.e., $\mathbf{n}_p|s^2 \sim \mathcal{N}(\mathbf{0}_D, s^2\mathbf{I}_D)$, with $\mathbf{I}_D$ the $D \times D$ identity matrix and $\mathbf{0}_D$ the D-dimensional zero vector. It is worth noting that the proposed model can be easily adapted to handle non-whiteness or even non-Gaussian noises. Following the approach proposed by [11], a conjugate inverse-gamma distribution and a non-informative Jeffreys prior are used as a prior distributions for the noise variance s^2 and the associated hyperparameter

$$s^2|\delta \sim \mathcal{IG}(1, \delta), \quad p(\delta) \propto \frac{1}{\delta}\mathbb{1}_{\mathbb{R}^+}(\delta) \tag{2}$$

where $\propto$ means proportional to and $\mathbb{1}_{\mathbb{R}^+}(\cdot)$ is the indicator function on $\mathbb{R}^+$. These choices ensure a straightforward estimation of the noise parameters. The observation model is then complemented by clustering and classification models described in the following sections.

Clustering model – As a bridge between the low-level task (i.e. unmixing) and the high-level task (i.e. classification), an additional clustering step is introduced in the model. More precisely, capitalizing on [11], the hyperspectral pixels are assumed to belong to K distinct clusters. To identify this belonging, each pixel is assigned a cluster label $z_p \in \mathcal{K} \triangleq \{1, \ldots, K\}$. Within a given cluster, the pixels are assumed to share common statistical behavior, i.e., abundance vectors are assumed to be characterized by identical 1st and 2nd order moments, justifying the following *a priori* distribution

$$\mathbf{a}_p|z_p = k, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k \sim \mathcal{N}(\boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k). \tag{3}$$

In this work, the mean vector $\boldsymbol{\psi}_k$ and covariance matrice $\boldsymbol{\Sigma}_k$ are assumed to be unknown and are also included within the Bayesian model to be estimated. Thus, as unknown parameters, they are also assigned prior distributions. First, Dirichlet distributions are chosen as priors for $\boldsymbol{\psi}_k$ ($k \in \mathcal{K}$)

$$\boldsymbol{\psi}_{k,r} \sim \text{Dir}(\mathbf{1}). \tag{4}$$

This choice allows the positivity and sum-to-one constraints classically used in SU to be imposed on the mean behavior of the abundance vectors. The $\boldsymbol{\Sigma}_k$ covariance matrix is chosen as $\boldsymbol{\Sigma}_k = \text{diag}(\sigma^2_{k,1}, \ldots, \sigma^2_{k,R})$ and conjugate inverse-gamma *a priori* distributions are assigned to the variance $\sigma^2_{k,r}$, assumed to be a priori independent

$$\sigma^2_{k,r} \sim \mathcal{IG}(a_\sigma, b_\sigma) \tag{5}$$

where $a_\sigma = 1$ and $b_\sigma = 0.01$ are chosen to define non-informative priors.

One of the main contributions of the proposed model lies in the prior model designed for the cluster labels $\mathbf{z} = [z_1, \ldots, z_P]$. A non-homogeneous MRF [16] is designed to promote two behaviors, namely, spatial coherence of the clustering and consistency between clusters and classes. This non-homogeneous MRF is composed of two terms, each associated with one of this behavior. Firstly, as in [11], a Potts-Markov potential [17] of granularity parameter β_1 is employed to promote spatial regularity of the cluster labels. Secondly, a local potential is introduced to promote coherence between cluster labels $\mathbf{z}$ and classification labels $\boldsymbol{\omega}$. This potential is parametrized by a $K \times J$ interaction matrix $\mathbf{Q}$. Thus, the prior conditional probability of z_p is defined as follows

$$\mathbb{P}[z_p = k|\mathbf{z}_{\mathcal{V}(p)}, \omega_p, q_{k,\omega_p}] \propto \exp\left(V_1(k, \omega_p, q_{k,\omega_p}) + \sum_{p' \in \mathcal{V}(p)} V_2(k, z_{p'})\right) \tag{6}$$

where $\mathcal{V}(p)$ stands for the set of indexes of pixels neighboring the pth pixel (a conventional 4-neighbor structure in our case) and $q_{k,j}$ is the kth component of the jth column of $\mathbf{Q}$. The two terms $V_1(\cdot)$ and $V_2(\cdot)$ are the potential of coherence with classification and the Potts-Markov potential defined by, respectively,

$$V_1(k, j, q_{k,j}) = \log(q_{k,j})$$
$$V_2(k, z_{p'}) = \beta_1\delta(k, z_{p'})$$

with $\delta(\cdot, \cdot)$ the Kronecker function. The matrix $\mathbf{Q}$ gathers a set of coefficients that encodes the relation of each pair $(k, j) \in \mathcal{K} \times \mathcal{J}$ of cluster and classification labels. More precisely, a high value of $q_{k,j}$ promotes the assignation, for a given pixel of class label $\omega_p = j$, a cluster label $z_p = k$. More generally, the coefficients defining a given column of $\mathbf{Q}$ provide an implicit description of a given class in terms of cluster contributions. Thus, Dirichlet distribution is assigned as a prior for each column $\mathbf{q}_j$ of $\mathbf{Q}$ assumed to be independent

$$\mathbf{q}_j \sim \text{Dir}(\mathbf{1}). \tag{7}$$

It is worth noting that, in the special case where $\beta = 0$ (i.e., no spatial regularization is imposed on the cluster labels), the choice of this Dirichlet distribution leads to the following posterior conditional distribution

$$\mathbf{q}_j|\mathbf{z}, \boldsymbol{\omega} \sim \text{Dir}(n_{1,j} + 1, \ldots, n_{K,j} + 1) \tag{8}$$

where $n_{k,j} = \#\{p|z_p = k, \omega_p = j\}$ is the number of pixels belonging to cluster k and class j. In particular, the posterior mean of $q_{k,j}$ can be written as $\mathbb{E}[q_{k,j}|\mathbf{z}, \boldsymbol{\omega}] = \frac{n_{k,j}+1}{\sum_{i=1}^K n_{i,k}+K}$ which is an empirical estimator of $\mathbb{P}[z_p = k|\omega_p = j]$.

Robust classification model – The prior probabilities for the classification labels $\boldsymbol{\omega}$ are defined similarly to the prior probabilities of the cluster labels $\mathbf{z}$ defined in the previous paragraph. Two potentials are tailored to define an appropriate non-homogeneous MRF as a prior model for the $\boldsymbol{\omega}$. The first potential is a spatial regularization similar to the potential $V_2(\cdot)$. The second potential exploits the external ground-truth information $\mathbf{c}_{\mathcal{L}}$ available for pixels whose indexes belong to $\mathcal{L}$ and reduces to a non-informative potential for pixels whose indexes belong to $\mathcal{U}$. Thus, this prior probability is defined as

$$\mathbb{P}[\omega_p = j|\boldsymbol{\omega}_{\mathcal{V}(p)}, c_p, \eta_p] \propto \exp\left(W_1(j, c_p, \eta_p) + \sum_{p' \in \mathcal{V}(p)} W_2(j, \omega_{p'})\right)$$

with

$$W_1(j, c_p, \eta_p) = \begin{cases} \begin{cases} \log(\eta_p), & \text{if } j = c_p \\ \log(\frac{1-\eta_p}{J-1}), & \text{otherwise} \end{cases} & \text{if } p \in \mathcal{L} \\ -\log(J) & \text{otherwise} \end{cases}$$

and

$$W_2(j, \omega_{p'}) = \beta_2 \delta(j, \omega_{p'}).$$

The potential $W_1(\cdot)$ is parametrized by $\eta_p \in (0, 1)$, a user-provided hyperparameter reflecting the confidence the user owns in the classification label c_p for the pth pixel. In the case of a high confidence in this external data ($\eta_p \approx 1$), the estimated classification label tends to be equal to the user-provide one, i.e., $\omega_p = c_p$. In the case of a lower confidence, a pixel can be assigned an estimated classification label ω_p different from the label c_p provided by the end-user. Thus, the proposed hierarchical model allows this ground-truthed external information to be corrected, resulting in a supervised classification which is robust to the presence of mislabeling.

3. GIBBS SAMPLER

Bayesian estimators associated with the parameters defining the model introduced in the previous sections are approximated thanks to a MCMC algorithm [18]. This algorithm generates samples asymptotically distributed according to the joint posterior distribution of the parameters using Gibbs moves. These samples are then used to approximate the maximum a posteriori (MAP) estimators of the cluster and classification labels, which consists in retaining the most recurrent labels. Then, the minimum mean square error (MMSE) estimators of the remaining parameters is approximated by empirical averages over the samples. This Gibbs sampling strategy consists in sampling according to the conditional posterior distributions of each parameter. These distributions are derived in the following paragraphs. More details are available in [19].

Abundances – Given the mixture model (1) and the prior (3), the abundance vectors are a posteriori distributed according to the following multivariate Gaussian distributions

$$\begin{aligned} &p(\mathbf{a}_p|\mathbf{y}_p, z_p = k, s^2, \boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k) \\ &\quad \propto |\boldsymbol{\Lambda}_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{a}_p - \boldsymbol{\mu}_k)^t \boldsymbol{\Lambda}_k^{-1}(\mathbf{a}_p - \boldsymbol{\mu}_k)\right) \end{aligned}$$

with $\boldsymbol{\mu}_k = \boldsymbol{\Lambda}_k(\frac{1}{s^2}\mathbf{M}^t\mathbf{y}_p + \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\psi}_k)$ and $\boldsymbol{\Lambda}_k = (\frac{1}{s^2}\mathbf{M}^t\mathbf{M} + \boldsymbol{\Sigma}_k^{-1})^{-1}$.

Cluster labels – As the cluster label z_p is a discrete random variable, its sampling can be achieved by evaluating the conditional probabilities associated with all possible values of $z_p \in \mathcal{K}$

$$\begin{aligned} &\mathbb{P}(z_p = k|\boldsymbol{\psi}_k, \boldsymbol{\Sigma}_k, \omega_p = j, q_{k,j}) \\ &\quad \propto |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{a}_p - \boldsymbol{\psi}_k)^t \boldsymbol{\Sigma}_k^{-1}(\mathbf{a}_p - \boldsymbol{\psi}_k)\right) \\ &\quad \times q_{k,j} \exp\left(\beta_1 \sum_{p' \in \mathcal{V}(p)} \delta(k, z'_p)\right). \end{aligned} \quad (9)$$

Interaction matrix – The conditional distribution of each column $\mathbf{q}_j$ ($j \in \mathcal{J}$) of the interaction matrix $\mathbf{Q}$ can be expressed as follows

$$p(\mathbf{q}_j|\mathbf{z}, \mathbf{Q}_{\setminus j}, \boldsymbol{\omega}) \propto \frac{\prod_{k=1}^K q_{k,j}^{n_{k,j}}}{C(\boldsymbol{\omega}, \mathbf{Q})} \mathbb{1}_{\mathbb{S}}(\mathbf{q}_j)$$

where $C(\boldsymbol{\omega}, \mathbf{Q})$ is the partition function of the MRF (introduced as a normalization constant), $\mathbf{Q}_{\setminus j}$ denotes the matrix $\mathbf{Q}$ whose jth column has been removed and $\mathbb{1}_{\mathbb{S}}(\cdot)$ is the indicator function of the probability simplex which ensures the positivity and sum-to-one constraints. In particular, when $\beta_1 = 0$ (i.e., no spatial regularization is imposed on the cluster labels), this conditional posterior distribution reduces to the Dirichlet distribution (8), which is easy to sampled from. More advanced sampling strategies should be considered when $\beta_1 > 0$ [19].

Classification map – Similarly to the cluster labels, the classification labels are sampled by evaluating their conditional probabilities for all possible labels $j \in \mathcal{J}$, while distinguishing the cases when an external data c_p is available or not for the considered pth pixel. More precisely, when $p \in \mathcal{U}$, this probability reads

$$\begin{aligned} &\mathbb{P}[\omega_p = j|z_p, \mathbf{z}_{\nu(p)}, \mathbf{q}_j, \boldsymbol{\omega}_{\mathcal{V}(p)}, c_p, \eta_p] \\ &\quad \propto \frac{q_{z_p,j}\pi_j \exp\left(\beta_2 \sum_{p' \in \nu(p)} \delta(j, \omega_{p'})\right)}{\sum_{k'=1}^K q_{k',j} \exp\left(\beta_1 \sum_{p' \in \nu(p)} \delta(k', z_{p'})\right)}. \end{aligned}$$

Conversely, when $p \in \mathcal{L}$, this posterior probability is

$$\begin{aligned} &\mathbb{P}[\omega_p = j|z_p, \mathbf{z}_{\nu(p)}, \mathbf{q}_j, \boldsymbol{\omega}_{\mathcal{V}(p)}, c_p, \eta_p] \\ &\quad \propto \frac{q_{z_p,j} \exp\left(\beta_2 \sum_{p' \in \nu(p)} \delta(j, \omega_{p'})\right)}{\sum_{k'=1}^K q_{k',j} \exp\left(\beta_1 \sum_{p' \in \nu(p)} \delta(k', z_{p'})\right)} \times \begin{cases} \eta_p, \text{when } \omega_p = c_p \\ \frac{1-\eta_p}{C-1}, \text{otherwise.} \end{cases} \end{aligned}$$

4. EXPERIMENTS

Synthetic images – To assess the effectiveness of the proposed model, experiments are first conducted on synthetic data. These synthetic images are generated from a clustering map drawn from a Potts-MRF. The classification map is then obtained by grouping together several clusters. Abundance vectors are generated from a Dirichlet distribution of fixed parameters for each cluster. Finally, pixels of the hyperspectral images are generated using the mixing model (1) with real spectra composed of $D = 413$ spectral bands and a Gaussian noise with SNR$=$ 30dB. To illustrate, two particular instances of the cluster and classification maps generated according to this protocol are represented in Fig. 2. The first case corresponds to a 100×100 image composed of $R = 3$ endmembers, $K = 3$ clusters and $J = 2$ classes (Image 1). The second case is a 200×200 image with $R = 9$ endmembers, $K = 12$ clusters and $J = 5$ classes (Image 2).

For both images (Images 1 & 2), the upper quarter of the classification map has been used as external training data $\{c_p\}_{p \in \mathcal{L}}$. To evaluate the robustness of the proposed model face to mislabeling, these labels have been corrupted by replacing the correct label class by another with a probability equal to a particular corruption rate. The confidence η_p ($p \in \mathcal{L}$) in the provided ground truth has been set equal to the percentage of correct labels. Classification results have been compared to those obtained by conducting a mixture discriminant analysis (MDA) [20]. MDA has been applied following two different ways: either directly on the pixel spectra, or on the abundance vectors estimated with the proposed model. Fig. 3 shows the quality of the classification evaluated with Cohen's kappa as a function of the corruption rate. The obtained results underline the expected robustness of the model.

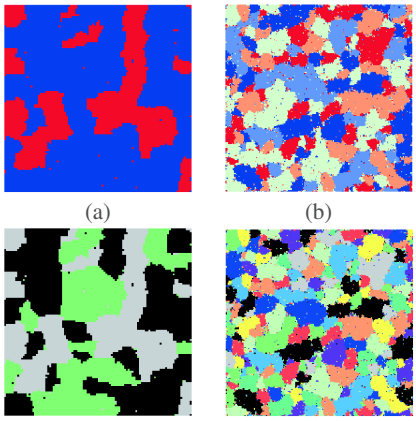

Fig. 2: Top, Image 1: classification (a) and clustering (b) maps. Bottom, Image 2: classification (c) and clustering (d) maps.

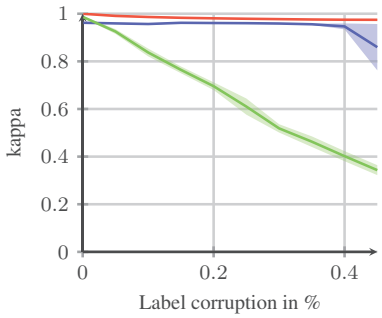

Fig. 3: Cohen's kappa as a function of label corruption: MDA with measured reflectance (green), MDA with abundance vectors (blue) and proposed model (red). Shaded areas correspond to standard deviation resulting from 20 trials.

Moreover, to illustrate the richness of the proposed model in term of possible interpretation, Fig. 4 represents the $\mathbf{Q}$ matrices estimated for Images 1 & 2. These matrices lead to explicit descriptions of the data structure by providing the distribution of the clusters with respect to the different classes. For example, class $\sharp 5$ in Image 2 gathers clusters $\sharp 3$ and $\sharp 5$.

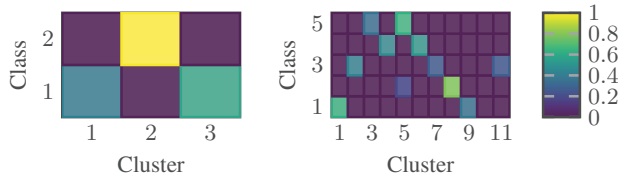

Fig. 4: Estimated $\mathbf{Q}$ matrix for Image 1 (left) and Image 2 (right).

Real images – Finally, experiments are conducted on a real 600×600 hyperspectral image composed of $D = 349$ spectral bands (after removing the bands of low SNR) obtained within the MUESLI mission[1]. First, $R = 7$ endmembers have been extracted by conducting a vertex component analysis [21]. The classification ground-truth provided by the experts after a field campaign is composed of $L = 6$ classes (summer crops, straw cereals, wooded area, buildings, bare/hayed land, meadow) and the left half of the ground-truth is used as external data $\mathbf{c}_{\mathcal{L}}$ with a confidence $\eta_p = 95\%$ ($\forall p \in \mathcal{L}$).

[1] http://fauvel.mathieu.free.fr/pages/muesli.html

The number of clusters K has been set to a high value, i.e., $K = 40$. The proposed algorithm is expected to empty most of these clusters. Results in term of classification accuracy obtained by the proposed method are compared to those obtained with a state-of-the-art random forest (RF) classifier, known to be particularly robust to labeling errors [22]. Parameters of the RF classifier are optimized using 5-folds cross-validation (50 trees, maximum depth of 20). The quantitative results are averaged over 10 trials.

Table 1: Classification results averaged over 10 trials ($\pm$ standard deviation).

	Cohen's kappa	Time (s)
Proposed model	**0.737 ($\pm$ 0.030)**	6651 ($\pm$ 62)
Random forest	0.695 ($\pm$ 0.003)	**16 ($\pm$ 0.2)**

Experiment results reported in Table 1 show significant better classification results for the proposed model on this particular image. Nevertheless, this result is obtained at the cost of more extensive computations induced by the MCMC algorithm, as underlined in the same table. However, it is worth noting that the proposed method also provides additional parameters of interest, in terms of abundance and cluster maps. To illustrate, results obtained for a particular trial are displayed in Fig. 5.

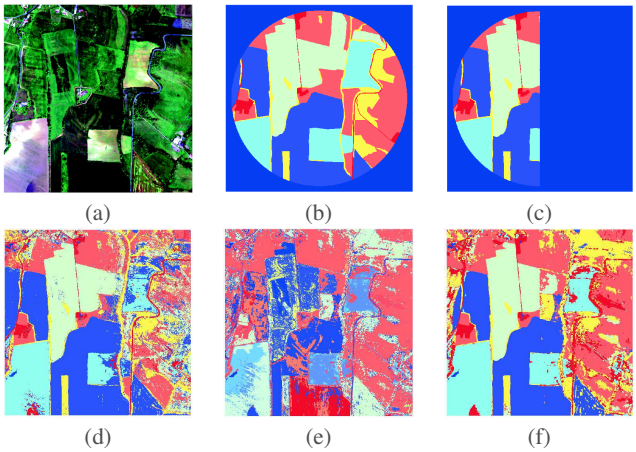

Fig. 5: Real data: (a) pseudo-colored image, (b) expert ground-truth, (c) training ground-truth, (d) RF classification, (e) obtained clustering and (f) obtained classification (with $\beta_1 = 0$ and $\beta_2 = 1.0$).

5. CONCLUSION

This paper introduced a new Bayesian model to perform spectral unmixing, clustering and robust classification jointly. Through the clustering step, the two well-admitted hyperspectral analysis methods, namely unmixing and classification, were conducted in a unified framework, benefiting from low-level and high-level descriptions of the data simultaneously. Interestingly, akin to any conventional supervised classification setup, external ground-truth data could be provided. However, the proposed model allowed corrupted ground-truth labels to be taken into account and corrected, resulting in a supervised classification robust to mislabeling. Results conducted on synthetic and real hyperspectral datasets illustrated good performance in term of classification and underlined the robustness of the model in case of label errors in training data. Future works will focus on the generalization of the proposed model to handle other low-level tasks, i.e., different from spectral unmixing.

6. REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 5, pp. 354–379, 2012.

[2] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, and others, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.

[3] 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, pp. 4355–4368, 2009.

[4] D. C. Heinz and C. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 529–545, 2001.

[5] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, pp. 45–54, 2014.

[6] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, pp. 542–546, 2011.

[7] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, pp. 652–675, 2013.

[8] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, pp. 4865–4876, 2011.

[9] A. Villa, J. Li, A. Plaza, and J. M. Bioucas-Dias, "A new semi-supervised algorithm for hyperspectral image classification based on spectral unmixing concepts," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*. IEEE, 2011, pp. 1–4.

[10] J. Li, I. Dópido, P. Gamba, and A. Plaza, "Complementarity of discriminative classifiers and spectral unmixing techniques for the interpretation of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, pp. 2899–2912, 2015.

[11] 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, pp. 5–16, 2013.

[12] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4775–4789, 2016.

[13] D. F. Nettleton, A. Orriols-Puig, and A. Fornells, "A study of the effect of different types of noise on the precision of supervised learning techniques," *Artif. Intell. Rev.*, vol. 33, pp. 275–306, 2010.

[14] C. Pelletier, S. Valero, J. Inglada, N. Champion, C. Marais Sicre, and G. Dedieu, "Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series," *Remote Sens.*, vol. 9, p. 173, 2017.

[15] C. Bouveyron and S. Girard, "Robust supervised classification with mixture models: Learning from data with uncertain labels," *Pattern Recognit.*, vol. 42, pp. 2649–2658, 2009.

[16] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Springer Science & Business Media, 2009.

[17] F.-Y. Wu, "The potts model," *Rev. Mod. Phys.*, vol. 54, p. 235, 1982.

[18] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, ser. Springer Texts in Statistics. New York, NY: Springer New York, 2004.

[19] A. Lagrange, M. Fauvel, S. May, and N. Dobigeon, "Hierarchical Bayesian image analysis: From low-level modeling to robust supervised learning," 2017. [Online]. Available: https://arxiv.org/abs/1712.00368

[20] T. Hastie and R. Tibshirani, "Discriminant Analysis by Gaussian Mixtures," *J. Roy. Stat. Soc. Ser. B*, vol. 58, pp. 155–176, 1996.

[21] 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, pp. 898–910, 2005.

[22] A. Folleco, T. M. Khoshgoftaar, J. V. Hulse, and L. Bullard, "Identifying learners robust to low quality data," in *Proc. IEEE Int. Conf. on Inf. Reuse and Integr.*, 2008, pp. 190–195.