EPSC Abstracts
Vol. 6, EPSC-DPS2011-1266, 2011
EPSC-DPS Joint Meeting 2011

Efficient Extraction and Classification of Spectra

Albrecht Schmidt (1), Frédéric Schmidt (2), Erwan Tréguier (1), Saïd Moussaoui (3)
(1) European Space Agency, ESAC, Spain, (2) IDES, UMR 8148, Université Paris Sud, CNRS, Orsay, France (3) IRCCYN, UMR CNRS 6597, Ecole Centrale Nantes, France

1. Abstract

Imaging spectrometers deliver very large amounts are data which call for automatic summarisation for exploratory data analysis. In the frequent absence of ground truth for planetary data, unsupervised analysis methods can provide unbiased information about the data. In this work, we investigate the use of unsupervised analysis based on non-negative matrix approximation [3, 4] combined with subsequent classification [2] to provide scientists with succinct summaries. Since typically there often is no ground truth to compare to, unsupervised rather than supervised methods allow to extract new information from data sets. We designed particularly efficient methods to cope with the large data volumes which are typical for this type of instrument.

2. Methods

The first step consists of hyperspectral unmixing, that is, extracting source spectra which make up the compound spectrum measured by the instrument [1] (see Figures 1 and 2). By considering P pixels of an hyperspectral image acquired at L frequency bands, the observed spectra are gathered in a $P \times L$ data matrix $\mathbf{X}$. Each row of this matrix contains a measured spectrum at a pixel with spatial index $p = 1, \ldots, P$. According to the linear mixing model, the p^{th} spectrum, $1 \leq p \leq P$, can be expressed as a linear combination of $r_i, 1 \leq r_i \leq R$, pure spectra of the surface components. Using matrix notations, this linear spectral mixing model can be written as

$$\boldsymbol{X} \approx \boldsymbol{A} \cdot \boldsymbol{S}, \tag{1}$$

where non-negative matrices $A \in M^{P \times R}$, and $S \in M^{R \times L}$ approximate $X \in M^{P \times L}$ in the sense that $\frac{1}{2}||AS - X||^2$ is minimised, where $M^{\cdot \times \cdot}$ is the space of matrices of respective dimensions with non-negative entries. The rows of matrix $\boldsymbol{S}$ now contain the pure surface spectra of the R components, and each element A_{pr} of matrix $\boldsymbol{A}$ corresponds to the abundance of the r^{th} component in pixel with spatial index p. For a qualitative and quantitative description of the observed scene composition, the estimation problem consists of finding matrices $\boldsymbol{S}$ and $\boldsymbol{A}$ which allow to explain the data matrix and, at the same time, have a coherent physical interpretation.

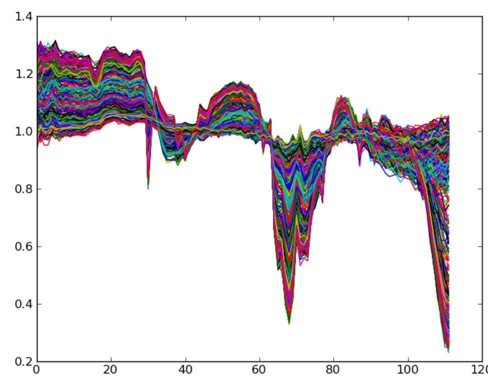

Figure 1: Overview of compound spectral shapes.

The second step consists of classifying the sources that were extracted in the previous step. For summarising the information contained in a set of spectral images, we look at a variety of unsupervised methods to see for which kind of data sets they deliver interesting results.

3. Summarisation

After the second step the classifications are looked at and a summary is created. Currently, this is done by selecting an archetype [2] for each class, such as the first spectrum in Figure 2; the instance selected as archetype typically depends on the classification method used.

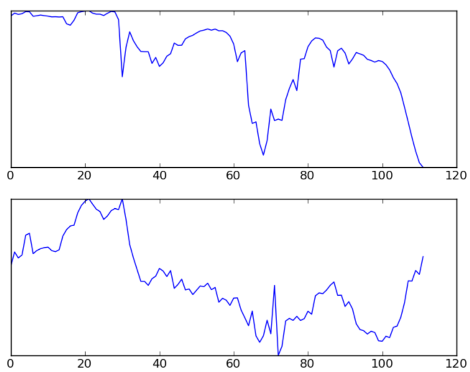

Figure 2: Extracted individual spectra.

References

[1] Cichocki, A., Amari, S.-I.: Adaptive blind signal and image processing – Learning algorithms and applications, Wiley Interscience, 2002

[2] Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009

[3] Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization, Advances in neural information processing systems, 2001

[4] Schmidt, F., Schmidt, A., Tréguier, E., Guiheneuf, M., Moussaoui, S., Dobigeon, N.: Implementation strategies for hyperspectral unmixing using Bayesian source separation, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4003–4013, Nov. 2010