

# PLANT PRODUCTION SYSTEM MONITORING VIA MULTIPLE SIGNAL CLASSIFICATION AND SPARSE REGRESSION

Marian-Daniel Iordache<sup>1,4</sup>, Ben Somers<sup>1</sup>, Laurent Tits<sup>2</sup>, José M. Bioucas-Dias<sup>3</sup>,  
Antonio Plaza<sup>4</sup> and Pol Coppin<sup>2</sup>

<sup>1</sup> Flemish Institute for Technological Research (VITO), Boeretang 200, BE-2400 Mol, Belgium

<sup>2</sup> Dept. of Biosystems, M3-BIORES, Katholieke Universiteit Leuven,  
W. de Croylaan 34, BE-3001 Leuven, Belgium

<sup>3</sup> Instituto de Telecomunicações and Instituto Superior Técnico, TULisbon, 1049-001, Lisbon, Portugal

<sup>4</sup> Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications,  
University of Extremadura, E-10071 Caceres, Spain

## ABSTRACT

In this paper, we propose a dictionary pruning technique for signal unmixing, which is a recent unmixing concept aimed at extracting the signal contribution of an individual class of endmembers to each observed pixel in a hyperspectral scene. We focus on a specific application in the area of plant production system monitoring. The pruning technique allows to infer the physical state of vegetation in a region of interest on the ground, by selecting appropriate spectra from a very large spectral library designed to model a wide variety of such possible states. On the one hand, *Multiple Endmember Spectral Mixture Analysis* (MESMA), a combinatorial technique, was intensively used to solve the signal unmixing problem. Sparse unmixing, a fast approach dealing with spectral libraries, was recently proposed as a reliable technique for spectral unmixing. The performances of both techniques are influenced by the large number of spectra collected in the libraries. The aim of this paper is two-fold: 1) to present a subspace-based method for library pruning, such that only spectra likely to be present in the image are used in unmixing, which is expected to improve MESMA results, and 2) to introduce sparse unmixing as a reliable signal unmixing technique, given the natural ability of these methods to deal with applications in which a low number of atoms is used to accurately describe every observed vector. The results obtained in a simulated environment show the potential of the proposed scheme to boost the signal unmixing performance both in terms of accuracy and running time.

**Index Terms**— Hyperspectral unmixing, signal unmixing, sparse regression, spectral library, MESMA

## 1. INTRODUCTION

Many research efforts were dedicated to the inference of physical parameters of vegetation on the ground through remote sensing techniques. Hyperspectral images are intensively used in this type of applications, as they are able to capture detailed information about

the observed data at fine spectral level. However, using the observed spectra directly to evaluate vegetation indices might result in inaccurate results, as most of the pixels in hyperspectral data are mixed. Unmixing becomes then necessary, with the goal of separating the vegetation signal from the signal given by other types of materials (*e.g.*, soils). Recently, the use of large spectral libraries (collections of pure spectra acquired in a laboratory or on the field) became a valuable tool in unmixing [1], as it overcomes well-known drawbacks of the endmember extraction techniques: the unknown number of endmembers in one image and (often) the lack of pure pixels in the image.

*Multiple Endmember Spectral Mixture Analysis* (MESMA) (see [2] and the references therein) is a typical combinatorial signal unmixing method using spectral libraries. Given possible classes of endmembers present in the image (*e.g.*, tree, soil, weed), MESMA tests, in each pixel, all the possible combinations of endmembers from each class and retains as a solution the combination which ensures the lowest reconstruction error. However, as the number of possible combinations is usually extremely large, the number of tested (random) combinations is limited because of time restrictions. Thus, the method becomes faster, but the probability to miss the ideal spectra combination increases.

Another class of unmixing techniques making use of spectral libraries is based on the sparsity of the mixtures, as only a few of the spectra collected in the library are present in one pixel [1]. Recently, fast sparse unmixing techniques were designed to ensure a low reconstruction error of the observed spectrum while keeping the number of active spectra in the solution as low as possible. As opposed to MESMA, sparse unmixing algorithms, of which *Sparse Unmixing via variable Splitting and Augmented Lagrangian* (SUnSAL) [3] is an example, take full advantage of the benefits of convex optimization, such that they are able to cope with tens of millions of observed signals, as it is the case in hyperspectral applications. On the other hand, the number of active endmembers can not be imposed explicitly.

It is easy to infer that both aforementioned methods suffer from the large number of spectra contained in the spectral libraries: MESMA because of the extremely large number of possible combinations that should be tested, sparse unmixing from the high mutual coherence of the real spectral libraries (*i.e.*, the largest cosine between any two columns of the matrix), which was shown to strongly affect the performance of this type of algorithms. Thus, the need

---

This work has been supported by the European Community's Marie Curie Research Training Networks Programme under contract MRTN-CT-2006-035927, Hyperspectral Imaging Network (HYPER-I-NET). Funding from the Portuguese Science and Technology Foundation, project PEst-OE/EEI/LA0008/2011, and from the Spanish Ministry of Science and Innovation (CEOS-SPAIN project, reference AYA2011-29334-C02-02) is also gratefully acknowledged.

to reduce the size of the spectral library used in unmixing becomes critical.

Based on the fact that the observed data lives in a lower-dimensional subspace, this paper presents a method to actively select a set of spectra from the spectral library able to describe the observed vectors with high accuracy. Both MESMA and SUnSAL should benefit from such a library pruning, which is expected to improve the accuracy and/or the running time of these algorithms. A short review of MESMA and SUnSAL algorithms is given in Section 2. The library pruning method is described in Section 3. Experiments in a simulated dataset are shown in Section 4. The last section of the paper is dedicated to concluding remarks and hints to the future work.

## 2. COMBINATORIAL VS. SPARSE TECHNIQUES

In this section, we review MESMA and SUnSAL algorithms. Let  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$  be the hyperspectral image composed by  $n$  observed  $L$ -dimensional spectral vectors representing the reflectance values acquired by a hyperspectral sensor in different spectral bands. We denote by  $\mathbf{A}$  the  $L \times m$  matrix representing the spectral library, where  $L$  is the number of bands and  $m$  is the number of pure spectra. The linear mixing model for one pixel can be written as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (1)$$

where  $\mathbf{x}$  is the vector of fractional abundances and  $\mathbf{n}$  is a vector collecting the errors that affect the measurement process (*e.g.*, noise). The fractional abundances are subject to the so-called *non-negativity constraint* (ANC) and the *sum-to-one constraint* (ASC) (the abundances cannot be negative and they should sum to one) [4].

### 2.1. MESMA based unmixing

MESMA exploits the heterogeneous composition of a given library  $\mathbf{A}$ . For each pixel, it randomly selects combinations of spectra belonging to different classes and it solves a fully constrained least-squares optimization problem (*i.e.*, using the selected spectra, it minimizes the reconstruction error, while the ANC and ASC are satisfied). The unmixing solution is given by the combination which leads to the lowest reconstruction error of the observed pixel. It is well-known that, when the number of classes in the library is large, MESMA is infeasible due to the enormous number of combinations to be tested. Even if the number of classes is low, the number of tries (combinations) is limited due to running time constraints. Thus, on the one hand, MESMA benefits from the use of large libraries as the possibility to encounter a combination modeling with accuracy the real spectrum of the observed pixel increases. On the other hand, the increase in the number of possible tries dramatically limits the abilities of MESMA, because of the limited number of combinations that should be tested in a reasonable amount of time.

### 2.2. Sparse unmixing - the SUnSAL algorithm

Sparse unmixing exploits the fact that the mixed pixels contain usually a low number of endmembers compared to the number of spectra in the spectral library. The unmixing is formulated as a sparse regression problem, expressed as the minimization of a convex objective function. A way of writing the sparse unmixing as a convex optimization problem is the following:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \\ \text{subject to:} \quad & \mathbf{x} \geq 0, \end{aligned} \quad (2)$$

where the first term is the data-fitting term and the second term is the  $\ell_1$ -norm of  $\mathbf{x}$  which imposes the sparsity, while  $\lambda$  is a regularization parameter which weights the two terms of the objective function. The algorithm that we use to solve the optimization problem (2) is the SUnSAL algorithm, which is a fast algorithm tailored to hyperspectral applications (in which millions or tens of millions of vectors are observed) and based on the Alternating Direction Method of Multipliers [5].

As opposed to MESMA, SUnSAL uses the full spectral library  $\mathbf{A}$  in the unmixing process, which limits its use when the number of spectra in  $\mathbf{A}$  is in the order of tens of thousands. However, for libraries containing hundreds of spectra, SUnSAL is faster and more accurate than methods which do not impose sparsity explicitly [1]. The SUnSAL performance is limited mainly by the high coherence of the real spectral libraries [6] [4], which in most applications approaches the value of one, as the number of spectra  $m$  is usually larger than the number of bands  $L$  (fat matrices). Thus, SUnSAL performances can be enhanced by using as input a reduced set of selected spectra, suitable for the specific observed data.

Signal unmixing focuses on finding an accurate representation, in each pixel, of the spectrum generated by some class of interest in the respective pixel (*e.g.*, vegetation). We consider the case when each spectrum from the library (also called *dictionary*) represents a possible state of the vegetation on the ground, related to different parameters (water content, LAI, chlorophyll content). Obviously, both SUnSAL and MESMA are able to solve the signal unmixing problem, subject to the aforementioned drawbacks. In the following section, we describe a dictionary pruning method intended to boost the performances of the two algorithms in signal unmixing.

## 3. DICTIONARY PRUNING FOR SIGNAL UNMIXING

The method we use for dictionary pruning is related to the CS-MUSIC algorithm detailed in recent works [7, 8], which addresses the so-called multiple measurement vector (MMV) problem. However, the application in hyperspectral unmixing is different from the original one, as the number of observations is much larger than the number of channels. The pruning method is, then, focused more on an accurate estimation of the subspace underlying the data instead of solving the issue of lack of sufficient data points. The method and the details behind are presented in [9, 10]. Here, we resume the main steps of the method: 1) subspace estimation; this step estimates the subspace in which the data lives using a dedicated subspace estimator (*e.g.*, the *Hyperspectral Signal Subspace Identification by Minimum Error* (HySime) [11]); theoretically, the endmembers should be included in this subspace; 2) projection of the library members onto the estimated subspace; it is an intermediate step; 3) computation of projection errors; this step measures the distance from each library member to the data subspace using the standard orthogonal projection operator; 4) assembling the new spectral library; this is done by retaining from the library the closest  $r$  members to the subspace. Once the pruning is done, the unmixing runs using the pruned library as input.

## 4. RESULTS IN A SIMULATED ENVIRONMENT

This section shows the performance of the proposed signal unmixing method with pruned libraries in a simulated dataset.

#### 4.1. Simulated datasets and performance discriminator

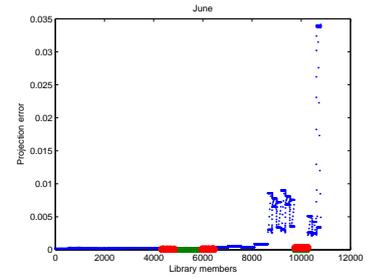
A time series of four simulated hyperspectral satellite images ( $18 \times 120$  pixels, 350–2500 nm, spectral resolution of 10 nm, spatial resolution of 30 m) were generated from a ray-tracing experiment in a fully calibrated virtual citrus orchard [12]. All images were simulated using an extended version of the Physically Based Ray-Tracer (PBRT [13]). The trees were constructed as triangular meshes by implementing the algorithm developed by Weber and Penn [14] and their material properties were described by a bidirectional scattering distribution function (BSDF) model [12]. Leaf spectral measurements (ASD, Boulder, CO) of 30 selected trees collected in the field in January, June, August and September were used to calibrate temporal spectral changes in the virtual citrus orchard. Temporal changes in the structure of the orchard were not modeled.

An additional set of 8640 in situ measured tree canopy and 2160 soil spectra were available. This dataset is a compilation of spectral measurements of the 30 selected trees (and surrounding soil patches) taken on a monthly basis during a two year period, thus incorporating the seasonal changes in phenology and incorporating most of the possible endmember states (dry-wet soil, healthy-stressed vegetation). This set of 10800 spectra was used as input spectral library in the experiments, with three goals: 1) analyzing if the pruning methodology is able to select correct spectra; 2) see if the proposed signal unmixing with the pruned library improve the ones obtained with the full library; 3) analyze the potential of sparse unmixing in the signal unmixing problem.

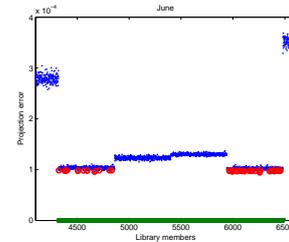
In our experiments, MESMA is applied both for the full and pruned libraries, while SUnSAL is applied for the pruned library only. Then, the vegetation spectrum in each pixel of the considered image is reconstructed based on the inferred fractional abundances of the endmembers. The performance discriminator we use is the average spectral angle deviation, which is an average of the spectral angle computed between the true and reconstructed spectra of vegetation over all pixels. The results shown for MESMA are average results over five runs. The subspace estimator we use in the experiments is HySime<sup>1</sup>, which returns a set of eigenvectors for the observed image and the estimated number of endmembers in the image  $k$ . For the subsequent processing, we retained  $k + 20$  eigenvectors to describe the data subspace. For each image, the number of retained spectra was set to 50, both for tree and soil, resulting in a pruned library with 100 spectra (approximately 100 times less than the full library).

#### 4.2. Performance assesment

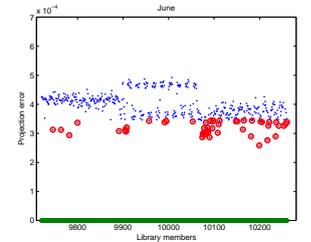
Firstly, we analyze the ability of the proposed pruning scheme to select appropriate signatures from the original library, in order to model the observed data. The experiments showed that, for every of the four scenes, correct endmembers were retained by the pruning method. For illustrative purposes, Fig. 1.(a) shows the projection errors of the library atoms onto the subspace inferred by HySime, in the simulated dataset for June (blue: projection errors corresponding to each library member; green: library members corresponding to June; red: projection errors of the retained members). Figs. 1.(b)–(c) show a zoomed version of the same plot for the signatures corresponding to the June dataset (tree and soil signatures, respectively). Note that all the retained members correspond to actual endmembers, which is also the case for the other three datasets. We conclude that the pruning method is able to select appropriate members from the large library. We analyze now the impact of the pruning methodology over the signal unmixing accuracy. The algorithms described



(a) Projection errors – full library



(b) Projection errors – trees



(c) Projection errors – soils

**Fig. 1.** Projection errors and selected members for the June dataset

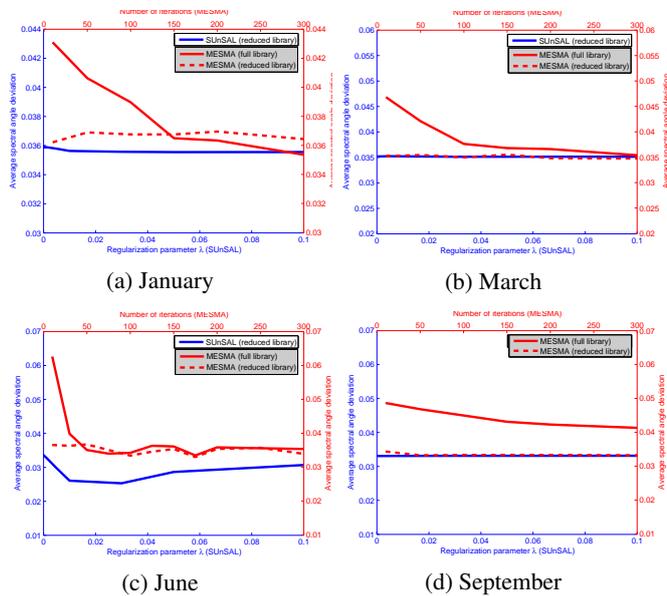
in Section 2 are applied subsequently to the four datasets, according to the methodology described in Subsection 4.1. SUnSAL is tested for different values of the regularization parameter  $\lambda$  ranging from  $\lambda = 0$  (which produces the classical NCLS solution) to  $\lambda = 0.1$ . The number of iterations set for MESMA varies between 10 and 300. Fig. 2 shows the average spectral angle deviation obtained by the considered methods for all considered simulation conditions (red: MESMA, blue: SUnSAL). It can be seen in Fig. 2 that the accuracy of MESMA running with the full library increases with the number of iterations, as expected. However, the MESMA solution becomes much more predictable when the pruning methodology is applied (the performance does not vary significantly with the number of iterations), which means that MESMA benefits from pruning in terms of running time, as the number of iterations can be greatly decreased without loss of accuracy. On the one hand, the performances of MESMA using full and reduced library approach similar values with the increase of the number of iterations, but the pruning methodology ensures increased confidence in the results obtained for lower number of iterations. On the other hand, SUnSAL shows the best performance in terms of average spectral angle deviation, while not being strongly dependent on the regularization parameter  $\lambda$  (a small value of  $\lambda$  ensures a good performance). As it is already known that sparse unmixing was designed as a fast technique, we show in Table 1 a comparison of the running times for SUnSAL and MESMA (for a selection of values corresponding to  $\lambda$  and to the number of iterations  $n_i$ ). The experiments were done using Matlab7 on a computer equipped with an Intel Core Duo@2.5GHz processor and 4GB of RAM memory. Note that the execution time of MESMA depends strongly on the number of iterations, while SUnSAL is practically independent w.r.t.  $\lambda$  in this respect. Obviously, SUnSAL outperforms MESMA in terms of running time.

To summarize this Section, we make a few observations. On the one hand, the tested methodology for library pruning is able to retain appropriate members from the library to explain the observed data. MESMA benefits from the pruning, as it is able to attain near-

<sup>1</sup>Available online at: <http://www.lx.it.pt/biucas/code.htm>

**Table 1.** Running times of the considered algorithms for several parameter settings. MESMA times are average times over five runs

Algorithm	MESMA (full library)			MESMA (pruned library)			SUnSAL (pruned library)		
Parameter	$n_i = 10$	$n_i = 100$	$n_i = 200$	$n_i = 10$	$n_i = 100$	$n_i = 200$	$\lambda = 10^{-5}$	$\lambda = 10^{-3}$	$\lambda = 10^{-1}$
TIME [s]	307.7	$2.84 \cdot 10^3$	$5.61 \cdot 10^3$	302.2	$2.84 \cdot 10^3$	$5.62 \cdot 10^3$	1.96	1.98	1.9



**Fig. 2.** Spectral angle deviation achieved by the considered algorithms in the four datasets

optimal performances for a lower number of iterations. On the other hand, the sparse unmixing technique is much faster and it obtains better accuracies than MESMA when the pruned library is used, which shows its potential as a reliable signal unmixing technique.

## 5. CONCLUSIONS

In this paper, a methodology for library pruning in hyperspectral signal unmixing was tested. The results obtained in a simulated environment showed the ability of the method to correctly select members of the original spectral library to describe the observed data. After pruning, both MESMA, which is a combinatorial method, and SUnSAL, a sparse unmixing technique, showed improved performance compared to the results obtained with MESMA using the full spectral library. The results indicate that sparse unmixing has the potential to become a powerful signal unmixing technique for the considered type of applications, due to the accuracy obtained and to the low running time needed. Future work will include tests with more heterogeneous real scenes, the exploitation of structured dictionaries in signal unmixing and parallel implementation of the algorithms.

## 6. REFERENCES

- [1] D. Iordache, J. Bioucas-Dias, and A. Plaza, “Sparse unmixing of hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, 2011.
- [2] L. Tits, B. Somers, and P. Coppin, “The potential and limitations of a clustering approach for the improved efficiency of multiple endmember spectral mixture analysis in plant production system monitoring,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, pp. 2273–2286, 2012.
- [3] J. Bioucas-Dias and M. Figueiredo, “Alternating direction algorithms for constrained sparse regression: application to hyperspectral unmixing,” *2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHIS-PERS)*, 2010.
- [4] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2011.
- [5] J. Eckstein and D. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Mathematical Programming*, vol. 5, pp. 293–318, 1992.
- [6] A. Bruckstein, D. Donoho, and M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” *SIAM Review*, vol. 51, pp. 34–81, 2009.
- [7] J.M. Kim, O.K. Lee, and J.C. Ye, “Compressive music with optimized partial support for joint sparse recovery,” *arXiv:1102.3288v2*, June 2011.
- [8] J.M. Kim, O.K. Lee, and J.C. Ye, “Compressive music: a missing link between compressive sensing and array signal processing,” *arXiv:1004.4398v5*, June 2011.
- [9] D. Iordache, J. Bioucas-Dias, A. Plaza, and B. Somers, “Music-csr: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression,” *IEEE Transactions on Geoscience and Remote Sensing (submitted)*, 2013.
- [10] D. Iordache, J. Bioucas-Dias, and A. Plaza, “Dictionary pruning in sparse unmixing of hyperspectral data,” *4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPER)*, Shanghai, China, 2012.
- [11] J. Bioucas-Dias and J. Nascimento, “Hyperspectral subspace identification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.
- [12] J. Stuckens, B. Somers, W.W. Verstraeten, and P. Coppin, “The impact of common assumptions on canopy radiative transfer simulations: A case study in citrus orchards,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 110, no. 1/2, pp. 1–21, January 2009.
- [13] M. Pharr and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.
- [14] J. Weber and J. Penn, “Creation and rendering of realistic trees,” in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1995, SIGGRAPH ’95, pp. 119–128, ACM.