Robust Linear Spectral Unmixing Using Anomaly Detection

Yoann Altmann, *Member, IEEE*, Steve McLaughlin, *Fellow, IEEE*, and Alfred Hero, *Fellow, IEEE*

Abstract—This paper presents a Bayesian algorithm for linear spectral unmixing of hyperspectral images that accounts for anomalies present in the data. The model proposed assumes that the pixel reflectances are linear mixtures of unknown endmembers, corrupted by an additional nonlinear term modeling anomalies, and additive Gaussian noise. A Markov random field is used for anomaly detection based on the spatial and spectral structures of the anomalies. This allows outliers to be identified in particular regions and wavelengths of the data cube. A Bayesian algorithm is proposed to estimate the parameters involved in the model yielding a joint linear unmixing and anomaly detection algorithm. Simulations conducted with synthetic and real hyperspectral images demonstrate the accuracy of the proposed unmixing and outlier detection strategy for the analysis of hyperspectral images.

Index Terms—Hyperspectral imagery, unsupervised spectral unmixing, Bayesian estimation, MCMC, anomaly detection.

I. Introduction

SPECTRAL unmixing (SU) of hyperspectral images (HSI) has been the subject of intensive interest over the last two decades. It consists of distinguishing the materials and quantifying their proportions in each pixel of an observed image. The SU problem has been widely studied for applications where pixel reflectances are linear combinations of pure component spectra (called endmembers) [1], [2]. However, as explained in [2], the linear mixing model (LMM) can be inappropriate for some hyperspectral images, such as those containing sand-like materials or where relief is present in the scene. Moreover, LMM-based methods can also fail when the data are corrupted by (sparse) outliers, especially when extracting the endmembers from the scene. Nonlinear mixing models (NLMMs) have been proposed in the hyperspectral image literature and can be divided into two main classes [3], [4]. The first class of NLMMs consists of physical models based on the nature of the environment (e.g., intimate mixtures [5] and multiple scattering effects

[6]–[8]). The second contains more flexible models allowing a wider range of nonlinearities to be approximated [9], [10].

Here, we consider a general mixing model for spectral unmixing which assumes that the observed pixels result from a convex combination of the endmembers of the scene, corrupted by an additive term modelling deviations from the classical LMM (e.g., outliers, variability, nonlinear effects) and additive Gaussian Noise. The number of endmembers is assumed to be known whereas their spectral signatures are unknown. It is interesting to note that many nonlinear models in the literature, including polynomial models [6]–[8] can be expressed in a similar manner. Here, the additional terms are assumed to be *a-priori* independent of the endmembers and/or their proportions (abundances), as in [11], [12]. This class of models for robust linear SU allows for general deviations from the LMM to be handled in blind source separation methods, i.e., nonlinear effects, outliers or possible endmember variability [13]. In [12], spatial and spectral sparsity structures were considered for the additional term since deviations from the LMM can occur in specific regions or spectral bands of the HSI. This is typically the case when outliers are present, but also when nonlinear effects (relief) occurs and when the reflectance of materials present has significant variations in particular spectral ranges (e.g. due to natural variability of vegetation). In this paper, we extend [12] by introducing a probabilistic 3D Ising model for spatial and spectral influence of outliers thus allowing for more flexible group-sparsity structures for the support sets of anomalies, whereas [12] assumed the support sets of outliers to have a fixed structure. Moreover, the algorithm presented in this paper allows the estimation of the Ising model parameters directly from the data.

Adopting a Bayesian framework, we assign prior distributions to the unknown model parameters to include available information (such as parameter constraints) within the estimation procedure. In particular, an Ising Markov random field is introduced to model spatial and spectral correlations for the anomalies. The joint posterior distribution of the unknown parameter vector is then derived. Since classical Bayesian estimators cannot be easily computed from this joint posterior, a Markov chain Monte Carlo (MCMC) method is used to generate samples according to this posterior. More precisely, we construct an efficient stochastic gradient MCMC (SGMCMC) algorithm [14] that simultaneously estimates the endmember and abundance matrices along with the Ising hyperparameters.

The main contributions of this work are threefold:

1) We develop a new hierarchical outlier model taking into account spatial and spectral correlations through Markovian dependencies, this contrasts with the model

proposed in [12] which considered a fixed outlier structure. This flexible model is embedded within the mixing model for robust unsupervised linear spectral unmixing *via* anomaly detection.

2) An adaptive MCMC algorithm is proposed to compute the Bayesian estimates of interest and perform Bayesian inference. This algorithm is equipped with a stochastic optimisation adaptation mechanism that automatically adjusts the parameters of the Markov random field by maximum marginal likelihood estimation, thus removing the need to set the regularisation parameters by cross-validation.

3) We show the benefits of the proposed flexible model for linear spectral unmixing of synthetic and real hyperspectral images. Specifically, we demonstrate the ability of the proposed algorithm to detect structured anomalies thus enhancing endmember and abundance estimation.

The remaining sections of the paper are organized as follows. Section II introduces the mixing model for robust linear SU of HSIs, followed by Section III which summarizes the likelihood and the priors assigned to the unknown parameters of the model. The resulting joint posterior distribution and the Gibbs sampler used to sample from it are summarized in Section IV. A generalization of the proposed Bayesian model for robust Bayesian subspace identification is proposed in Section V. Some simulation results conducted on synthetic data are shown and discussed in Section VI. Conclusions and future work are reported in Section VIII.

II. PROBLEM FORMULATION

We consider a set of N observed pixels/spectra $\mathbf{y}_n = [y_{1,n}, \ldots, y_{L,n}]^T, n \in \{1, \ldots, N\}$ where L is the number of spectral bands. Each of these spectra is assumed to result from a linear combination of R unknown endmembers $\mathbf{m}_r$, corrupted by possible additive outliers and Gaussian noise. The observation model can be expressed as

$$\mathbf{y}_n = \sum_{r=1}^{R} \mathbf{m}_r a_{r,n} + \mathbf{r}_n + \mathbf{e}_n$$

$$= \mathbf{M}\boldsymbol{a}_n + \mathbf{r}_n + \mathbf{e}_n, \quad n = 1, \ldots, N \tag{1}$$

where $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ is the spectrum of the rth material present in the scene and $a_{r,n}$ is its corresponding proportion (abundance) in the nth pixel. In (1), $\mathbf{e}_n$ is an additive independently but non identically distributed zero-mean Gaussian noise sequence with diagonal covariance matrix $\boldsymbol{\Sigma}_0 = \mathrm{diag}(\boldsymbol{\sigma}^2)$, denoted as $\mathbf{e}_n \sim \mathcal{N}(\mathbf{e}_n; \mathbf{0}_L, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_L^2]^T$ is the vector of the L noise variances and $\mathrm{diag}(\boldsymbol{\sigma}^2)$ is an $L \times L$ diagonal matrix containing the elements of the vector $\boldsymbol{\sigma}^2$. Moreover, $\mathbf{r}_n$ denotes the outlier vector of the nth pixel. Note that the usual matrix and vector notations $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R]$ and $\boldsymbol{a}_n = [a_{1,n}, \ldots, a_{R,n}]^T$ have been used in the second row of (1).

As a consequence of physical constraints, the abundance vectors $\boldsymbol{a}_n$ satisfy the following positivity and sum-to-one constraints

$$\sum_{r=1}^{R} a_{r,n} = 1, \ a_{r,n} > 0, \quad \forall r \in \{1, \ldots, R\}. \tag{2}$$

The problem investigated in this paper is to estimate the endmember matrix $\mathbf{M}$, the abundance matrix $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$, the noise variances in $\boldsymbol{\sigma}^2$ and the outlier matrix $\mathbf{R} = [\mathbf{r}_1 \ldots, \mathbf{r}_N]$ from the observation matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$. To solve this problem, we propose a hierarchical Bayesian model and a sampling method to estimate the unknown parameters.

III. ROBUST BAYESIAN LINEAR UNMIXING (RBLU)

A. Likelihood

Eq. (1) implies that $\mathbf{y}_n | \mathbf{M}, \boldsymbol{a}_n, \mathbf{r}_n, \boldsymbol{\sigma}^2 \sim \mathcal{N}(\mathbf{y}_n; \mathbf{M}\boldsymbol{a}_n + \mathbf{r}_n, \boldsymbol{\Sigma}_0)$. Assuming independence between noise sequences of the N observed pixels, the likelihood of the observation matrix $\mathbf{Y}$ can be expressed as $f(\mathbf{Y} | \mathbf{M}, \mathbf{A}, \mathbf{R}, \boldsymbol{\sigma}^2) \propto$

$$|\boldsymbol{\Sigma}_0|^{-N/2} \mathrm{etr} \left[-\frac{(\mathbf{Y} - \mathbf{M}\mathbf{A} - \mathbf{R})^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{Y} - \mathbf{M}\mathbf{A} - \mathbf{R})}{2} \right] \tag{3}$$

where $\propto$ means "proportional to" and $\mathrm{etr}(\cdot)$ denotes the exponential trace.

B. Parameter priors

1) Prior for the abundance matrix $\mathbf{A}$: Each abundance vector can be written as $\boldsymbol{a}_n = [\mathbf{c}_n^T, a_{R,n}]^T$ with $\mathbf{c}_n = [a_{1,n}, \ldots, a_{R-1,n}]^T$ and $a_{R,n} = 1 - \sum_{r=1}^{R-1} a_{r,n}$. The LMM constraints (2) impose that $\mathbf{c}_n$ belongs to the simplex $\mathcal{S} = \left\{ \mathbf{c} \middle| c_r \geq 0, \forall r \in 1, \ldots, R-1, \sum_{r=1}^{R-1} c_r \leq 1 \right\}$. To reflect the lack of prior knowledge about the abundances, a uniform prior is assigned for each vector $\mathbf{c}_{n,n} \in \{1, \ldots, N\}$, i.e., $f(\mathbf{c}_n) \propto \mathcal{SS}(\mathbf{c}_n)$, where $\mathcal{SS}(\cdot)$ is the indicator function defined on the simplex $\mathcal{S}$. When prior knowledge about the abundances are available, the uniform prior can be replaced by more informative priors such as (mixtures of) Dirichlet distributions [15] or Gaussian mixtures using logistic coefficients [16]. Assuming prior independence between the N abundance vectors $\{\boldsymbol{a}_n\}_{n=1,\ldots,N}$ leads to the following joint prior distribution

$$f(\mathbf{C}) = \prod_{n=1}^{N} f(\mathbf{c}_n), \tag{4}$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N]$ is an $(R-1) \times N$ matrix.

2) Prior for the endmember matrix $\mathbf{M}$: To reflect the lack of prior knowledge about the endmembers, we use the following multivariate truncated Gaussian prior

$$f(\mathbf{M} | \xi) \propto \prod_{r=1}^{R} \mathcal{N}_{(\mathbb{R}^+)^L} (\mathbf{m}_r; \mathbf{0}, \xi \mathbf{I}_L) \tag{5}$$

where ξ is fixed to a large value, to ensure endmember positivity while using a weakly informative prior. Note that (5) is considered in order to handle the case where the data are not normalized. If the data are actual reflectance values, a prior can be introduced ensuring that the endmember spectra belong to $(0, 1)$, such as a uniform, beta [17] or Gaussian distribution. Note that the prior can also include prior information from an endmember extraction algorithm, as in [18], [19].

3) Prior for the noise variances: A Jeffreys' prior is chosen for the noise variance in each spectral band σ_ℓ^2, i.e., $f(\sigma_\ell^2) \propto \sigma_\ell^{-2} \mathbf{1}_{\mathbb{R}^+}(\sigma_\ell^2)$ where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ denotes the indicator function defined on $\mathbb{R}^+$, which reflects the absence of knowledge about these parameters. Again, these non-informative priors can be easily replaced by conjugate inverse-Gamma priors to include prior knowledge available about the noise levels. Assuming prior independence between the noise variances, we obtain

$$f(\boldsymbol{\sigma}^2) = \prod_{\ell=1}^{L} f(\sigma_\ell^2). \tag{6}$$

4) Priors of the outliers: As in [10], [12], the outliers are assumed to be sparse, i.e., at most of the pixels and spectral bands, the outliers are expected to be exactly equal to zero. To model the outlier sparsity, we factorize the outlier matrix as

$$\mathbf{R} = \mathbf{Z} \odot \mathbf{X}, \tag{7}$$

where $\mathbf{Z} \in \{0,1\}^{L \times N}$ is a label matrix, $\mathbf{X} \in \mathbb{R}^{L \times N}$ and $\odot$ denotes the Hadamard (termwise) product. This decomposition allows one to decouple the location of the sparse components from their values. More precisely, $z_{\ell,n} = [\mathbf{Z}]_{\ell,n} = 1$ if an outlier is present in the ℓth spectral band of the nth observed pixel with value equal to $r_{\ell,n} = x_{\ell,n}$. A conjugate Gaussian prior is used for $\mathbf{X}$, i.e.,

$$f(\mathbf{X}|s^2) = \prod_{\ell,n} \mathcal{N}(x_{\ell,n}; 0, s^2), \tag{8}$$

where s^2 controls the prior energy of the outliers. Note that (8) allows the outliers to be negative. Other conjugate priors, such as exponential or truncated Gaussian priors, could be used instead of (8), e.g., to enforce outlier positivity. The next paragraph presents the prior considered for the label matrix $\mathbf{Z}$.

5) Label matrix: For many applications, the locations of outliers are likely to be spectrally (e.g., water absorption bands) and/or spatially (e.g. weakly represented components, shadowing effects) correlated. An effective way to take correlated outliers/nonlinear effects into account is to consider Markov random fields (MRF) to build a prior for the label matrix $\mathbf{Z}$ [10]. MRFs assume that the distribution of a label $z_{\ell,n}$ conditionally to the other labels of the image equals the distribution of this label vector conditionally to its neighbors, i.e., $\mathrm{P}(z_{\ell,n}|\mathbf{Z}_{\backslash z_{\ell,n}}) = \mathrm{P}(z_{\ell,n}|\mathbf{Z}_{\mathcal{V}_{\ell,n}})$, where $\mathcal{V}_{\ell,n}$ is the index set of the neighbors of $z_{\ell,n}$, $\mathbf{Z}_{\backslash z_{\ell,n}}$ denotes the matrix $\mathbf{Z}$ whose element $z_{\ell,n}$ has been removed and $\mathbf{Z}_{\mathcal{V}_{\ell,n}}$ is the subset of $\mathbf{Z}$ composed of the elements whose indexes belong to $\mathcal{V}_{\ell,n}$. In this study, we consider that the spatial and spectral correlations can be different and thus consider two different neighborhoods. We decompose the neighborhood $\mathcal{V}_{\ell,n}$ as $\mathcal{V}_{\ell,n} = \mathcal{V}_{\ell,n}^L \cup \mathcal{V}_{\ell,n}^N$ where $\mathcal{V}_{\ell,n}^N$ (resp. $\mathcal{V}_{\ell,n}^L$) denotes the spatial (resp. spectral) neighborhood of $z_{\ell,n}$. In this paper, we consider an Ising model that can be expressed as

$$\mathrm{P}(\mathbf{Z}|\boldsymbol{\beta}') = \frac{1}{B(\boldsymbol{\beta}')} \exp\left[\boldsymbol{\beta}^T \boldsymbol{\phi}(\mathbf{Z}) + \phi_0(\mathbf{Z}, \beta_0)\right] \tag{9}$$

where $\boldsymbol{\beta} = [\beta_N, \beta_L]^T$, $\boldsymbol{\beta}' = [\boldsymbol{\beta}^T, \beta_0]^T$ and

$$\begin{cases} \phi_L(\mathbf{Z}) &= \sum_{n,\ell} \sum_{z_{\ell',n} \in \mathcal{V}_{\ell,n}^L} \delta(z_{\ell,n} - z_{\ell',n}), \\ \phi_N(\mathbf{Z}) &= \sum_{n,\ell} \sum_{z_{\ell,n'} \in \mathcal{V}_{\ell,n}^N} \delta(z_{\ell,n} - z_{\ell,n'}), \\ \phi(\mathbf{Z}) &= [\phi_L(\mathbf{Z}), \phi_N(\mathbf{Z})]^T, \\ \phi_0(\mathbf{Z}, \beta_0) &= \beta_0 \sum_{n,\ell}(1 - z_{\ell,n}) + (1 - \beta_0) \sum_{n,\ell} z_{\ell,n}, \end{cases}$$

and $\delta(\cdot)$ denotes the Kronecker delta function. Moreover, $\beta_N > 0$ and $\beta_L > 0$ are hyperparameters that control the spatial and spectral granularity of the MRF and $0 \leq \beta_0 \leq 1$ is an additional parameter that models the probability of having outliers in the image. Specifically, the higher the value of β_0, the lower the probability of outliers in the data. The estimation of the proposed Ising model hyperparameters will be discussed in the next section. Different spectral and spatial neighbourhoods can be used in (9). In this paper, we consider a 4-neighbour structure to account for the spatial correlation and a 2-neighbour structure for the spectral dimension.

C. Outlier variance s^2

The following conjugate inverse-Gamma prior is assigned to s^2

$$s^2 \sim \mathcal{IG}(\gamma, \nu), \tag{10}$$

where (γ, ν) are fixed to $(\gamma, \nu) = (10^{-3}, 10^{-3})$ to ensure a weakly informative prior. This choice of hyperparameters reflects the lack of prior information about the outliers variance. However, (10) could be replaced by more informative priors by suitably adapting (γ, ν).

D. Joint posterior distribution

Assuming the parameters $\mathbf{M}, \mathbf{A}, \mathbf{Z}, \mathbf{X}$ and $\boldsymbol{\sigma}^2$ are a priori independent, the joint posterior of the parameter vector $\boldsymbol{\theta} = \{\mathbf{M}, \mathbf{A}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\sigma}^2\}$ and the parameter s^2 can be expressed as

$$f(\boldsymbol{\theta}, s^2 | \mathbf{Y}, \boldsymbol{\phi}, \boldsymbol{\beta}') \propto f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|s^2, \xi, \boldsymbol{\beta}) f(s^2|\gamma, \nu) \tag{11}$$

where

$$f(\boldsymbol{\theta}|s^2, \xi, \boldsymbol{\beta}') = f(\mathbf{M}|\xi) f(\mathbf{A}) f(\boldsymbol{\sigma}^2) f(\mathbf{X}|s^2) \mathrm{P}(\mathbf{Z}|\boldsymbol{\beta}') \tag{12}$$

and $\boldsymbol{\phi} = [\xi, \gamma, \nu]^T$ a vector of model hyperparameters. The MRF parameter vector $\boldsymbol{\beta}'$ will be determined by maximum marginal likelihood estimation during the inference procedure. The directed acyclic graph (DAG) summarizing the structure of proposed Bayesian model is depicted in Fig. 1.

Next we describe an MCMC method for sampling from this posterior distribution to estimate the unknown model parameters.

IV. BAYESIAN INFERENCE

The Bayesian model defined in Section III specifies the joint posterior density for the unknown parameters $\boldsymbol{\theta}, s^2$ given the observations $\mathbf{Y}$ and the hyperparameters ξ, γ, ν and $\boldsymbol{\beta}'$. This
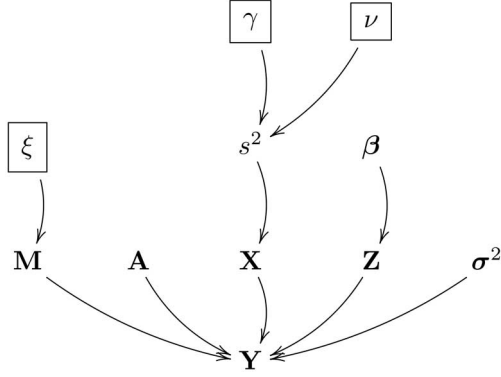
Fig. 1. Directed acyclic graph representing the proposed hierarchical Bayesian model (fixed quantities appear in boxes).

posterior distribution models the complete knowledge about the unknown parameters given the observed data and the prior information available. We propose the following Bayesian estimators for hyperspectral unmixing and nonlinearity detection: the marginal posterior mean or minimum mean square error estimator of the abundance and endmember matrices

$$\left(\widehat{\mathbf{A}}_{\text{MMSE}}, \widehat{\mathbf{M}}_{\text{MMSE}}\right) = \text{E}\left[\mathbf{A}, \mathbf{M} \,\middle|\, \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}'\right], \quad (13)$$

where the expectation is taken with respect to the marginal posterior density $f(\mathbf{A}, \mathbf{M} | \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}')$ (by marginalizing out $\mathbf{Z}, \mathbf{X}, \boldsymbol{\sigma}^2$ and s^2, this density takes into account their uncertainty), the marginal maximum *a posteriori* (MMAP) estimator for the outlier support $\mathbf{Z}$

$$z_{n,\ell}^{MMAP} = \underset{z_{n,\ell} \in \{0,1\}}{\text{argmax}} \, f(z_{n,\ell} | \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}'), \quad (14)$$

and, conditionally on the estimated outliers location, the minimum mean square error estimator of the outlier values

$$r_{n,\ell}^{MMSE} = \text{E}\left[x_{n,\ell} \,\middle|\, z_{n,\ell} = z_{n,\ell}^{MMAP}, \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}'\right], \quad (15)$$

where

$$f(z_{n,\ell} | \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}') = \int f(\boldsymbol{\theta}, s^2 | \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}') \mathrm{d}\boldsymbol{\theta}_{\backslash z_{n,\ell}} \mathrm{d}s^2,$$

where $\text{E}[\cdot]$ denotes the expectation with respect to the conditional marginal density

$$f(x_{n,\ell} | z_{n,\ell}, \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}') \frac{\int f(\boldsymbol{\theta}, s^2 | \mathbf{Y}, \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}') \mathrm{d}\boldsymbol{\theta}_{\backslash z_{n,\ell}} \mathrm{d}s^2}{f(z_{n,\ell} | \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}')}.$$

Note that the outlier estimator (15) is sparse by construction (*i.e.*, $\text{E}\left[x_{n,\ell} \,\middle|\, z_{n,\ell} = 0, \mathbf{Y}, \boldsymbol{\phi}, \hat{\boldsymbol{\beta}}'\right] = 0$).

Computing (13), (14) and (15) is challenging because it requires access to the joint marginal density of $(\mathbf{M}, \mathbf{A})$, the univariate marginal densities of $z_{n,\ell}$ and the joint marginal densities of $(x_{n,\ell}, z_{n,\ell})$, which in turn require computing the posterior (11) and performing an integration over a very high-dimensional space. Fortunately, these can be efficiently approximated with arbitrarily good accuracy by Monte Carlo integration. More precisely, it is possible to compute (13), (14)

and (15) by first using an MCMC computational method to generate samples asymptotically distributed according to (11) and subsequently using these samples to approximate the required marginal probabilities and expectations. Note that in (13), (14) and (15), we have set $\boldsymbol{\beta}' = \hat{\boldsymbol{\beta}}'$, which denotes the maximum marginal likelihood estimator of the Ising regularisation hyperparameter vector $\boldsymbol{\beta}'$ given the observed data $\mathbf{Y}$, i.e.,

$$\hat{\boldsymbol{\beta}}' = \underset{\boldsymbol{\beta}' \in \mathcal{B}}{\text{argmax}} f\left(\mathbf{Y} | \boldsymbol{\phi}, \boldsymbol{\beta}'\right). \quad (16)$$

This is an empirical Bayes approach for specifying $\boldsymbol{\beta}'$ where hyperparameters with unknown values are replaced by point estimates computed from observed data (as opposed to being fixed a priori or integrated out of the model by marginalisation). As explained in [14], this strategy has several important advantages for MRF hyperparameters such as $\boldsymbol{\beta}'$ having intractable conditional distributions. In particular, it allows automatic adjustment of the value of $\boldsymbol{\beta}'$ to each image, producing significantly better estimation results than using a single fixed value of $\boldsymbol{\beta}'$ for all images. Furthermore it has significantly lower computational cost compared to that of competing approaches, such as including $\boldsymbol{\beta}'$ in the model and subsequently marginalizing it during the inference procedure [20].

A. Bayesian estimation algorithm

Exact computation of the MMSE and MMAP estimators (13), (14) and (15) is very challenging because it involves calculating expectations with respect to posterior marginal densities, which in turn require evaluating the full posterior (11) and integrating it over a very high-dimensional space. Exact computation of $\hat{\boldsymbol{\beta}}'$ is also difficult because it involves solving an intractable optimisation problem (it is not possible to evaluate the exact marginal likelihood $f(\mathbf{Y} | \alpha_3)$ or its gradient $\nabla f(\mathbf{Y} | \alpha_3)$). Here we follow the approach proposed in [14] and design a stochastic optimisation and simulation algorithm to compute (13), (14) and (15) simultaneously. We construct an SGMCMC algorithm that simultaneously estimates $\hat{\boldsymbol{\beta}}'$ and generates a chain of N_{MC} samples $\{\mathbf{M}^{(t)}, \mathbf{A}^{(t)}\}_{t=1}^{N_{\text{MC}}}$ asymptotically distributed according to the marginal density $f(\mathbf{M}, \mathbf{A} | \mathbf{Y}, \hat{\boldsymbol{\beta}}')$ (this algorithm is summarised in Algorithm 1 below). Once the samples have been generated, the estimators (13), (14) and (15) are approximated by Monte Carlo integration [21, Chap. 10], i.e.,

$$\widehat{\mathbf{M}}_{MMSE} = \frac{1}{N_{\text{MC}} - N_{\text{bi}}} \sum_{t=N_{\text{bi}}+1}^{N_{\text{MC}}} \mathbf{M}^{(t)},$$

$$\widehat{\mathbf{A}}_{MMSE} = \frac{1}{N_{\text{MC}} - N_{\text{bi}}} \sum_{t=N_{\text{bi}}+1}^{N_{\text{MC}}} \mathbf{A}^{(t)},$$

$$z_{n,\ell}^{MMAP} = \begin{cases} 0 & \text{if } \text{card}(\mathcal{Z}_{n,\ell}) \leq (N_{\text{MC}} - N_{\text{bi}})/2 \\ 1 & \text{else,} \end{cases}$$

$$r_{n,\ell}^{MMSE} =$$

$$\begin{cases} \dfrac{1}{\text{card}(\mathcal{Z}_{n,\ell})} \sum_{z_{n,\ell}^{(t)} \in \mathcal{Z}_{n,\ell}} r_{n,\ell}^{(t)} & \text{if } z_{n,\ell}^{MMAP} = 1 \\ 0 & \text{else,} \end{cases}$$

with $\mathcal{Z}_{n,\ell} = \left\{ z_{n,\ell}^{(t)} | t \in \{N_{\text{bi}} + 1, \ldots, N_{\text{MC}}\}, z_{n,\ell}^{(t)} = 1 \right\}$ and where the samples from the first N_{bi} iterations (corresponding to the transient regime or burn-in period) are discarded. The main steps of this algorithm are detailed in below.

Algorithm 1. RBLU algorithm

1: <u>Fixed input parameters</u>: Number of endmembers R, number of burn-in iterations N_{bi}, total number of iterations N_{MC}

2: <u>Initialization $(t = 0)$</u>
   - Set $\mathbf{A}^{(0)}, \mathbf{M}^{(0)}, \mathbf{X}^{(0)}, \mathbf{Z}^{(0)}, \boldsymbol{\sigma}^{2(0)}, s^{2(0)}, \boldsymbol{\beta}^{(0)}$

3: <u>Iterations $(1 \leq t \leq N_{\text{MC}})$</u>

4: Sample $\mathbf{Z}^{(t)}$ from (19)

5: Sample $\mathbf{M}^{(t)}$ from (21)

6: Sample $\mathbf{A}^{(t)}$ from (23)

7: Sample $\mathbf{X}^{(t)}$ from (24)

8: Sample $\boldsymbol{\sigma}^{2(t)}$ from (26)

9: Sample $s^{2(t)}$ from (27)

10: **if** $t < N_{\text{bi}}$ **then**

11:     Sample $\mathbf{Z}' \sim \mathcal{K}(\mathbf{Z}|\mathbf{Z}^{(t)}, \boldsymbol{\beta}'^{(t-1)})$

12:     Set $\beta_N^{(t)} = \mathcal{P}_{[0,B_t]}\left(\beta_N^{(t-1)} + \Delta\beta_N\right)$ with

13:     $\Delta\beta_N = \delta_t \left[\frac{d}{d\beta_N} \log f(\mathbf{Z}|\boldsymbol{\beta}') - \frac{d}{d\beta_N} \log f(\mathbf{Z}'|\boldsymbol{\beta}') \right].$

14:     Set $\beta_L^{(t)} = \mathcal{P}_{[0,B_t]}\left(\beta_L^{(t-1)} + \Delta\beta_L\right)$ with

15:     $\Delta\beta_L = \delta_t \left[\frac{d}{d\beta_L} \log f(\mathbf{Z}|\boldsymbol{\beta}') - \frac{d}{d\beta_L} \log f(\mathbf{Z}'|\boldsymbol{\beta}') \right].$

16:     Set $\beta_0^{(t)} = \mathcal{P}_{[0,1]}\left(\beta_0^{(t-1)} + \Delta\beta_0\right)$ with

17:     $\Delta\beta_0 = \delta_t \left[\frac{d}{d\beta_0} \log f(\mathbf{Z}|\boldsymbol{\beta}') - \frac{d}{d\beta_0} \log f(\mathbf{Z}'|\boldsymbol{\beta}') \right].$

18: **else**

19:     Set $\boldsymbol{\beta}'^{(t)} = \boldsymbol{\beta}'^{(t-1)}$

20: **end if**

21: Set $t = t + 1$.

---

1) Sampling the labels: It can be seen from (11) that

$$f(z_{n,\ell} = i | \mathbf{Y}, \boldsymbol{\theta}_{\backslash z_{n,\ell}}, \boldsymbol{\beta}', s^2) \propto \tilde{\pi}_{n,\ell}^{(i)}, \quad \forall(n,\ell), \qquad (17)$$

where $i \in \{0, 1\}$ and

$$\log\left(\tilde{\pi}_{n,\ell}^{(i)}\right) = -\frac{(y_{\ell,n} - \mathbf{m}_{\ell,:}\boldsymbol{a}_n - i x_{\ell,n})^2}{2\sigma_\ell^2} - \boldsymbol{\beta}^T \boldsymbol{\phi}(\mathbf{Z}) - \phi_0(\mathbf{Z}, \beta_0). \qquad (18)$$

Consequently, the label $z_{n,\ell}$ can be drawn from its conditional distribution by drawing randomly from $\{0, 1\}$ with probabilities given by

$$f(z_{n,\ell} = i | \mathbf{Y}, \boldsymbol{\theta}_{\backslash z_{n,\ell}}, \boldsymbol{\beta}', s^2) = \frac{\tilde{\pi}_{n,\ell}^{(i)}}{\tilde{\pi}_{n,\ell}^{(0)} + \tilde{\pi}_{n,\ell}^{(1)}}. \qquad (19)$$

2) Sampling the endmembers: It can be easily shown that

$$f(\mathbf{M}|\mathbf{Y}, \boldsymbol{\theta}_{\backslash\mathbf{M}}, s^2, \phi, \boldsymbol{\beta}') = \prod_{\ell=1}^{L} f(\mathbf{m}_{\ell,:}|\mathbf{Y}, \boldsymbol{\theta}_{\backslash\mathbf{m}_{\ell,:}}, s^2, \xi), \qquad (20)$$

i.e., the rows of $\mathbf{M}$, denoted as $\{\mathbf{m}_{\ell,:}\}$ are *a posteriori* independent (conditioned on the other parameters). Moreover,

$$\mathbf{m}_{\ell,:}\Big|\mathbf{Y}, \boldsymbol{\theta}_{\backslash\mathbf{m}_{\ell,:}}, s^2, \xi \sim \mathcal{N}_{(\mathbb{R}^+)^R}\left(\mathbf{m}_{\ell,:}; \tilde{\mathbf{m}}_{\ell,:}, \mathbf{S}_\ell^{(\mathbf{M})}\right). \qquad (21)$$

where

$$\tilde{\mathbf{m}}_{\ell,:} = \sigma_\ell^{-2} \tilde{\mathbf{y}}_{\ell,:} \mathbf{A}^T \mathbf{S}_\ell^{(\mathbf{M})}$$
$$\mathbf{S}_\ell^{(\mathbf{M})} = \left(\sigma_\ell^{-2} \mathbf{A}\mathbf{A}^T + \xi^{-2}\mathbf{I}_L\right)^{-1} \qquad (22)$$

and $\tilde{\mathbf{y}}_{\ell,:}$ is the ℓth row of the $L \times N$ matrix $\widetilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{R}$. Sampling from (21) can be achieved efficiently by using the Hamiltonian method recently proposed in [22] or by successive sampling from R truncated Gaussian distributions (via Gibbs sampling).

3) Sampling the abundances: In a similar fashion to obtaining (20), it can be easily shown that

$$f(\mathbf{C}|\mathbf{Y}, \boldsymbol{\theta}_{\backslash\mathbf{C}}, s^2, \phi, \boldsymbol{\beta}') = \prod_{n=1}^{N} f(\mathbf{c}_n|\mathbf{y}_n, \boldsymbol{\theta}_{\backslash\mathbf{c}_n}, s^2), \qquad (23)$$

i.e., the columns of $\mathbf{A}$ are *a posteriori* independent (conditioned on the other parameters). Moreover, the conditional distribution of $\mathbf{c}_n|\mathbf{y}_n, \boldsymbol{\theta}_{\backslash\mathbf{c}_n}, s^2$ is a multivariate Gaussian distribution restricted to the simplex $\mathcal{S}$, which can be sampled efficiently using the method proposed in [22].

4) Sampling the latent variable matrix $\mathbf{X}$: In a similar manner to the abundance matrix in (23), the elements of X are *a posteriori* independent (conditioned on the other parameters) and can be sampled independently. Since the prior (8) is conjugate to the Gaussian distribution, the full conditional distribution of $x_{\ell,n}$ is given by

$$x_{\ell,n}\big|y_{\ell,n}, \boldsymbol{\theta}_{\backslash x_{\ell,n}}, s^2, \boldsymbol{\beta}' \sim \mathcal{N}\left(x_{\ell,n}; \tilde{x}_{\ell,n}, \tilde{\sigma}_{\ell,n}^2\right) \qquad (24)$$

where

$$\tilde{x}_{\ell,n} = z_{\ell,n}(y_{\ell,n} - \mathbf{m}_{\ell,:}\boldsymbol{a}_n)\frac{\tilde{\sigma}_{\ell,:}^2}{\sigma_\ell^2}$$
$$\tilde{\sigma}_{\ell,n}^2 = \frac{\sigma_\ell^2 s^2}{\sigma_\ell^2 + z_{\ell,n}s^2}. \qquad (25)$$

5) Sampling the noise variances: Sampling the noise variances can be easily achieved by sampling from the following L independent inverse-Gamma distributions

$$\sigma_\ell^2\Big|\mathbf{Y}, \boldsymbol{\theta}_{\sigma_\ell^2}, s^2 \sim \mathcal{IG}\left(\sigma_\ell^2; \frac{N}{2}, \frac{\|\mathbf{y}_{\ell,:} - \mathbf{m}_{\ell,:}\mathbf{A} - \mathbf{r}_{\ell,:}\|^2}{2}\right) \qquad (26)$$

6) Sampling the outlier variance s^2: Finally, in a similar fashion to the noise variances, it can be shown that

$$s^2\Big|\mathbf{Y}, \boldsymbol{\theta}, \phi \sim \mathcal{IG}\left(\frac{NL}{2} + \gamma, \nu + \sum_{n,\ell} \frac{x_{\ell,n}^2}{2}\right) \qquad (27)$$

7) Updating the Ising regularisation model parameter vector $\boldsymbol{\beta}'$: If the marginal likelihood $f(\mathbf{Y}|\phi, \boldsymbol{\beta}')$ was tractable,

we could update $\boldsymbol{\beta}'$ from one MCMC iteration to the next by using a classic gradient descent step

$$\boldsymbol{\beta}'^{(t)} = \boldsymbol{\beta}'^{(t-1)} + \delta_t \nabla \log f(\mathbf{Y}|\boldsymbol{\phi}, \boldsymbol{\beta}'^{(t-1)}),$$

with $\delta_t = t^{-3/4}$, such that $\boldsymbol{\beta}'^{(t)}$ converges to $\hat{\boldsymbol{\beta}}'$ as $t \to \infty$. However, this gradient has two levels of intractability, one due to the marginalisation of $(\boldsymbol{\theta}, s^2)$ and another one due to the intractable normalising constant of the Ising model. We address this difficulty by following the approach proposed in [14]; that is, by replacing $\nabla \log f(\mathbf{Y}|\boldsymbol{\phi}, \boldsymbol{\beta}'^{(t)})$ with an estimator computed with the samples generated by the MCMC algorithm at iteration t, and a set of two auxiliary variables $(\mathbf{Z}') \sim \mathcal{K}(\mathbf{Z}|\mathbf{Z}^{(t)}, \boldsymbol{\beta}'^{(t-1)})$ generated with an MCMC kernel $\mathcal{K}$ with target density (9) (in our experiments we used a Gibbs sampler implemented using a colouring scheme such that half of the elements of $\mathbf{Z}'$ are generated in parallel). The updated value $\boldsymbol{\beta}'^{(t)}$ is then projected onto the domain $\mathcal{B}_t = [0, B_t] \times [0, B_t] \times [0, 1]$ to guarantee the constraints of β_N, β_L and β_0 and the stability of the stochastic optimisation algorithm, where B_t is an arbitrarily large upper bound on β_N and β_L that can be increased at every iteration. In our experiments we have used $B_t = 10$.

It is worth mentioning that if it was possible to simulate the auxiliary variables $\mathbf{Z}'$ exactly from (9) then the estimator of $\nabla \log f(\mathbf{Y}|\boldsymbol{\phi}, \boldsymbol{\beta}'^{(t-1)})$ used in Algorithm 1 would be unbiased and as a result $\boldsymbol{\beta}'^{(t)}$ would converge exactly to $\hat{\boldsymbol{\beta}}'$. However, exact simulation from (9) is not computationally feasible and therefore we resort to the MCMC kernel $\mathcal{K}$ and obtain a biased estimator of $\nabla \log f(\mathbf{Y}|\boldsymbol{\phi}, \boldsymbol{\beta}'^{(t-1)})$ that drives $\boldsymbol{\beta}'^{(t)}$ to a neighbourhood of $\hat{\boldsymbol{\beta}}'$ [14]. We found that computing this is significantly less expensive than alternative approaches, e.g., using an approximate Bayesian computation algorithm [20], and that it leads to very accurate unmixing results.

Note that in Algorithm 1, $\mathcal{P}_{[0, B_t]}(\cdot)$ denotes the projection onto $[0, B_t]$. In this paper, we propose an MCMC method that sequentially updates $\mathbf{Z}, \mathbf{M}, \mathbf{A}, \mathbf{X}, \boldsymbol{\sigma}^2$ and s^2 at each sampler iteration. However, some of these variables, such as $(\mathbf{Z}, \mathbf{X}), (\mathbf{M}, \mathbf{X})$ or $(\mathbf{A}, \mathbf{X})$ could be updated simultaneously or could be re-sampled within a given sampler iteration, thus improving the mixing properties of the proposed sampling scheme, c.f. [12].

V. GENERALIZATION TO ROBUST BAYESIAN PRINCIPAL COMPONENT ANALYSIS

As mentioned previously, one of the main contributions of this paper is the introduction of a 3D Ising model to model the possible correlation between the outlier locations. This outlier model has been applied to linear SU in Section III. However, this model can be applied to many other blind source separation (subspace identification) and inverse problems (nonlinear spectral unmixing) where the observations are corrupted by additive Gaussian noise and outliers. In this section, we discuss the generalization of the robust Bayesian principal component analysis model studied in [12], [23]. As in [23], we consider the following observation model (expressed in matrix form)

$$\mathbf{Y} = \mathbf{L} + \mathbf{R} + \mathbf{E} \tag{28}$$
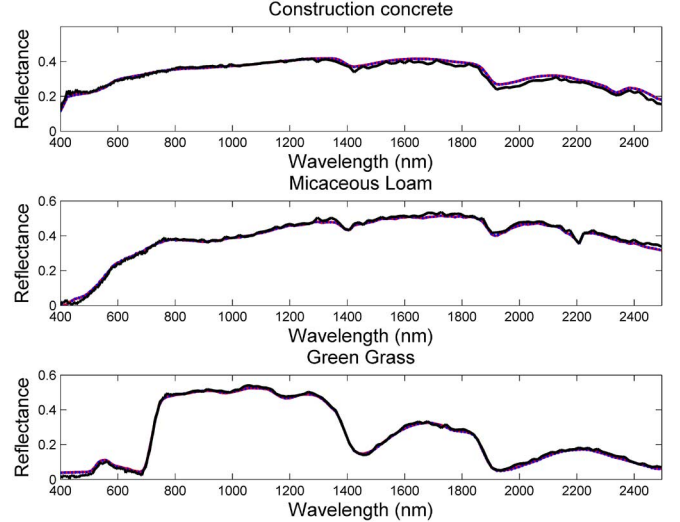

Fig. 2. Actual endmembers (red lines) used to generate the synthetic images and endmembers estimated by VCA (black lines) and RBLU (dashed blue lines) for I_2.

where $\mathbf{R}$ (resp. $\mathbf{E}$) are $L \times N$ matrices representing the outliers (resp. the additive Gaussian noise) and $\mathbf{L} = \mathbf{D}\boldsymbol{\Lambda}\mathbf{W}^T$ corresponds to a low rank matrix. The $L \times R$ matrices $\mathbf{D}$ and $\mathbf{W}$ are matrices of the left- and right-singular vectors, respectively, and $\boldsymbol{\Lambda} = \text{diag}([\lambda_1, \ldots, \lambda_R])$ is a diagonal matrix consisting of the singular values $\{\lambda_r\}_{r=1,\ldots,R}$. Following the model considered in [12], [23], we can set $\lambda_r = \zeta_r \eta_r$ where $\zeta_r \in \{0, 1\}$, which allows estimation of the dimension of the principal subspace. Since the columns of $\mathbf{E}$ are i.i.d., i.e.,

$$\mathbf{e}_n \sim \mathcal{N}(\mathbf{e}_n; \mathbf{0}_L, \boldsymbol{\Sigma}_0) \quad \forall n, \tag{29}$$

we can assign appropriate prior distributions to $\boldsymbol{\sigma}^2, \mathbf{D}, \mathbf{W}, \{\zeta_r\}$ and $\{\eta_r\}$, and consider the Bayesian model proposed in Section III for the outlier matrix $\mathbf{R}$ (i.e., Eqs. (7)–(10)). The resulting *robust Bayesian principal component analysis model* can then be used to estimate the data principal subspace in the presence of non-i.i.d. additive Gaussian noise and possibly correlated outliers. As the posterior distribution of the anomalous outliers can also be estimated, the model also allows the outlier matrix to be estimated. The sampling strategy associated with Bayesian model is similar to those proposed in [12], [23] and that presented in Section IV and is not further developed here due to space constraints.

VI. SIMULATIONS USING SYNTHETIC DATA

A. Linear mixtures corrupted by outliers

The performance of the proposed method, referred to as "RBLU" (Robust Bayesian linear unmixing), is investigated on two synthetic 60×60 pixel hyperspectral images composed of $R = 3$ endmembers and observed at $L = 207$ different spectral bands (see Fig. 2). The abundances of the two images are uniformly distributed in the simplex, defined by the positivity and sum-to-one constraints, and the noise variances set to $\sigma_\ell^2 = 10^{-4}, \forall \ell$, corresponding to an average SNR of 30 dB without

TABLE I
ESTIMATION PERFORMANCE

			RBLU	VCA-FCLS	o-FCLS
I_1	SAM ($\times 10^{-2}$)	$\mathbf{m}_1$	**0.21**	0.68	-
		$\mathbf{m}_2$	**0.17**	0.92	-
		$\mathbf{m}_3$	**0.26**	1.96	-
	RNMSE ($\times 10^{-2}$)		0.68	1.60	**0.67**
I_2	SAM ($\times 10^{-2}$)	$\mathbf{m}_1$	**0.19**	4.03	-
		$\mathbf{m}_2$	**0.20**	3.08	-
		$\mathbf{m}_3$	**0.29**	4.26	-
	RNMSE ($\times 10^{-2}$)		**0.74**	8.27	6.95

TABLE II
OUTLIER DETECTION (I_2): CONFUSION MATRIX

	$z = 0$	$z = 1$	Total
$\hat{z} = 0$	652724	7190	659914
$\hat{z} = 1$	789	84497	85286
Total	653513	91687	745200

TABLE III
ABUNDANCE RNMSE ($\times 10^{-2}$) FOR DIFFERENT OUTLIER ENERGIES
AND PROPORTIONS

Noise variance	Outlier fraction	$s^2 = 0.01$		$s^2 = 0.1$	
		RBLU	o-FCLS	RBLU	o-FCLS
$\sigma^2 = 10^{-4}$	10%	0.77	2.00	**0.74**	6.95
	20%	0.80	3.06	**0.77**	8.89
	30%	**0.86**	3.66	0.87	10.34
$\sigma^2 = 10^{-3}$	10%	**2.19**	2.77	2.21	6.00
	20%	**2.36**	3.65	2.39	8.90
	30%	**2.59**	4.09	2.62	10.57

any anomaly addition. The first image I_1 does not contain outliers whereas the parameter s^2 controlling the outlier power has been set to $s^2 = 0.1$ for the second image I_2. The label matrix of I_2 has been generated using (9) with $\boldsymbol{\beta} = [0.25; 0.25; 0.55]^T$ which leads to approximately 10% of actual outliers in $\mathbf{R}$. The proposed method has been applied to the images with $N_{\mathrm{MC}} = 1000$ iterations (including $N_{\mathrm{bi}} = 300$). The endmember matrix was initialized using VCA [24] and the abundance matrix was initialized using FCLS [1]. The combination VCA-FCLS is also used for performance comparison.

The quality of the unmixing procedures can be measured by comparing the estimated and actual abundance vector using the root normalized mean square error (RNMSE) defined by

$$\mathrm{RNMSE} = \sqrt{\frac{1}{NR} \sum_{n=1}^{N} \| \hat{\boldsymbol{a}}_n - \boldsymbol{a}_n \|^2} \qquad (30)$$

where $\boldsymbol{a}_n$ and $\hat{\boldsymbol{a}}_n$ are the actual and estimated abundance vectors for the nth pixel of the image. The quality of endmember estimation is evaluated by the spectral angle mapper (SAM) defined as

$$\mathrm{SAM} = \arccos \left(\frac{\langle \hat{\mathbf{m}}_r, \mathbf{m}_r \rangle}{\| \hat{\mathbf{m}}_r \| \, \| \mathbf{m}_r \|} \right) \qquad (31)$$

where $\mathbf{m}_r$ is the rth actual endmember and $\hat{\mathbf{m}}_r$ its estimate. The smaller the value of $|\mathrm{SAM}|$, the closer the estimated endmembers to their actual values.

Table I compares the performance of the proposed method and the VCA-FCLS unmixing strategy and shows that the proposed methods outperforms VCA-FCLS in terms of abundance and endmember estimation. Moreover, the confusion matrix of the proposed outlier detection method in Table II illustrates the ability of the method to identify the corrupted data.

Table III compares the abundance estimation performance of RBLU to o-FCLS (which assumes perfectly known endmembers) for different outlier corruption scenarios (proportions and variances) and two noise settings, $\sigma^2 = 10^{-4}$ and $\sigma^2 = 10^{-3}$ (which correspond to SNR of approximately 30dB and

20dB, respectively, when considering data without anomalies). This table shows a general performance degradation of the algorithms when the number of outliers increases. However, although RBLU also estimates the endmembers (jointly with the abundances), the performance degradation is less severe for RBLU than for o-FCLS by virtue of RBLU's outlier detection ability. It is interesting to note that RBLU is also less sensitive than o-FCLS to variations in the outlier variance (o-FCLS abundance estimation performance decreases when the outlier variance increases). In addition, as the proportion of outliers increases, sparsity ceases to be a reliable discriminant and it becomes increasingly difficult to detect the outlier samples. Similarly, if the variances of the noise and the anomalies are similar RBLU will not be able to detect the potential anomalies, confounding the outliers with a fictitious Gaussian noise having larger variance.

For a 64-bit Matlab R2014b implementation on a 3 GHz Intel Xeon quad-core workstation, RBLU required 30min to analyse each image composed of 60×60 pixels (3s for VCA-FCLS). Although RBLU can provide significantly improved endmember and abundance estimates, it has a higher computational cost than VCA-FCLS.

Next we discuss the unmixing problem when both linearly and nonlinearly mixed pixels are present in the image.

B. Unmixing of linear and nonlinear mixtures

As mentioned previously, many nonlinear models of the spectral unmixing literature can be expressed as (1). This is the case, for instance, for polynomial models [6]–[8] introduced to handle multiple scattering effects. The proposed model (1) seems particularly well adapted to detecting nonlinear effects that are often spatially localized, e.g., regions with significant relief variations, and/or spectrally concentrated such as nonlinear interactions (if present) varying smoothly over wavelengths. To evaluate the performance of the RBLU algorithm in terms of endmember estimation, abundance estimation and nonlinearity detection, we synthesised a 60×60 pixel image composed of the three endmembers considered in the previous paragraph. Most of the pixels (75%) were generated according to the classical LMM while the remaining 25% were generated according to a bilinear model, namely, the generalized bilinear model [8]

$$\mathbf{y}_n = \mathbf{M}\boldsymbol{a}_n + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \gamma_{i,j,n} a_{i,n} a_{j,n} \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{e}_n, \qquad (32)$$
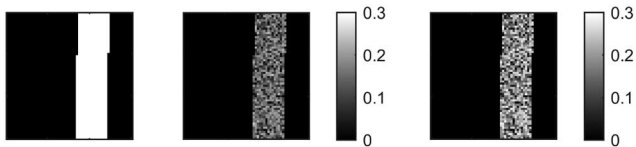
Fig. 3. Left: Actual location of the linearly (black pixels) and nonlinearly (white pixels) mixed pixels. Actual (middle) and estimated (right) anomaly energy in each pixel of the synthetic image composed of linearly and nonlinearly mixed pixels.

TABLE IV

ABUNDANCE AND ENDMEMBER ESTIMATION FOR THE SYNTHETIC IMAGE COMPOSED OF LINEARLY AND NONLINEARLY MIXED PIXELS

	RNMSE ($\times 10^{-2}$)	SAM ($\times 10^{-2}$)		
		$\mathbf{m}_1$	$\mathbf{m}_2$	$\mathbf{m}_3$
o-FCLS	4.44	–	–	–
VCA+FCLS	4.46	**0.35**	**0.55**	0.90
NfindR + FCLS	5.96	0.65	3.09	0.80
o-GBM	2.28	–	–	–
VCA+GBM	**2.15**	**0.35**	**0.55**	0.90
NfindR + GBM	2.26	0.65	3.09	0.80
BLU	4.03	1.76	1.94	1.33
RBLU	3.81	2.79	3.56	**0.59**

where $0 \leq \gamma_{i,j,n} \leq 1$ characterizes the level of interaction between the endmembers $\mathbf{m}_i$ and $\mathbf{m}_j$ in the nth pixel. This choice of nonlinear model is motivated by the fact the polynomial (and in particular bilinear) models, introduced to account for multiple scattering effects, have demonstrated improved performance for certain types of urban and vegetated areas [6]–[8]. The abundances of each pixel (linearly or nonlinearly) mixed were uniformly drawn from the simplex defined by the abundance positivity and sum-to-one constraints. The nonlinearity parameters in (32) were fixed to $\gamma_{i,j,n} = 1$, which corresponds to the choice made in Fan's model [25]. All pixels have been corrupted by a zero-mean additive Gaussian noise i.i.d, with variance σ^2, corresponding to an average SNR of 28dB. Note that although the generated abundance vectors are spatially independent, the positions of the generated anomalies are spatially correlated due to the spatial organization of the linearly and nonlinearly mixed pixels (see Fig. 3 (left)). The RBLU algorithm was applied to the data using $N_{\mathrm{MC}} = 2000$ iterations (including $N_{\mathrm{bi}} = 500$).

Table IV compares the estimation performance of RBLU to the results obtained with 1) BLU [18], 2) VCA+FCLS [1], [24], 3) NfindR + FCLS [1], [26], 4)o-FCLS [1], 5) NfindR [26] followed by the gradient-based inversion step proposed in [8] based on the GBM (32) (referred to as "NfindR + GBM" in the table), 6) VCA [24] followed by the gradient-based inversion step proposed in [8] based on the GBM (32) (referred to as "VCA + GBM" in the table), and 7) the GBM method [8] applied to the data assuming the endmembers are known (referred to as "o-GBM" in the table for oracle-GBM). Due to its anomaly detection ability, RBLU generally provides better abundance estimates than LMM-based methods. Although RBLU does not rely on the GBM, its provides better abundance estimates than BLU, in particular for nonlinearly mixed pixels. For this data set containing pure pixels, N-FindR and VCA

TABLE V

COMPUTATIONAL TIME (IN SECONDS) FOR THE SYNTHETIC IMAGE COMPOSED OF LINEARLY AND NONLINEARLY MIXED PIXELS

o-FCLS	1
VCA+FCLS	3
NfindR + FCLS	3
o-GBM	350
VCA+GBM	354
NfindR + GBM	358
BLU	583
RBLU	1526

can provide better endmember estimates than RBLU, which would be different in the absence of pure pixels. Indeed, in a similar manner to BLU, the proposed RBLU algorithm does not require the pure pixel assumption. However, due to space constraints, simulations conducted on synthetic data which do not contain pure pixels are not presented in this paper which focuses on the anomaly detection capability of RBLU (the interested reader is invited to consult [27] for additional results on data which do not contain pure pixels). We discuss this point in the next section in the context of the RBLU analysis of real hyperspectral images.

Fig. 3 (middle and right) shows the actual and estimated outlier energy in each pixel (i.e., $\|\hat{\mathbf{r}}_n\|_2^2$) and illustrates the ability of RBLU to detect bilinear mixtures in an image containing both linear and non-linear mixtures.

Finally, Table V compares the execution time required by the different methods to analyse the 3600 pixels of the image composed of linearly and nonlinearly mixed pixels on the same basis as before. Although the GBM-based method proposed [8] takes longer than FCLS, it processes the pixels independently and successively. Its run-time could thus be improved (*e.g.*, using parallelization) and its computational complexity could approach that of FCLS. BLU and RBLU are unsupervised unmixing algorithms, which jointly estimate the endmembers and abundances and are based on MCMC methods, which are more computationally demanding. Although some sampling steps can be performed in parallel, the underlying Gibbs samplers are intrinsically sequential processes that require a sufficient number of iterations to explore the posterior distribution of interest. RBLU is more computationally demanding than BLU as it includes additional sampling steps (labels and anomaly values) and the estimation of the Ising model regularization parameters. However, the results presented in this paper illustrate the benefits of the (structured) anomaly detection ability of RBLU on the endmember and abundance estimation performance.

VII. SIMULATIONS USING REAL HYPERSPECTRAL DATA

A. Moffett data set

The first real image considered in this section is composed of $L = 189$ spectral bands and was acquired in 1997 by the AVIRIS satellite. The acquired image covers a region over Moffett Field, CA. A subimage of size 50×50 pixels has been chosen here to evaluate the performance of RBLU. The AVIRIS
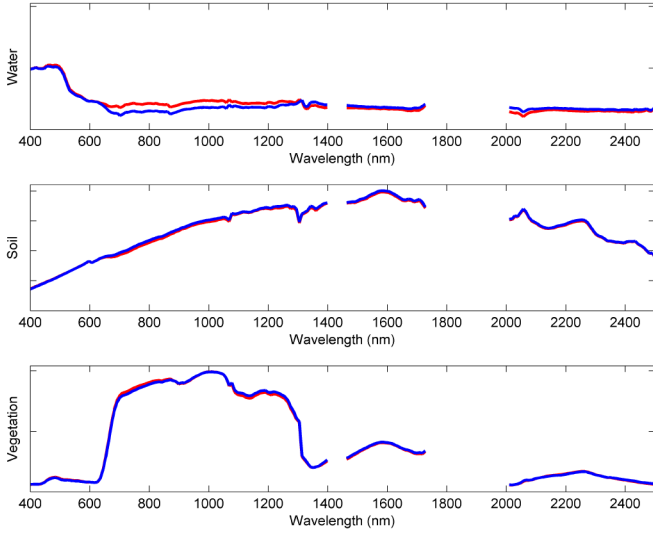
Fig. 4. The $R = 3$ endmembers extracted from the Moffett image by N-FindR (red) and RBLU (blue).
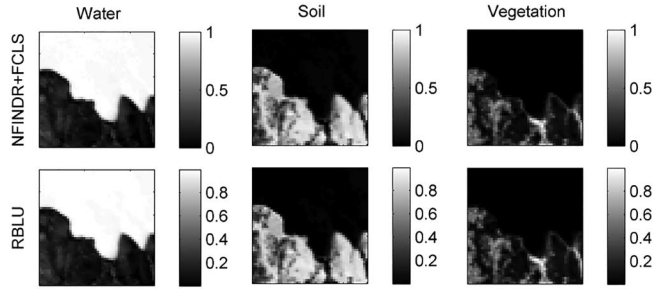

Fig. 5. The $R = 3$ abundance maps associated with the Moffett image and estimated by FCLS (top) and RBLU (bottom).
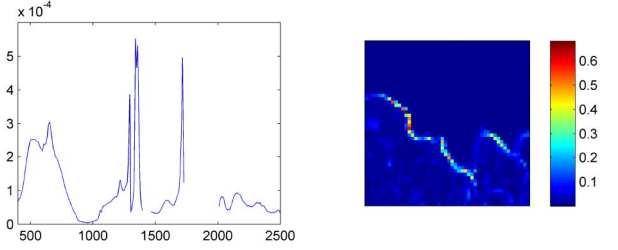

Fig. 6. Left: Average spectral energy of the outliers in the Moffett scene estimated by RBLU. Right: Estimated outlier energy $\|\hat{\mathbf{r}}_n\|_2^2$ in each pixel of the Moffett image.
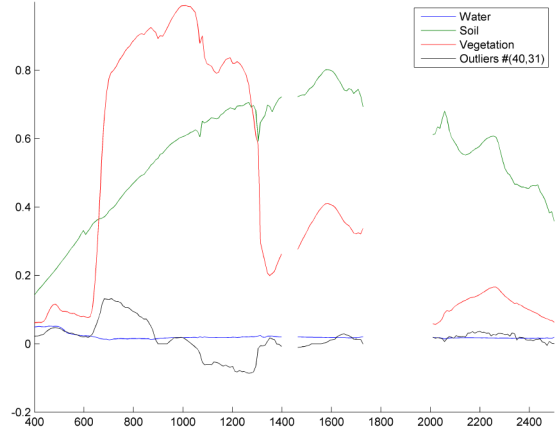

Fig. 7. Endmembers and outlier signature of the pixel $(40, 31)$ estimated by RBLU for the Moffett image.
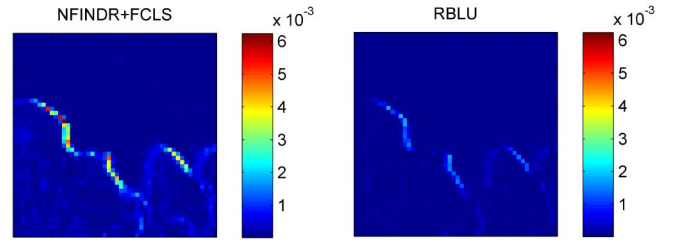

Fig. 8. Reconstruction errors of the Moffett image using N-FindR+FCLS (left) and RBLU.

Moffet Field dataset has been previously used for comparing methods of linear [8], [18], [28]–[30] and nonlinear [8] unmixing. The subimage of interest is mainly composed of water, vegetation, and soil. As in in the previous subsection, the RBLU algorithm was applied with $N_{\mathrm{MC}} = 2000$ iterations (including $N_{\mathrm{bi}} = 500$).

Fig. 4 depicts the endmembers estimated by N-FindR and RBLU. Fig. 5 compares the abundance maps provided by RBLU to those obtained with N-FindR followed by FCLS. These figures show that the endmembers and abundance maps are all in good agreement. In addition to the endmember and abundance estimates, RBLU provides spectral and spatial information about the possible outliers/nonlinearities. Fig. 6 (left) shows average spectral outlier energy over wavelength for the Moffett subimage. In addition to significant outlier levels near the water absorption bands, around 1400 nm and 1800 nm, RBLU identifies important deviations from the linear mixing model for the spectral bands between 400 nm and 800 nm. Fig. 6 (right) displays the estimated outlier energy (i.e., $\|\hat{\mathbf{r}}_n\|_2^2$) map over all pixels in the Moffett scene and shows that the deviations from the linear mixing model are mainly located in the coastal area, which is in agreement with the results obtained in [8], [28]. Fig. 7 compares the estimated spectrum of the outliers in the pixel $(40, 31)$ (located in the coastal area) to the estimated endmembers. Although RBLU promotes groups of

outliers, it does not explicitly enforce spatial nor spectral dependencies for the outlier values. However, the outliers in this pixel seem to be spectrally correlated. These results show that RBLU is able to distinguish structured outliers from Gaussian noise. These results also show that the outlier spectrum and the vegetation spectral signature (in red) seem correlated, especially in the visible spectrum. A possible explanation for this correlation could be local significant changes of vegetation, e.g., chlorophyll and water content, additional vegetation species, multiple scattering effects. Finally, the reconstruction errors obtained by N-FindR+FCLS and RBLU are compared in Fig. 8. Due to its outlier detection ability, RBLU provides lower reconstruction errors in the coastal area, which is the region where outliers/anomalies are the most dominant.

B. Villelongue data set

The second real image considered was acquired in 2010 in the Madonna project and collected by the Hyspex hyperspectral

Fig. 9. True color image of the Villelongue area (left) and sub-image of interest (right).
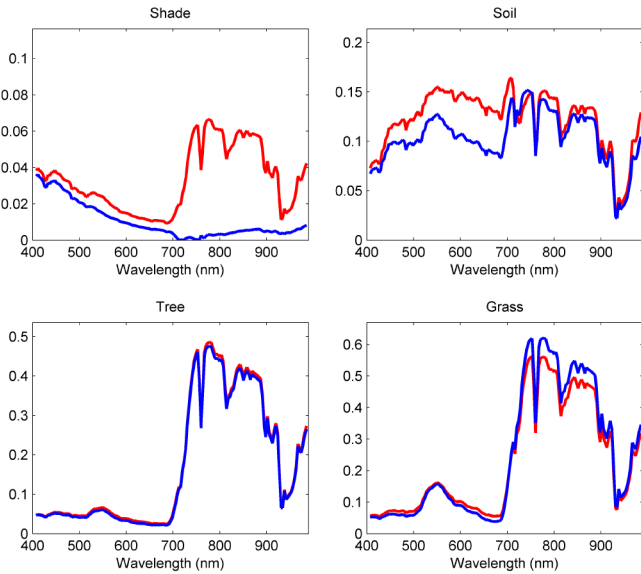

Fig. 10. The $R = 4$ endmembers extracted from the Villelongue image by N-FindR (red) and RBLU (blue).

Fig. 11. The $R = 4$ abundance maps associated with the Villelongue image and estimated by FCLS (right) and RBLU (left).

scanner over Villelongue, France ($00°03'$W and $42°57'$N). $L = 160$ spectral bands were recorded from the visible to near infrared with a spatial resolution of 0.5 m. This dataset has previously been studied in [10], [31], [32] and is mainly composed of forested and urban areas. More details about the data acquisition and pre-processing steps can be found in [31]. A sub-image of size 300×250 pixels is chosen here to evaluate the proposed unmixing procedure and is depicted in Fig. 9. The scene is composed mainly of trees and grass, resulting in $R = 4$ endmembers (soil, grass, trees and shade).

Fig. 10 compares the $R = 4$ endmembers estimated by N-FindR and RBLU for this second real image. Although it is difficult to objectively assess the performance of the two EEAs for this image, it is interesting to note that the results obtained by the two methods are similar for the tree, soil and grass spectra. The shade signature identified by RBLU has a lower amplitude than the spectrum estimated by N-FindR, which is probably due to the absence of completely shadowed pixels in the image (as discussed above, RBLU does not rely on the pure pixel assumption, in contrast to N-FindR). The two methods however lead to abundance maps (depicted in Fig. 11) that are in agreement with the true color image in Fig. 9. In particular, the two algorithms are able to identify the path (soil) in the scene. This path is barely visible in Fig. 9 but its presence can be confirmed using the Google Map image for this region (see Fig. 12).
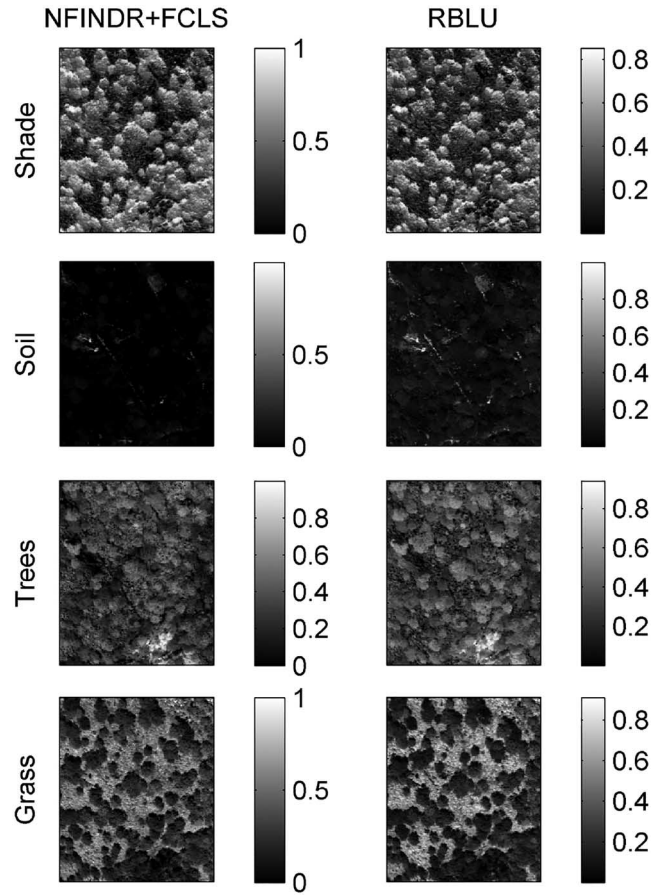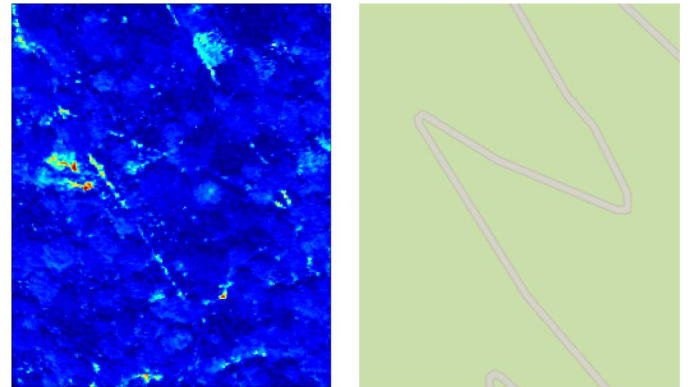

Fig. 12. Soil abundance map estimated by RBLU for the Villelongue image (left) and corresponding Google Map image (right) highlighting the presence of a path in the region of interest.

Fig. 13 (left) depicts the anomaly energy map estimated by RBLU over the Villelongue image and highlights two main regions where significant deviations from the linear mixing model occur. The first region, on the top of Fig. 13 (left) is located where trees are identified and the deviations are likely to be due to the presence of different tree species. Note that this region can also be identified in Fig. 13 (right) (lighter green region). The second region representing a line in the centre of Fig. 13 (left) is more difficult to interpret and is barely visible in the true color image. However, it is interesting to note
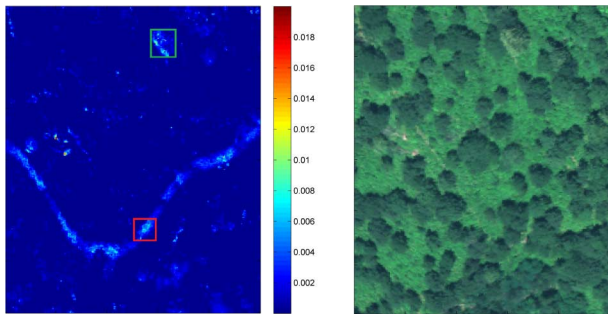
Fig. 13. Left: Anomaly energy map estimated by RBLU for the Villelongue scene. Right: sub-image of interest in true colors.
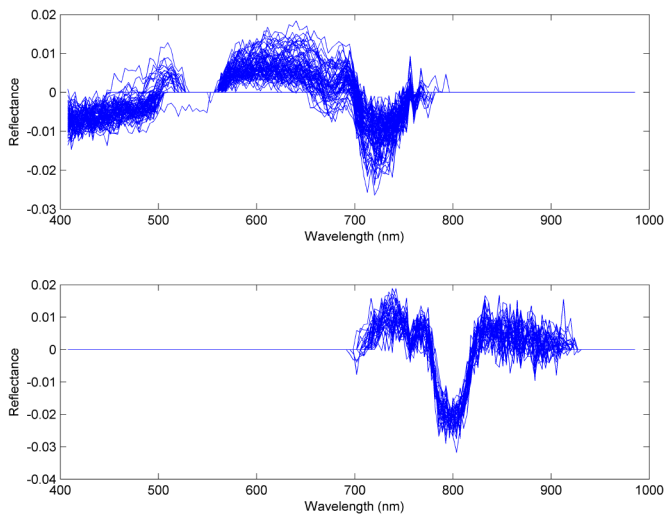

Fig. 14. Anomaly spectra in the green (top) and red (bottom) boxes depicted in Fig 13.

TABLE VI
COMPUTATIONAL TIME (IN SECONDS) TO ANALYSE THE REAL IMAGES
USING RBLU AND N-FINDR+FCLS

	Moffett	Villelongue
RBLU	1590	18360
N-FindR+FCLS	3	31

that the anomalies in this region form a stripe whose energy is higher in the pixels composed of grass than it is in those containing trees. Consequently, it is reasonable to postulate that the physical phenomena causing deviations from the linear mixing model occur on the ground or below the surface, i.e., not in the canopy. One possible cause of these spectral signature changes could be the presence of different kinds of surface vegetation. Finally, Fig. 14 presents examples of anomaly spectra of the first (green box in Fig. 13) and second (red box in Fig. 13) spatial regions. This figure shows that the deviations from the linear mixing model occur in specific spectral bands and are relatively similar within each spatial region. Moreover, the anomalies are spectrally different in the two regions, which confirms they are probably due to different physical phenomena.

Finally, Table VI compares the Matlab execution time required to analyse the Moffett and Villelongue data using RBLU and N-FindR+FCLS.

VIII. CONCLUSION

In this paper, we have proposed a Bayesian algorithm for robust linear spectral unmixing of hyperspectral images that performs joint endmember estimation, abundance estimation and outlier/anomaly detection. Appropriate prior distributions were used to enforce endmember and abundance positivity in addition to abundance sum-to-one constraints. Moreover, a 3D spatial-spectral Ising Markov random field was used to model correlations between outliers. Finally, an adaptive MCMC method was proposed to sample from the resulting posterior distribution in order to estimate the unknown model parameters. Simulations conducted on synthetic data showed superior performance of the proposed method for linear SU and the detection of outliers in hyperspectral images. The proposed method was also applied to real hyperspectral images and provided interesting results in terms of outlier analysis. Future work might include generalization of the proposed method for non-Gaussian observation noise.

REFERENCES

[1] D. C. Heinz and C.-I Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 529–545, Mar. 2001.

[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[3] 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.

[4] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.

[5] B. W. Hapke, "Bidirectional reflectance spectroscopy. I. Theory," *J. Geophys. Res.*, vol. 86, pp. 3039–3054, 1981.

[6] B. Somers *et al.*, "Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1183–1193, 2009.

[7] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE Image Signal Process. Remote Sens. XV*, 2009, vol. 7477, p. 74770I.

[8] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[9] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, Jan. 2013.

[10] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Residual component analysis of hyperspectral images - Application to joint nonlinear unmixing and nonlinearity detection," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2148–2158, May 2014.

[11] N. Dobigeon and C. Févotte, "Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process. Evol. Remote Sens. (WHISPERS)*, Gainesville, FL, USA, Jun. 2013, pp. 1–4.

[12] G. E. Newstadt, A. O. Hero, and J. Simmons, "Robust spectral unmixing for anomaly detection," in *Proc. IEEE-SP Workshop Stat. Signal Process.*, Gold Coast, VIC, Australia, Jul. 2014, pp. 109–112.

[13] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.

[14] M. Pereyra, N. Whiteley, C. Andrieu, and J.-Y. Tourneret, "Maximum marginal likelihood estimation of the granularity coefficient of a Potts-Markov random field within an MCMC algorithm," in *Proc. IEEE-SP Workshop Stat. Signal Process.*, Gold Coast, VIC, Australia, Jul. 2014, pp. 121–124.

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

[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] X. Du, A. Zare, P. Gader, and D. Dranishnikov, "Spatial and spectral unmixing using the beta compositional model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1994–2003, Jun. 2014.

[18] 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. 2657–2669, Nov. 2009.

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

[20] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Estimating the granularity coefficient of a Potts-Markov random field within an MCMC algorithm," *IEEE Trans. Image Processing*, vol. 22, no. 6, pp. 2385–2397, Jun. 2013.

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

[22] A. Pakman and L. Paninski, "Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians," *ArXiv e-prints*, Aug. 2012.

[23] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3419–3430, Dec. 2011.

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

[25] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Remote Sens. Environ.*, vol. 30, no. 11, pp. 2951–2962, Jun. 2009.

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

[27] Y. Altmann, S. McLaughlin, and A. Hero, "Robust linear spectral unmixing using anomaly detection," *ArXiv e-prints*, May 2015 [Online]. Available: http://arxiv.org/abs/1501.03731

[28] O. Besson, N. Dobigeon, and J.-Y. Tourneret, "Minimum mean square distance estimation of a subspace," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5709–5720, Dec. 2011.

[29] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1403–1413, Jun. 2010.

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

[31] D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon, "Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential of hyperspectral imagery," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Jul. 2011, pp. 3672–3675.

[32] Y. Altmann, N. Dobigeon, S. McLaughlin, and J. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2442–2453, May 2013.

Yoann Altmann (S'10–M'14) was born in Toulouse, France, in 1987. He received the degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications, Toulouse, France, and the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse (INP Toulouse), Toulouse, France, both in 2010, and the Ph.D. degree in signal processing from INP Toulouse in 2013. Since 2014, he has been with the Heriot-Watt University, Edinburgh, as a Post-Doctoral Researcher. He conducts his research within the Institute of Sensors, Signals and Systems, School of Engineering and Physical Sciences. His research interests include statistical signal and image processing and Bayesian inverse problems with applications to remote sensing and biomedical imaging.

Steve McLaughlin (F'11) was born in Clydebank, Scotland, in 1960. He received the B.Sc. degree in electronics and electrical engineering from the University of Glasgow, Glasgow, Scotland, in 1981, and the Ph.D. degree in signal processing in communications from the University of Edinburgh, Edinburgh, Scotland, in 1990. From 1981 to 1984, he was a Development Engineer in industry involved in the design and simulation of integrated thermal imaging and fire control systems. From 1984 to 1986, he worked on the design and development of high-frequency data communication systems. In 1986, he joined the Department of Electronics and Electrical Engineering, University of Edinburgh, as a Research Fellow where he studied the performance of linear adaptive algorithms in high noise and nonstationary environments. In 1988, he joined the academic staff at Edinburgh, and from 1991 until 2001, he held a Royal Society University Research Fellowship to study nonlinear signal processing techniques. In 2002, he was a personal Chair in Electronic Communication Systems at the University of Edinburgh. In October 2011, he joined Heriot-Watt University, Edinburgh, Scotland, as a Professor of Signal Processing and Head of the School of Engineering and Physical Sciences. His research interests include the adaptive signal processing and nonlinear dynamical systems theory and their applications to biomedical, energy, and communication systems. Prof. McLaughlin is a Fellow of the Royal Academy of Engineering, Royal Society of Edinburgh, and Institute of Engineering and Technology.

Alfred Hero (F'98) received the B.S. degree (*summa cum laude*) from Boston University, Boston, MA, USA, in 1980, and the Ph.D. degree from Princeton University, Princeton, NJ, USA, in 1984, both in electrical engineering. Since 1984, he has been with the University of Michigan, Ann Arbor, MI, USA, where he is the R. Jamison and Betty Williams Professor of Engineering. His primary appointment is in the Department of Electrical Engineering and Computer Science, and also, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. From 2008 to 2013, he held the Digiteo Chaire d'Excellence, sponsored by Digiteo Research Park in Paris, located at the Ecole Superieure d'Electricite, Gif-sur-Yvette, France. He has held other visiting positions at LIDS Massachusetts Institute of Technology, Cambridge, MA, USA (2006), Boston University (2006), I3S University of Nice, Sophia-Antipolis, France (2001), Ecole Normale Supérieure de Lyon, Lyon, France (1999), Ecole Nationale Supérieure des Télécommunications, Paris, France (1999), Lucent Bell Laboratories* (1999), Scientific Research Labs of the Ford Motor Company, Dearborn, MI, USA (1993), Ecole Nationale Superieure des Techniques Avancees (ENSTA), Palaiseau, France, Ecole Superieure d'Electricite, Paris, France (1990), and M.I.T. Lincoln Laboratory, Lexington, MA, USA (1987–1989). Dr. Hero has been plenary and keynote speaker at several workshops and conferences. He was the President of the IEEE Signal Processing Society (2006–2007). He was a member of the IEEE TAB Society Review Committee (2009), the IEEE Awards Committee (2010–2011), and served on the Board of Directors of the IEEE (2009–2011) as Director of Division IX (Signals and Applications). He served on the IEEE TAB Nominations and Appointments Committee (2012–2014). He is currently a member of the Big Data Special Interest Group (SIG) of the IEEE Signal Processing Society. Since 2011, he has been a member of the Committee on Applied and Theoretical Statistics (CATS) of the US National Academies of Science. His research interests include the data science of high dimensional spatio-temporal data, statistical signal processing, machine learning, and applications to networks, including social networks, multimodal sensing and tracking, database indexing and retrieval, imaging, biomedical signal processing, and biomolecular signal processing. He was the recipient of the University of Michigan Distinguished Faculty Achievement Award (2011), several best paper awards including: an IEEE Signal Processing Society Best Paper Award (1998), a Best Original Paper Award from the Journal of Flow Cytometry (2008), a Best Magazine Paper Award from the IEEE Signal Processing Society (2010), a SPIE Best Student Paper Award (2011), an IEEE ICASSP Best Student Paper Award (2011), an AISTATS Notable Paper Award (2013), an IEEE ICIP Best Paper Award (2013), an IEEE Signal Processing Society Meritorious Service Award (1998), an IEEE Third Millenium Medal (2000), an IEEE Signal Processing Society Distinguished Lecturership (2002), and an IEEE Signal Processing Society Technical Achievement Award (2014).