Geophysical Research Abstracts
Vol. 14, EGU2012-12524, 2012
EGU General Assembly 2012

Compressing Data by Source Separation

A. Schmidt (1), E. Tréguier (1), F. Schmidt (2), and S. Moussaoui (3)

(1) European Space Agency, SRE-ODR, Villanueva de la Cañada, Spain (albrecht.schmidt@esa.int), (2) IDES, Université Paris-Sud, Orsay, France (frederic.schmidt@u-psud.fr), (3) Ecole Centrale de Nantes, IRCCyN, Nantes, France (said.moussaoui@irccyn.ec-nantes.fr)

We interpret source separation of hyperspectral data as a way of applying lossy compressing. In settings where datacubes can be interpreted as a linear combination of source spectra and their abundances and the number of sources is small, we try to quantify the trade-offs and the benefits of source separation and its implementation with non-negative source factorisation. While various methods to implement non-negative matrix factorisation have been used successfully for factoring hyperspectral images into physically meaningful sources which linearly combine to an approximation of the original image. This is useful for modelling the processes which make up the image. At the same time, the approximation opens up the potential for a significant reduction of the data by keeping only the sources and their corresponding abundances, instead of the original complete data cube. This presentation will try to explore the potential of the idea and also to establish limits of its use.

Formally, the setting is as follows: we consider P pixels of a hyperspectral image which are acquired at L frequency bands and which are represented as a PxL data matrix X. Each row of this matrix represents a spectrum at a pixel with spatial index p=1..P; this implies that the original topology may be disregarded. Since we work under the assumption of linear mixing, the p-th spectrum, 1<=p<=P, can be expressed as a linear combination of r, 1<=r<=R, source spectra. Thus, X=AxS+E, E being an error matrix to be minimised, and X, A, and S only have non-negative entries. The rows of matrix S are the estimations of the R source spectra, and each entry of A expresses the contribution of the r-th component to the pixel with spatial index p.

There are applications where we may interpret the rows of S as physical sources which can be combined using the columns of A to approximate the original data. If the source signals are few and strong (but not even necessarily meaningful), the data volume that has to be transmitted or looked at can be reduced significantly.