

# Identification des spectres infrarouges des populations de poussière carbonées interstellaires par séparation aveugle de sources

Olivier BERNÉ<sup>1,2</sup>, Yannick DEVILLE<sup>2</sup>, Christine JOBLIN<sup>1</sup>

<sup>1</sup>Centre d'Etude Spatiale des Rayonnements, CNRS et Université Paul Sabatier Toulouse 3,  
Observatoire Midi-Pyrénées, 9 Av. du Colonel Roche, 31028 Toulouse

<sup>2</sup>Laboratoire d'Astrophysique de Toulouse-Tarbes, CNRS et Université Paul Sabatier Toulouse 3,  
Observatoire Midi-Pyrénées, 14 Av. Edouard Belin, 31400 Toulouse  
berne@cesr.fr, yannick.deville@ast.obs-mip.fr,  
joblin@cesr.fr

**Résumé** – L'utilisation de méthodes de séparation de sources pour l'astronomie reste un domaine à explorer. Nous avons appliqué deux algorithmes de séparation de sources (i.e. Non-Negative Matrix Factorization et *FastICA*) à des données de spectro-imagerie infrarouge du milieu interstellaire obtenues avec le télescope spatial *Spitzer*. Ces méthodes nous ont permis d'extraire les spectres infrarouges propres à trois populations de nanoparticules carbonées interstellaires distinctes. Ces spectres peuvent ensuite être utilisés pour interpréter l'émission dans l'infrarouge moyen de la poussière dans d'autres galaxies proches.

**Abstract** – The use of Blind Signal Separation methods (ICA and other approaches) for the analysis of astrophysical data remains quite unexplored. In this paper, we present a new approach for analyzing the infrared emission spectra of interstellar dust, obtained with NASA's Spitzer Space Telescope, using *FastICA* and Non-negative Matrix Factorization (NMF). Using these two methods, we were able to unveil the *source* spectra of three different types of carbonaceous nanoparticles present in interstellar space. These spectra can then constitute a basis for the interpretation of the mid-infrared emission spectra of interstellar dust in the Milky Way and nearby galaxies.

## 1 Introduction

Dans la Voie Lactée, les étoiles sont espacées entre elles de plusieurs années-lumière. Cet espace, que l'on nomme *Milieu Interstellaire*, est majoritairement vide mais on y trouve également de grands nuages de gaz et de poussière. Bien qu'ils ne représentent que 10% de la masse de notre galaxie, ces nuages jouent un rôle fondamental dans son évolution physique et chimique. Lorsqu'ils sont illuminés par le rayonnement UV des étoiles, les grains de poussière ré-émettent dans l'infrarouge (IR) l'énergie qu'ils ont absorbée en UV. Un excellent moyen de sonder la nature de ces grains est par conséquent la spectroscopie IR. Cependant, les spectres de ces nuages sont généralement constitués d'un mélange d'émissions dues à plusieurs populations de poussière. En appliquant des méthodes de séparation aveugle de sources (*FastICA* et *NMF*) à des données de spectro-imagerie infrarouge obtenues avec le télescope spatial *Spitzer*, nous montrons qu'il est possible d'extraire le spectre de chacune de ces populations. L'identification de ces spectres permet de mieux cerner la nature et l'évolution des grains interstellaires dans notre galaxie. Nous montrons également qu'il est possible d'interpréter les spectres observés de galaxies proches à l'aide des spectres élémentaires que nous avons extraits dans notre galaxie.

## 2 Observations

Nous avons observé des régions de Photodissociation (PDR), dans lesquelles les photons UV d'une étoile illuminent un nuage interstellaire. Dans ces régions, l'intensité du rayonnement reçu par les grains diminue avec la distance à l'étoile, et par conséquent la forme du spectre d'émission infrarouge de la poussière varie de manière significative suivant la position dans le nuage. Afin de profiter du maximum d'information sur l'émission des grains dans ces régions, nous en avons observé une dizaine en spectro-imagerie infrarouge avec le télescope spatial *Spitzer* de la NASA [2]. Pour chaque position sur une image de PDR, nous disposons d'un spectre infrarouge entre 5 et 35  $\mu\text{m}$  (voir Fig. 1). Les spectres comportent de larges bandes, dues à l'émission de grandes molécules aromatiques, et un continuum dû à la présence de nanograins dont la nature est à ce jour méconnue. Mathématiquement, ces données peuvent être vues comme des matrices 3D  $C(p_x, p_y, \lambda)$ , définissant les variations de l'émission infrarouge en fonction de la longueur d'onde  $\lambda$ , pour une position sur le ciel donnée  $(p_x, p_y)$ . Les dimensions typiques de ces observations sont de  $30 \times 30 \times 250$  points.

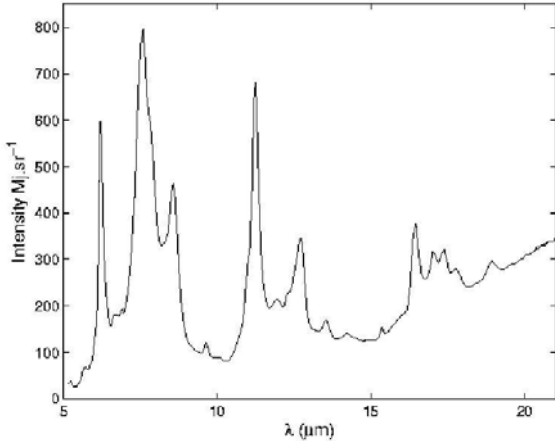
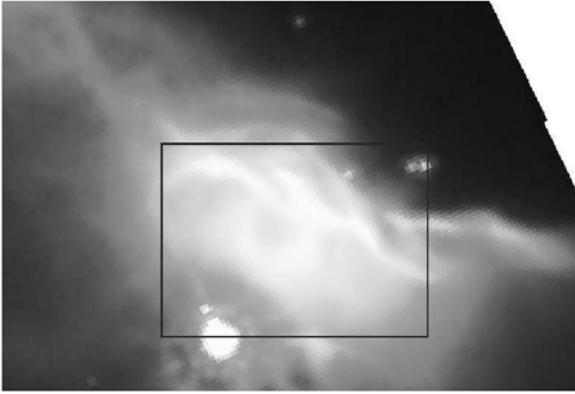


FIG. 1 – *En haut* : Vue d’ensemble à  $8\mu\text{m}$  d’une région de Photodissociation dans la nébuleuse de l’Iris (NGC 7023). L’étoile illumine le nuage qui se trouve en haut. La région encadrée montre la zone observée en spectro-imagerie. *En bas* : Un des spectres IR observés dans la région encadrée. On distingue les larges bandes d’émission dues aux molécules aromatiques.

### 3 Analyse par séparation aveugle de sources

Les matrices 3D (cubes) définies ci-dessus contiennent un ensemble de spectres, qui sont le résultat de combinaisons linéaires de spectres *sources* propres à chacune des population de grains présentes dans le nuage. On se trouve par conséquent en présence d’un cas typique de problème de séparation aveugle de sources (SAS), dans lequel on cherche à identifier les spectres *sources* inconnus, à partir des observations qui sont des combinaisons linéaires aux coefficients inconnus de ces *sources*. Ce mélange linéaire instantané s’écrit mathématiquement :

$$X = AS \quad (1)$$

où  $X$  est la matrice des signaux observés,  $A$  la matrice de mélange, et  $S$  la matrice contenant les signaux *sources*. Dans notre cas, la matrice de signaux observés est 3D. Pour se retrouver dans le cas 2D défini en Eq. (1), on place l’ensemble des spectres d’un cube dans les lignes de la matrice d’observations  $X$ , faisant ainsi disparaître la structure spatiale des données. Nous avons utilisé deux mé-

thodes totalement différentes afin d’identifier les spectres de la matrice  $S$  : *FastICA* et la *Non-negative Matrix Factorization* (NMF). *FastICA* est un algorithme très répandu, qui a déjà été utilisé pour des applications astrophysiques (par ex. [5]). Cette méthode statistique est basée sur l’indépendance et la non-gaussianité des signaux *sources*. Nous avons également appliqué une méthode de NMF à ces données, ce qui constitue à notre connaissance une première en astronomie. Contrairement à *FastICA* qui utilise l’indépendance des signaux *sources*, la NMF est uniquement basée sur la non-négativité de ces signaux et des coefficients avec lesquels ces signaux sont combinés. La matrice des observations  $X$  est approximée par le produit :

$$X \approx WH \quad (2)$$

où  $W$  est une matrice non-négative des poids et  $H$  la matrice positive des *sources* estimées. La qualité de l’approximation réalisée dans l’Eq. (2) peut être optimisée en minimisant la divergence  $D(X|WH)$  définie par :

$$D(X|WH) = \sum_{ij} (X_{ij} \log \frac{X_{ij}}{(WH)_{ij}} - X_{ij} + (WH)_{ij}). \quad (3)$$

L’ajustement de  $W$  et  $H$  peut être réalisé par une descente en gradient mais cela comporte un certain nombre de contraintes comme le montre [4]. Par conséquent, Lee et Seung ont développé un nouvel algorithme basé sur une règle d’actualisation itérative des matrices  $W$  et  $H$  :

$$H_{a\mu} \leftarrow H_{a\mu} \frac{\sum_i W_{ia} X_{i\mu} / (WH)_{i\mu}}{\sum_k W_{ka}}, \quad (4)$$

et

$$W_{ia} \leftarrow W_{ia} \frac{\sum_{\mu} H_{a\mu} X_{i\mu} / (WH)_{i\mu}}{\sum_{\nu} H_{a\nu}}. \quad (5)$$

En partant de matrices  $W$  et  $H$  aléatoires, ces règles garantissent la non-croissance de la divergence définie en Eq. (3) et optimisent l’approximation de l’équation (2). [3] a mis en évidence le fait que la NMF peut s’avérer problématique dans certains cas, car elle ne fournit pas de solution unique. Dans les analyses que nous avons réalisées sur 7 cubes de données différents, en changeant jusqu’à 200 fois les conditions initiales (matrices aléatoires  $W$  et  $H$  initiales), la NMF tend toujours vers la même solution. Nous avons d’abord comparé les performances de *FastICA* et de la NMF pour notre application avec des signaux simulés, en présence de bruit. Il apparaît que la NMF permet de mieux restituer les sources artificiellement mélangées en présence de bruit, ce qui nous conduit à la conclusion que la NMF est plus appropriée dans le cas précis de notre application, bien que *FastICA* fournisse des résultats satisfaisants. La partie suivante présente les résultats obtenus sur les données réelles avec la NMF.

### 4 Résultats

Dans sept régions de photodissociation nous avons réussi à extraire les spectres de deux, voire trois populations de poussière grâce à la NMF. Les spectres correspondant aux trois populations identifiées sont représentés en Fig. 2. Nous attribuons [6] ces trois spectres à l’émission de trois

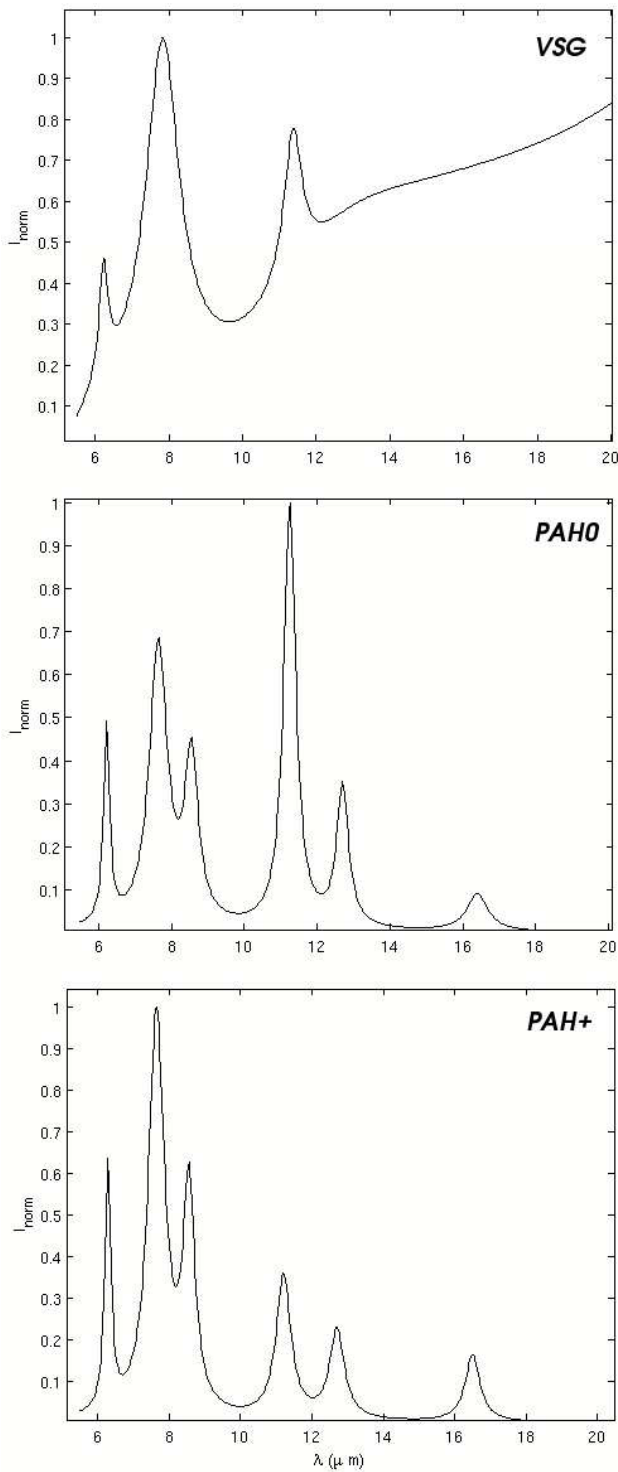


FIG. 2 – Spectres des trois populations de grains extraites par SAS des observations de régions de photodissociation.

grandes familles de nanoparticules carbonées : des très petits grains carbonés (VSG) porteurs du continuum et d'une large bande d'émission à  $7.8 \mu\text{m}$ , des molécules polycycliques aromatiques hydrogénées neutres ( $\text{PAH}^0$ ) dont l'émission est dominée par la bande à  $11.3 \mu\text{m}$ , et des PAH ionisés ( $\text{PAH}^+$ ) qui émettent majoritairement dans les bandes  $7.7$  et  $8.6 \mu\text{m}$ . Il est ensuite possible, pour un

cube de spectro-imagerie IR donné, de calculer en tout point du ciel défini dans ce cube par les coordonnées  $(p_x, p_y)$ , la corrélation  $E[\text{Obs}(p_x, p_y, \lambda)y_p]$ , entre le spectre observé  $\text{Obs}(p_x, p_y, \lambda)$  et le spectre *source* estimé  $y_p$ . Ceci permet donc de réaliser des cartes de corrélation, qui renseignent sur la distribution spatiale des populations de grains dans les PDRs (Fig. 3), ou dans toute autre région observée en spectro-imagerie infrarouge comme des galaxies proches. Les cartes obtenues pour les régions de photodissociation (voir l'exemple de NGC 7023 Fig. 4) montrent que les VSG se situent dans les régions les plus éloignées de l'étoile où le rayonnement UV n'est pas trop intense. En se rapprochant de l'étoile on trouve successivement les  $\text{PAH}^0$  puis les  $\text{PAH}^+$ . Ce résultat indique que les VSGs donnent naissance aux PAHs sous l'effet du rayonnement, et que ces derniers sont ensuite ionisés, toujours en raison de l'effet du rayonnement. D'un point de vue général, ces spectres fournissent des traceurs de l'intensité du champ de rayonnement. Une extension de ce travail consiste donc à utiliser ces spectres pour construire des cartes de corrélation à partir des cubes de spectro-imagerie pour remonter à une information sur les conditions de champ de rayonnement, par exemple à l'échelle de galaxies proches.

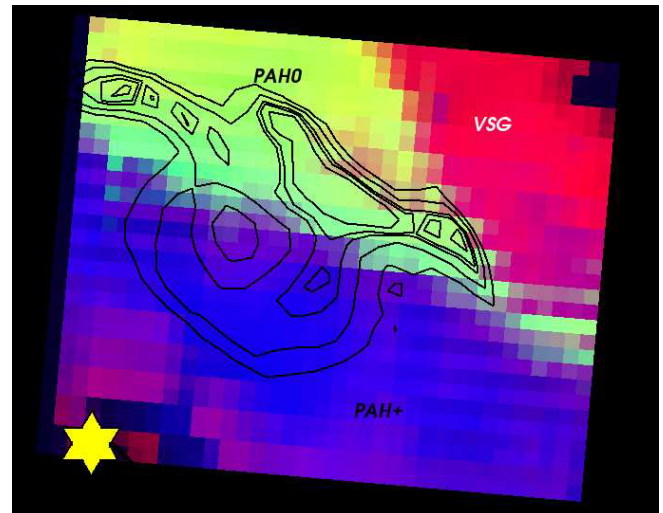


FIG. 3 – Distribution spatiale des population de poussière dans NGC 7023, identifiées par NMF à partir des données de spectro-imagerie. Les contours noirs représentent l'émission correspondant à l'image présentée en Fig. 1.

## 5 Un exemple d'extension : l'analyse de l'émission de la poussière dans la collision de galaxies des Antennes

"Les Antennes", Fig. 4, sont deux galaxies (NGC 4038 et 4039) en interaction. Cet objet comme de nombreux autres a été observé en spectro-imagerie par le télescope spatial Spitzer. À l'aide des spectres que nous avons extraits des observations des nuages de notre galaxie, il est

possible de construire la carte de distribution spatiale des trois types de poussière (VSG, PAH et  $\text{PAH}^+$ , voir Sec. 4) dans ces galaxies. Ces cartes de distribution sont présentées en Fig. 4 avec le même code de couleurs que pour la figure 3. Il apparaît clairement, grâce à cette analyse, que chacune des populations domine l'émission IR-moyen dans une région différente de ces galaxies. On trouve les VSG dans les bras, dans les grands nuages froids et denses tracés par le gaz moléculaire (CO). Les  $\text{PAH}^0$  semblent envelopper les bras alors que les  $\text{PAH}^+$  se situent dans les régions moins denses entre les bras. Il apparaît, en faisant l'analogie avec notre connaissance des régions galactiques, que les VSG sont détruits au bord des bras denses de ces galaxies et donnent naissance aux PAH libres. Dans les régions les plus diffuses, entre les bras, les PAH sont ionisés par le rayonnement.

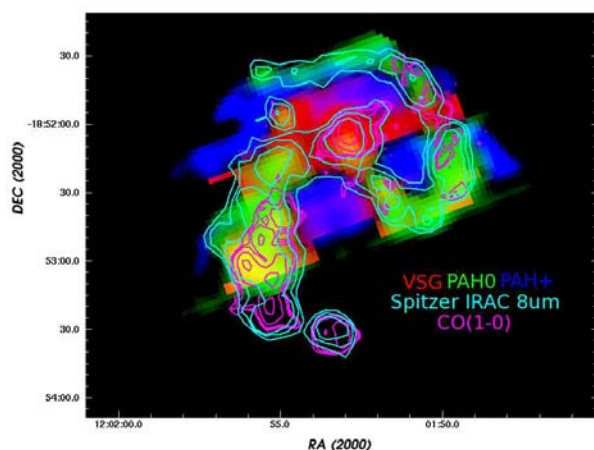
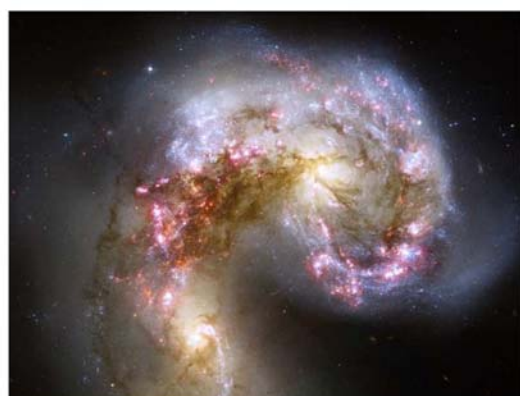


FIG. 4 – *En haut* : Les galaxies des Antennes vues par le Télescope Spatial Hubble. *En bas* : En couleurs, distributions spatiales des différentes composantes de poussière dans ces galaxies (VSG en rouge,  $\text{PAH}^0$  en vert et  $\text{PAH}^+$  en bleu). Les contours blancs indiquent l'émission infrarouge à  $8 \mu\text{m}$ , et les contours magenta l'émission du gaz moléculaire (CO) froid et dense d'après [7].

## 6 Conclusion

Grâce à l'utilisation de méthodes de séparation de sources, nous avons pu extraire les spectres de différents constituants de la poussière interstellaire. L'interprétation astrophysique des spectres extraits permet de mieux comprendre quelle est la nature et l'évolution de la poussière dans notre galaxie. Ces résultats sont essentiels pour comprendre l'évolution des galaxies elles-mêmes.

Nous avons montré ici que ces méthodes peuvent s'avérer particulièrement efficaces pour l'analyse de données astrophysiques. Ceci suggère que ces algorithmes peuvent s'avérer extrêmement utiles à l'avenir dans de nombreuses applications en astronomie, notamment dans le cadre de l'analyse de données de grands instruments (HERSCHEL, PLANCK, MUSE etc.).

## Références

- [1] O. Forni, F. Poulet, J.-P. Bibring et al., 2005, *Component Separation of OMEGA Spectra with ICA*, 36<sup>th</sup> Lunar and Planetary Science Conference, 1623
- [2] C. Joblin, A. Abergel, J.-P. Bernard, et al., 2005, *Very Small Particles and Chemistry in photodissociation regions, from ISO to Spitzer*, International Astronomical Union Symposium, 153
- [3] Moussaoui et al. Neurocomputing à paraître en 2007.
- [4] D. D. Lee & H. S. Seung, 2001, *Algorithms for Non-Negative Matrix Factorization*, NIPS, ed. MIT press, Vol. 13, 556
- [5] D. Nuzillard & A. Bijaoui, 2000, *Blind source separation and analysis of multispectral astronomical images*, Astronomy and Astrophysics Supplement Series, 147, 129
- [6] M. Rapacioli, M., C. Joblin, P. Boissel, 2005, *Spectroscopy of polycyclic aromatic hydrocarbons and very small grains in photodissociation regions*, Astronomy and Astrophysics, 429, 193-204
- [7] C. D. Wilson, N. Scoville, S. C. Madden, V. Charmandaris, 2000, *High-Resolution Imaging of Molecular Gas and Dust in the Antennae (NGC 4038/39) : Super Giant Molecular Complexes*, Astrophysical Journal, 542, 120-127