

# Learning mid-IR emission spectra of PAH populations from observations

Sacha Foschino<sup>1</sup>, Olivier Berné and Christine Joblin

IRAP, universit  de Toulouse, CNRS, CNES, UPS  
email: [sacha.foschino@irap.omp.eu](mailto:sacha.foschino@irap.omp.eu)

**Abstract.** Observations of the mid-infrared (mid-IR, 3–15  $\mu\text{m}$ ) spectra of photo-dissociation regions reveal ubiquitous, broad and intense emission bands, the aromatic infrared bands (AIBs), attributed to polycyclic aromatic hydrocarbons (PAHs). Studies of the AIBs showed spectral variations (e.g. in the band positions) between different astrophysical objects, or even within single object, thanks to hyperspectral images. The James Webb Space Telescope (JWST) will allow to get further spectral and spatial details compared to former space observatories. This will come with large data sets, which will require specific tools in order to perform efficient scientific analysis.

We propose in this study a method based on blind signal separation to reduce the analysis of such large data set to that of a small number of elementary spectra, spectrally representative of the data set and physically interpretable as the spectra of populations of mid-IR emitters. The robustness and fastness of the method are improved compared to former algorithms. It is tested on a ISO-SWS data set, which approaches the best the characteristics of JWST data, from which four elementary spectra are extracted, attributed to cationic, neutral PAHs, evaporating very small grains and large and ionized PAHs.

**Keywords.** ISM: molecules - ISM: lines and bands - Infrared: ISM

---

## 1. Challenges of JWST large data sets in the study of AIB

The observation of the Galaxy in the mid-IR region (3–15  $\mu\text{m}$ ) revealed broad and intense spectral structures in the spectrum of photodissociation regions at 3.3, 6.2, 7.7, 8.6, 11.2 and 12.8  $\mu\text{m}$ , the so called aromatic infrared bands (AIBs), attributed by [Leger & Puget \(1984\)](#) and [Allamandola \*et al.\* \(1985\)](#) to IR emission of polycyclic aromatic hydrocarbons (PAHs). A particularity of these spectral structures is their variabilities. Indeed, from one astrophysical object to another or even among a same object when a hyperspectral image is available, one can notice variations in the peak wavelength, width of AIBs or even in the flux ratio between bands.

These variations can be used to characterize the AIB carriers and to probe the physical and chemical conditions of the emitting medium. To analyze these variations one can focus on individual bands or use blind signal separation (BSS) approach. Former studies (e.g. [Rapacioli \*et al.\* \(2005\)](#) and [Bern  \*et al.\* \(2007\)](#)) used this approach to rationalize AIB variations as the linear combination of a certain number of elementary spectra. These elementary spectra can be interpreted as the spectra of populations of cationic and neutral PAHs and “evaporating very small grains” ([Pilleri \*et al.\* \(2012\)](#)). This can help propose a scenario of evolution of PAHs under UV radiation.

In these studies, BSS has been used on data sets of small sizes (data matrix of  $\sim 10\text{k}$  points) compared to those that will be provided by the JWST. Indeed, around 900 spectra of 40k spectral points each (i.e.  $\sim 40 \times 10^6$  points) will be provided by one field of view of its NIRSpec and MIRI integral field unit instruments combined data. Moreover,

the time cost of this kind of approach increases with the number of points of the data set. For the analysis of AIBs, the most suitable BSS method is called non-negative matrix factorization (NMF), used by Berné *et al.* (2007). But because of its iterative nature, NMF algorithms have to be initialized, in general randomly, resulting in different solutions (i.e. set of elementary spectra) at each run on the same data set. This implies the use of a Monte Carlo post-processing, increasing the time cost of this method. Moreover, former studies did not deal with contamination by gas lines and continuum. NMF being a linear method, extinction, gas lines or continuum emissions cannot be taken into account and have to be treated before running NMF on the data set in order to have a set of elementary spectra containing only AIBs and no other kind of emission.

In this study, we propose a method solving these issues and improving the robustness and the fastness of this kind of approach on large data sets.

## 2. Hybrid MASS-NMF method: a fast and robust approach

JWST will observe a variety of environments, a large majority presenting gas lines, continuum emission and AIBs. In order to apply BSS on AIBs only, we consider an observed spectrum to be the sum of each of these contributions. We then use a linear (i.e. fast) fitting algorithm (e.g. non negative least square algorithm) based on a list of Gaussians to model gas lines, a list of blackbodies affected by different extinctions and a list of Gaussians to model the AIBs. With this model, one can easily extract the contribution of AIBs and build a data matrix containing only pure AIB spectra.

Boulais *et al.* (2015) showed, based on synthetic data (i.e. ground truth) that the initialization of the NMF algorithm using the output of the geometrical method they developed, called MASS (for Maximum angle source separation), provides a solution as robust as the solution resulting from a Monte Carlo analysis. Thus, only one application of MASS and one application of the NMF on the data set provide a satisfying solution, dividing the time cost necessary by a factor of  $\sim 100$ . We combine MASS and NMF applied on the pure AIB matrix to constitute the new method.

As a test of this method, we used an ISO-SWS data set of 31 spectra from which we extracted 4 elementary spectra. The first three spectra are in good agreement with former studies. The fourth spectrum is attributed to a population initially invoked by Joblin *et al.* (2008) to account for the particular spectra of planetary nebulae. The output of the test shows promising results when it will be applied on large JWST data sets. The total method is performed on a  $31 \times 6799$  data matrix in 10 minutes on a laptop. We estimate that on JWST data set, the first step of AIB extraction will take 1 min 30 sec per spectrum and the run of MASS-NMF itself 10 minutes.

## Acknowledgment

We acknowledge support from the European Research Council FP7 ERC-2013-SyG NANOCOSMOS (G.A. no. 610256).

## References

- Allamandola, L. J., Tielens, A. G. G. M., & Barker, J. R. 1985, *A&A*, 290, L25
- Berné, O., Joblin, C., Deville, Y., Smith, J. D., Rapacioli, M., Bernard, J. P., Thomas, J., Reach, W., & Abergel, A. 2007, *A&A*, 469, 575
- Boulais, A., Deville, Y., & Berné, O. 2015, *2015 IEEE International Workshop of ECMSM*, 1
- Joblin, C., Szczerba, R., Berné, O., & Szyszka, C. 2008, *A&A*, 490, 189
- Leger, A. & Puget, J. L. 1984, *A&A*, 137, L5
- Pillari, P., Montillaud, J., Berné, O., & Joblin, C. 2012, *A&A*, 542, A69
- Rapacioli, M., Joblin, C., & Boissel, P. 2005, *A&A*, 429, 193