

Analysis of the emission of very small dust particles from Spitzer spectro-imagery data using blind signal separation methods^{*,**}

O. Berné¹, C. Joblin¹, Y. Deville², J. D. Smith³, M. Rapacioli⁴, J. P. Bernard¹, J. Thomas², W. Reach⁵, and A. Abergel⁶

¹ Centre d'Étude Spatiale des Rayonnements, CNRS et Université Paul Sabatier Toulouse 3, Observatoire Midi-Pyrénées, 9 Av. du Colonel Roche, 31028 Toulouse Cedex 04, France
e-mail: olivier.berne@cesr.fr

² Laboratoire d'Astrophysique de Toulouse-Tarbes, CNRS et Université Paul Sabatier Toulouse 3, Observatoire Midi-Pyrénées, 14 Av. Edouard Belin, 31400 Toulouse, France

³ Steward Observatory, University of Arizona, Tucson, AZ 85721, USA

⁴ Technische Universität Dresden, Institut für Physikalische Chemie und Elektrochemie, 01062 Dresden, Germany

⁵ Spitzer Science Center, MS 220-6, California Institute of Technology, Pasadena, CA 91125, USA

⁶ Institut d'Astrophysique Spatiale, CNRS et Université de Paris Sud, Bât. 109, 91405 Orsay Cedex, France

Received 21 August 2006 / Accepted 6 March 2007

ABSTRACT

Context. This work was conducted as part of the *SPECPDR* program, dedicated to the study of very small particles and astrochemistry, in Photo-Dissociation Regions (PDRs).

Aims. We present the analysis of the mid-IR spectro-imagery observations of Ced 201, NCG 7023 East and North-West and ρ Ophiuchi West filament.

Methods. Using the data from all four modules of the InfraRed Spectrograph onboard the *Spitzer Space Telescope*, we produced a spectral cube ranging from 5 to 35 μm , for each one of the observed PDRs. The resulting cubes were analysed using *Blind Signal Separation* methods (NMF and *FastICA*).

Results. For Ced 201, ρ Ophiuchi West filament and NGC 7023 East, we find that two signals can be extracted from the original data cubes, which are 5 to 35 μm spectra. The main features of the first spectrum are a strong continuum emission at long wavelengths, and a broad 7.8 μm band. On the contrary, the second spectrum exhibits the classical Aromatic Infrared Bands (AIBs) and no continuum. The reconstructed spatial distribution maps show that the latter spectrum is mainly present at the cloud surface, close to the star whereas the first one is located slightly deeper inside the PDR. The study of the spectral energy distribution of Ced 201 up to 100 μm suggests that, in cool PDRs, the 5–25 μm continuum is carried by Very Small Grains (VSGs). The AIB spectra in the observed objects can be interpreted as the contribution of neutral and positively-charged Polycyclic Aromatic Hydrocarbons (PAHs).

Conclusions. We extracted the 5 to 25 μm emission spectrum of VSGs in cool PDRs, these grains being most likely carbonaceous. We show that the variations of the mid-IR (5–35 μm) spectra of PDRs can be explained by the photo-chemical processing of these VSGs and PAHs, VSGs being the progenitors of free PAHs.

Key words. astrochemistry – dust, extinction – ISM: lines and bands – reflection nebulae – infrared: ISM – methods: numerical

1. Introduction

The Aromatic Infrared Bands (AIBs), observed in emission in UV irradiated interstellar matter, have been attributed to Polycyclic Aromatic Hydrocarbons (PAHs) by Léger & Puget (1984) and Allamandola et al. (1985). Later observations with the Infrared Space Observatory (ISO) (Kessler et al. 1996) have largely contributed to the study of these macromolecules and their evolution in the ISM, together with many fundamental studies involving both laboratory experiments and quantum chemistry calculations. Although these studies seem to confirm that the AIB carriers are PAH-like species, a precise identification of the mixture of the interstellar molecules has not been possible yet. In parallel to these efforts to identify the AIBs,

Cesarsky et al. (2000) gave strong evidence of the existence of Very Small carbonaceous Grains (VSGs), from the ISO observations of the reflection nebula Ced 201. They proposed that these grains, emitting broad bands and a continuum, are destroyed by UV radiation and/or shock waves freeing the AIB carriers in the process. In addition, they suggested that these grains could be present everywhere in the ISM, but detectable in the mid-IR only under special conditions. Photo-Dissociation Regions (PDRs) constitute the transition regions between UV-rich regions and molecular clouds. In PDRs with limited excitation, the infrared emission is dominated by dust particles. To emit at mid-IR wavelengths, these particles must be heated to high temperatures which implies that they are small (from ~ 1 to 10 nm) and not in thermal equilibrium with the incident radiation field. PDRs are extremely useful for probing the properties of very small dust particles, because the local UV radiation field varies as a function of the distance from the star. This change in the excitation conditions is expected to provoke changes in the dust processing and therefore significant variations in the observed infrared spectrum. Analysis of the observed spectral variations can provide valuable clues on the physical and chemical properties

* This work is based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA.

** Spectra are only available in electronic form (FITS files) at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via <http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/469/575>

of the small emitting particles. The study of such regions with a high spectral and spatial resolution was first made possible with ISOCAM onboard the Infrared Space Observatory (ISO) (Cesarsky et al. 1996). Most analyses of the ISOCAM spectral cubes used simple mathematical decompositions of the observed spectra in bands (the AIBs) and continua (Abergel et al. 2003). However, more sophisticated methods were used elsewhere to describe the observed spectrum in each pixel as the combination of the emission of several chemical populations. This approach requires the use of methods such as Single Value Decomposition (SVD is applied to the covariance matrix of the data, which is equivalent to Principal Component Analysis). Applying SVD to the observations of different PDRs, Boissel et al. (2001) and then Rapacioli et al. (2005) (RJB hereafter) were able to extract the spectra of different types of very small dust particles: PAH-like species and VSGs. Although pioneering, this work was restricted to two objects of the ISOCAM database, because of the limited efficiency of the method.

In the field of signal processing, a number of algorithms have been developed to extract original signals from observations which are mixtures of these signals. Such algorithms are referred to as *Blind Signal Separation* (BSS) methods. The most widely used are Independent Component Analysis (ICA) approaches, with algorithms like *FastICA* (Hyvarinen 1999). The use of ICA is so widespread that it is sometimes presented as the only class of BSS methods, although some other algorithms like Non-negative Matrix Factorisation (NMF) (Lee & Seung 1999) have proven to be efficient.

This paper presents the results of a study conducted with the newly available spectro-imagery data from the Infrared Spectrograph (IRS) (Houck et al. 2004) onboard *Spitzer*. Our goal is to use both the spectral and spatial information available in IRS data cubes to reveal the origin of the observed spectral variations. The IRS spectrograph provides several improvements when compared to ISOCAM, including sensitivity (about 0.06 mJy from 6 to 15 μm , and 0.4 mJy from 15 to 38 μm), a better spectral resolution ($\frac{\lambda}{\Delta\lambda} \sim 60\text{--}127$ in the low resolution mode) and a range of wavelengths extended to 35 μm . We use two classes of BSS algorithms, i.e. ICA and NMF, to extract the spectra of the different populations of very small particles. This work, carried out on four PDRs, extends previous studies in the following ways:

- i) we have gained in sensitivity and spectral resolution;
- ii) a wider range of wavelengths is available;
- iii) the BSS methods are more powerful than SVD-based methods;
- iv) as a consequence, our results involve fainter objects.

After describing the observations and data reduction in Sect. 2, we define the BSS methods we have used for our analysis in Sect. 3. We then present the results we have obtained applying these methods to the data (Sects. 4, 5) and provide evidence of their efficiency (Sect. 6). We show that the extracted spectra can be attributed to different populations of dust (Sect. 7). Using the extracted information we infer the properties and evolution of these populations in the studied PDRs in Sect. 8.

2. Observations

2.1. Selected PDRs

This study was conducted as part of the *SPECPDR*¹ program (Joblin et al. 2005) in which a total of 11 PDRs were observed

with all the instruments onboard *Spitzer*. The PDRs selected for this paper, Ced 201, the ρ -Ophiuchi filament, NGC 7023 East and North, have mild radiation fields (200 to 1300 in units of the Habing (1968) field, $G_0 = 1.6 \times 10^{-10} \text{ W cm}^{-2}$) and relatively simple geometries. They were chosen among the 11 observations because their IR emissions show the strongest spectral changes across the PDR, thus providing the best information on the nature of the carriers.

2.1.1. Ced 201

Ced 201 is a compact reflection nebula lying in the Cepheus constellation about 420 pc from the Sun (Casey 1991). Unlike common reflection nebulae, which are illuminated by their own star, Ced 201 is the result of a chance encounter between a B9.5 V star (BD+69°1231) and a molecular cloud, at a relative velocity of about 12 km s⁻¹ (Witt et al. 1987). Because of this peculiarity, Ced 201 is a very attractive object for the spectroscopic study of interstellar dust. Cesarsky et al. (2000) revealed the presence of very small carbonaceous grains, with a spectrum exhibiting bands and continuum emission, identified as the general mid-IR VSG spectrum by RJB.

2.1.2. NGC 7023 East and North-West

NGC 7023 is a well studied reflection nebula located at 430 pc from the sun (van den Ancker et al. 1997). Three PDRs are visible around the Herbig Be illuminating star, HD 200775. The brightest one, NGC 7023 North-West (NGC 7023-NW hereafter), lies about 40'' North-West of the star and another one lies around 70'' South. The last one, NGC 7023 East (NGC 7023-E hereafter), is located further East of the star ($\sim 170''$). The spectroscopic studies of the North-West and South PDRs were performed by Cesarsky et al. (1996) using ISOCAM, and by Werner et al. (2004) with IRS, leading to the discovery of new infrared features. In the present paper, we analyse the IRS spectro-imagery observations of NGC 7023-E and NGC 7023-NW.

2.1.3. The ρ -Ophiuchi filament

The ρ -Ophiuchi molecular cloud is a nearby star-forming region. The western part of this cloud is bounded by a filament structure (Oph-fil hereafter), forming an edge-on PDR, illuminated by the B2 HD147889 star (Abergel et al. 1999; Habart et al. 2003).

2.2. IRS observations and data reduction

The four PDRs were observed using IRS onboard *Spitzer*, in the low resolution ($\frac{\lambda}{\Delta\lambda} = 60\text{--}127$) mode. The data were obtained in the spectral mapping mode (with a spatial resolution from 3.6 to 10.2 arcsec), where the spacecraft moves the spectrograph's slit before each integration, in order to cover a given area of the extended PDR. Each full spectral cube was assembled using the CUBISM software (Smith et al. 2007). The original Basic Calibrated Data (BCD) files are from the S12 pipeline of the *Spitzer* Science Center. Calibration was achieved using the FLUXCON tables provided by the *Spitzer* Science Center together with the retrieved data. Additional corrections for extended sources were applied (Smith et al. 2004) and the remaining bad pixels were removed by hand. In the low spectral resolution mode, IRS is divided in two modules: the Short wavelengths Low resolution (SL) module, and Long wavelengths Low resolution (LL) module. Each module is sub-divided into two

¹ http://www.cesr.fr/joblin/SPECPDR_public/Home

orders: SL2 (5 to 8.7 μm) and SL1 (7.4 to 14.5 μm) for SL and LL2 (14 to 21.3 μm) and LL1 (19.5 to 38 μm) for LL. For Ced 201, both SL and LL modules were used, providing the data from 5 to 38 μm . For NGC 7023-NW we assembled the SL data from the PID 28 *Spitzer* program² with the LL2 order data obtained as “bonus” in the SPECPRD program, while the LL1 order of the IRS was observing NGC 7023-E. For NGC 7023-E and Oph-fil, the SPECPRD observations were only performed with the LL module, to complement the available and good-quality ISOCAM spectral cubes. The full spectral cubes, ranging from 5 to 38 μm , were constructed by assembling the cubes from each module, and reprojecting the data on a grid with the lowest resolution (i.e. the LL grid for Ced 201 and NGC 7023-NW, and the ISOCAM grid for NGC 7023-E and Oph-fil). The spectral data from 35 to 38 μm is disregarded because of its poor quality. The resulting set of data consists of four spectral cubes with wavelengths ranging from 5 to 35 μm . Each spectral cube is thus a 3-dimensional matrix $C(p_x, p_y, \lambda)$, which defines the variations of the recorded data with respect to the wavelength λ , for each considered position with coordinates (p_x, p_y) in the cube.

2.3. Zodiacal light

The contribution from zodiacal light to the observations can become significant when looking at relatively faint objects. This is the case in Ced 201. Fortunately, this nebula is situated in a region of the sky which is slightly contaminated by zodiacal light. However we removed the background by subtracting the spectrum from a nearby region of dark sky (including background emission and zodiacal light). The other objects show stronger emission and are thus less subject to relative contamination by zodiacal light (less than 10% at 20 μm). For these objects no correction was applied.

2.4. MIPS-SED and IRAS data

We have combined the Infrared-Red Astronomical Satellite (IRAS) photometry and Multi-Band Imaging Photometer (MIPS, onboard *Spitzer*) data with the IRS data to build the spectral energy distribution of Ced 201, between 5 and 100 μm . The IRS spectrum of Ced 201 was obtained by averaging the full IRS cube (see Sect. 2.2). The IRAS data was retrieved from the archive. Because the IRAS beam is larger than the whole region observed in the spectral mapping mode with IRS, it is in principle not possible to compare the fluxes observed with the two instruments. Thus we have made the assumption that the emission of Ced 201 is dominated by the region close to the illuminating star and covered by IRS observations. Therefore the flux calibration of the IRAS points at 12, 25, 60 and 100 μm points was achieved by scaling the 12 and 25 μm points to the IRS-LL data. The MIPS-SED cubes were built stacking the BCD files from the S12 pipeline. The MIPS-SED intensity spectrum was done by averaging the data over the common region with IRS.

3. Blind signal separation methods

3.1. Problem overview

Blind Signal Separation is commonly used to restore a set of unknown “source” signals from a set of observed signals which are

mixtures of these source signals, with unknown mixture parameters (Hyvarinen et al. 2001). It has e.g. been applied in acoustics for unmixing recordings, or in the biomedical field for separating mixed electromagnetic signals produced by the brain (Sajda et al. 2004). BSS is most often achieved using ICA methods. In particular, an algorithm called *FastICA* (Hyvarinen 1999) has proven to be efficient for recovering source signals. An alternative class of methods for achieving BSS is NMF, which was introduced in Lee & Seung (1999) and then extended by a few authors.

In the field of Astrophysics, ICA has been successfully used for spectra discrimination in infrared spectro-imaging of Mars ices (Forni et al. 2005), elimination of artifacts in astronomical images (Funaro et al. 2003) or extraction of cosmic microwave background signal in *Planck* simulated data (Maino et al. 2002). To our knowledge, NMF has not yet been applied to astrophysical problems. However, it has been used to separate spectra in other application fields, e.g. for magnetic resonance chemical shift imaging of the human brain (Sajda et al. 2004) or for analysing wheat grain spectra (Gobinet et al. 2005)

The simplest version of the BSS problem concerns so-called “linear instantaneous” mixtures. It is modelled as follows:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (1)$$

where \mathbf{X} is an $m \times n$ matrix containing n samples of m observed signals, \mathbf{A} is an $m \times r$ mixing matrix and \mathbf{S} is an $r \times n$ matrix containing n samples of r source signals. The observed signal samples are considered to be linear combinations of the source signal samples (with the same sample index). It is assumed that $r \leq m$ in most investigations, including this paper. The objective of BSS algorithms is then to recover the source matrix \mathbf{S} and/or the mixing matrix \mathbf{A} from \mathbf{X} .

The correspondence between the generic BSS data model (1) and the 3-dimensional spectral cube $C(p_x, p_y, \lambda)$ to be analysed in the present paper may be defined as follows. Each observed signal, consisting of all samples available for this signal, corresponds to a row of \mathbf{X} in Eq. (1). In this paper, the sample index is associated to the wavelength λ , and each observed signal consists of the overall spectrum recorded for a cube pixel (p_x, p_y) . Moreover, each observed spectrum is a linear combination of “source spectra”, which are respectively associated to each of the (unknown) types of dust particles that contribute to the recorded spectral cube. Therefore, the recorded spectra may here be expressed according to (1), with unknown combination coefficients in \mathbf{A} , unknown source spectra in \mathbf{S} and an unknown number r of source spectra. Sections 3.2 and 3.3 introduce the two BSS methods we have tested on these data, i.e. *FastICA* and an NMF algorithm.

3.2. *FastICA*

FastICA is a statistical BSS method intended for stationary, non-Gaussian and mutually statistically independent random signals (Hyvarinen 1999). It is expressed for zero-mean signals hereafter. In practice, such signals are obtained by first subtracting the sample mean of each observed signal to all signal samples in the corresponding row of \mathbf{X} in (1). For the sake of simplicity, the notation \mathbf{X} refers to these zero-mean signals hereafter.

The next step of *FastICA* consists in applying SVD to the covariance matrix of the observed data, i.e. in deriving a matrix \mathbf{Z} of transformed signals defined as

$$\mathbf{Z} = \mathbf{M}\mathbf{X} \quad (2)$$

² *Spectral Line Diagnostics of Shocks and Photon-Dominated Regions*, PI Giovanni Fazio.

where \mathbf{M} is selected so that: i) all signals in \mathbf{Z} are mutually uncorrelated, ii) each of these signals has unit power and iii) the number of signals in \mathbf{Z} is \hat{r} . The value of \hat{r} is selected as follows. When applied to the m signals in \mathbf{X} , SVD intrinsically yields m output components. Keeping all these components therefore corresponds to selecting $\hat{r} = m$. Instead, if $r < m$, one may choose to only keep the \hat{r} output components which have the highest powers, with \hat{r} selected so that $r \leq \hat{r} < m$ (see details on p. 129 of Hyvarinen et al. 2001). This reduces the dimensionality of the processed data and allows one to combine the following two features: i) using $\hat{r} \geq r$ still makes it possible to recover all source signals from \mathbf{Z} and ii) selecting $\hat{r} < m$ decreases in \mathbf{Z} the influence of noise components which exist in real observations \mathbf{X} but were not taken into account in the above data model (2).

The basic version of FastICA then extracts a first source signal from the matrix \mathbf{Z} . The criterion used consists of maximising the non-Gaussianity of an output signal defined as a linear combination of the signals in \mathbf{Z} . Therefore, denoting z a column of \mathbf{Z} corresponding to a given sample index, the corresponding sample of the output signal reads

$$y = d^T z, \quad (3)$$

where the column vector d is constrained to have unit norm. Several versions of the FastICA method have been defined, depending on which parameter is used to measure the non-Gaussianity of y . The most standard parameter is the absolute value of the non-normalised kurtosis, defined for a zero-mean signal y as

$$\text{Kurt}(y) = E[y^4] - 3(E[y^2])^2 \quad (4)$$

where $E[\cdot]$ stands for expectation. Various algorithms may then be used for adapting d so as to maximise that absolute kurtosis parameter. Before the FastICA algorithm was introduced, Delfosse & Loubaton (1995) optimised it by using a standard gradient ascent procedure. FastICA is an alternative, fixed-point, optimisation algorithm described in Hyvarinen (1999). It has been shown to yield much faster and more reliable convergence than gradient procedures. Moreover, it does not require one to select any tunable parameter (such as the adaptation gain of gradient algorithms). Once a first source signal has been extracted as an output signal y , *FastICA* removes its contributions from all observed signals contained by \mathbf{X} . This yields a matrix X' which contains only $r-1$ sources. The same procedure as above is then applied to X' in order to extract another source. This “deflation” procedure is repeated until all sources are extracted from the observations.

3.3. Non-negative matrix factorisation

Unlike ICA, NMF is based on assuming non-negative source signals and mixing coefficients, without requiring independence of the source signals. It aims to recover the r source signals by approximating the supposedly non-negative matrix \mathbf{X} with the following factorisation:

$$\mathbf{X} \approx \mathbf{W}\mathbf{H} \quad (5)$$

where \mathbf{W} is a $m \times r$ non-negative weight matrix and \mathbf{H} is a $r \times n$ non-negative matrix of approximated “source” signals. The approximate quantity in Eq. (5) can be optimised by adapting the non-negative matrices W and H so as to minimise the squared Euclidian distance $\|\mathbf{X} - \mathbf{W}\mathbf{H}\|^2$ or the divergence $D(\mathbf{X}|\mathbf{W}\mathbf{H})$ (Lee & Seung 2001), defined as

$$\|\mathbf{X} - \mathbf{W}\mathbf{H}\|^2 = \sum_{ij} (X_{ij} - (\mathbf{W}\mathbf{H})_{ij})^2 \quad (6)$$

and

$$D(\mathbf{X}|\mathbf{W}\mathbf{H}) = \sum_{ij} (X_{ij} \log \frac{X_{ij}}{(\mathbf{W}\mathbf{H})_{ij}} - X_{ij} + (\mathbf{W}\mathbf{H})_{ij}). \quad (7)$$

The adjustment of \mathbf{W} and \mathbf{H} can be achieved using classical gradient descent although this can be problematic (see Lee & Seung 2001). Therefore, the authors have proposed a new algorithm requesting that the Euclidian distance is non increasing under the iterative update rule

$$H_{a\mu} \leftarrow H_{a\mu} \frac{(\mathbf{W}^T \mathbf{X})_{a\mu}}{(\mathbf{W}^T \mathbf{W}\mathbf{H})_{a\mu}}, W_{ia} \leftarrow W_{ia} \frac{(\mathbf{X}\mathbf{H}^T)_{ia}}{(\mathbf{W}\mathbf{H}\mathbf{H}^T)_{ia}} \quad (8)$$

and that the divergence is also non increasing under the rule

$$H_{a\mu} \leftarrow H_{a\mu} \frac{\sum_i W_{ia} X_{i\mu} / (\mathbf{W}\mathbf{H})_{i\mu}}{\sum_k W_{ka}}, W_{ia} \leftarrow W_{ia} \frac{\sum_\mu H_{a\mu} X_{i\mu} / (\mathbf{W}\mathbf{H})_{i\mu}}{\sum_\nu H_{a\nu}}. \quad (9)$$

Thus, the following iterative algorithm can be derived from this result in order to minimise either the euclidian distance or divergence:

1. fix r ;
2. randomly initialise matrices \mathbf{W} and \mathbf{H} ;
3. update these matrices with the update rule (8) or (9);
4. if convergence is reached, then end. Otherwise go back to Step 3.

When convergence is reached, \mathbf{H} provides an approximation of “source” signals.

4. Application of the BSS methods

4.1. Suitability of BSS methods for analysis of *Spitzer-IRS* cubes

In order to apply the above BSS methods to the IRS data cubes, it is necessary to make sure that the “linear instantaneous” mixture condition is fulfilled (see Sect. 3.1). Here we consider that each observed spectrum is a linear combination of “source spectra”, which are due to the emission of different populations of dust particles. The main effect that can disturb the linearity of the model is radiative transfer as shown by Nuzillard & Bijaoui (2000), due to the non-linearity of the equations. In our case however, this effect is not expected to be a major concern. The mid-IR emission of dust particles occurs in the external layers of the cloud, at optical depth lower or comparable with H_2 emission (Joblin et al. 2007). Using the Meudon PDR code (Le Bourlot et al. 1993; Le Petit et al. 2006) we found that H_2 emission occurs at extinctions of $A_V \leq 0.2$ in NGC 7023-E and Oph-fil. For Ced 201 which is a peculiar case, as the star has penetrated the cloud, we use the visual extinction $A_V = 0.21$ of the illuminating star BD+69°1231 derived by Witt et al. (1987). From this we conclude that very small particles emit in the infrared at a maximum visual extinction of $A_V \sim 0.2$ in the studied PDRs. We can then roughly estimate $\tau_{9.7}$, the optical depth of the $9.7 \mu\text{m}$ silicate band which is the main absorption in the mid IR range. Typically, the $\tau_{9.7}/A_V$ ratio is between ~ 10 and ~ 20 (Mathis 1990), which yields $\tau_{9.7} < 2 \times 10^{-2}$. This value is very small, indicating that the effect of radiative transfer on the mid-IR spectra is very weak. Thus, the source spectra are not significantly affected by this re-absorption, and we conclude that the “linear instantaneous” mixture model is fulfilled. This is validated a posteriori as we show that the extraction has the same efficiency around $10 \mu\text{m}$, where the silicates absorb, as at the other wavelengths (see Fig. 10).

4.2. Preliminary tests

To test the BSS methods presented in Sect. 3, we generated two synthetic spectra: one representing a PAH-type emission spectrum, consisting of Lorentzian profiles at $6.2\ \mu\text{m}$, $7.7\ \mu\text{m}$, $8.6\ \mu\text{m}$ and $11.3\ \mu\text{m}$, and one representing the emission of VSGs (RJB), consisting of a broad emission band at $7.8\ \mu\text{m}$, the $11.3\ \mu\text{m}$ band and a second-order polynomial continuum. These spectra were mixed with a random 100×2 mixing matrix A , thus providing 100 artificial observed spectra, which are linear combinations of the initial spectra. These spectra were then analysed using *FastICA* and NMF. With both methods, we managed to recover the original spectra efficiently. In this case we have artificially mixed the sources, and therefore we can compare the extracted spectra with these sources to measure the efficiency of the methods. With both methods, the correlation between the original and unmixed signals is excellent (correlation coefficients higher than 0.995). However a realistic test must include the effect of noise. The noise in the IRS instrument is quite complex, and depends strongly on wavelength because of the different modules that are used. In order to simply quantify the effect of noise on the BSS we applied a white spatially homogeneous noise. Though this is not representative of the real IRS noise which would require a detailed study to estimate, this provides good knowledge on the response of the algorithms to noise. With both methods the results remain good for Signal to Noise Ratios (SNRs) over 2 (i.e. 3 dB). As a comparison, the estimated IRS instrumental noise using the SpecPET tool³ developed by the *Spitzer* Science Center, is between 20 and 100 (13–20 dB) depending on the wavelength for Ced 201. We have also estimated the noise using the spectrum of Ced 201 presented in Fig. 9 (position 2) and found a SNR of about 30 (14.8 dB) in the LL region which is consistent with the SpecPET estimation. Below $SNR = 2$, the efficiency of *FastICA* drops dramatically while NMF is still able to significantly recover the original signals (correlation coefficient greater than 0.85).

4.3. Application to *Spitzer* data

We have applied *FastICA* and NMF to the *Spitzer* data. The spectra from a given cube of observations $C(p_x, p_y, \lambda)$ are placed in the rows of the \mathbf{X} matrix defined in Eq. (1) or (5). To maintain efficiency, the noisiest spectra were removed. Thus, \mathbf{X} contains the m IRS spectra of a given PDR, over n points in wavelength. For NMF, Euclidian distance or divergence were minimised using Lee & Seung’s algorithm (Lee & Seung 2001) implemented with Matlab, to find \mathbf{W} and \mathbf{H} . The best results were found using the divergence criterion after about 1000 iterations (which takes less than one minute with a 3.2 GHz processor). The value of r (number of “source” spectra) is not imposed by the NMF method. The strategy for extracting the sources was as follows:

- Apply the algorithm with a minimum number of assumed sources $\hat{r} = 2$, to a given dataset, providing 2 sources.
 - If the found solutions are physically coherent and linearly independent we consider that at least $\hat{r} = 2$ sources can be extracted.
 - Else, we consider that the algorithm is not suited for analysis (never occurred in our case).
- Try the algorithm on the same dataset but with $\hat{r} = 3$ sources.
 - If the found solutions are physically coherent and linearly independent we consider that at least $\hat{r} = 3$ sources can be extracted.

→ Else, we consider that only two sources can be extracted, extraction was over at $\hat{r} = 2$ and thus $r = 2$.

- Same as previous step but with $\hat{r} = 4$ sources.
 - If the found solutions are physically coherent and linearly independent we consider that at least $\hat{r} = 4$ sources can be extracted.

→ Else we consider that only three sources can be extracted, extraction was over at $\hat{r} = 3$ and thus $r = 3$.

...

Physically incoherent spectra exhibit sparse peaks (spikes) which cannot be PDR gas lines. We found $r = 3$ for NGC 7023-NW and $r = 2$ for the other PDRs, implying that we could respectively extract 3 and 2 spectra from the data cubes. Similarly, with *FastICA*, all sources were extracted with $r = 2$ for Ced 201, Oph-fil and NGC 7023 E, and $r = 3$ for NGC 7023-NW. Thus we were able to extract the “source” signals from the original cubes using both methods. The choice of the method was therefore based on its suitability, considering the information we have on the sources. In the following section, we argue on why NMF is more consistent in our case.

4.4. Choice of the BSS method

NMF and *FastICA* have different constraints. The strength of *FastICA* is that the algorithm is guaranteed to converge towards a solution which yields separated spectra. However, this property only holds if the source signals are statistically independent, due to the non-gaussianity criterion imposed by *FastICA*. This criterion is not ideal in our application, where the “source” spectra are likely to be correlated due to chemical similarities between the different emitting populations (RJB). Moreover, the presence of noise can degrade the efficiency of the separation with *FastICA* (see Sect. 4.2). On the contrary, the non-negativity of the “sources” and mixing coefficients, required by NMF, are consistent with the non-negativity of emission spectra and of their linear combination. Thus NMF is likely to be more appropriate for the analysis of our data cubes. The only identified drawback of NMF is that its convergence point may depend on the random initialisation of the matrices (see Sect. 3.3). To test this effect, we ran NMF on the data of NGC 7023-E using 200 different initialisation conditions. We found that the outputs vary only slightly with the initial conditions. Since NMF is most likely to have the greatest efficiency for our application, we now detail the results obtained with this algorithm when applied to our data cubes.

5. Results

In this section we present the extracted spectra for Ced 201, Oph-fil, NGC 7023-E and 7023-NW (Figs. 1–4) and their distribution maps. These spectra are normalised to 1. Normalisation is required because intensity depends on the considered pixel. The molecular hydrogen lines have been subtracted in order to provide the spectra of dust only and the hydrogen/dust interaction will be the subject of a subsequent paper (Joblin et al. 2007).

5.1. Extracted spectra

From each of the cubes of Ced 201, Oph-fil and NGC 7023-E, we were able to extract two spectra. These independently found pairs of spectra have the same characteristics from one object to another. Moreover, for each object, the two spectra exhibit different features. The *Signal 1* spectra, presented in the upper part of Figs. 1–3, are characterised by a rising continuum combined

³ <http://ssc.spitzer.caltech.edu/tools/specpet/>

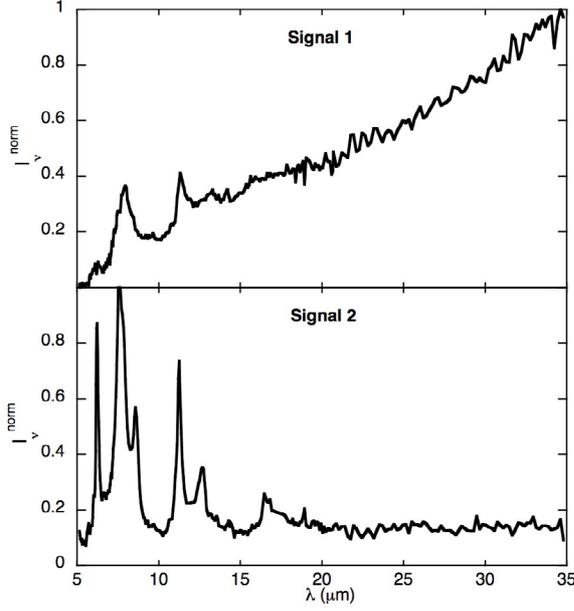


Fig. 1. BSS extracted spectra in Ced 201.

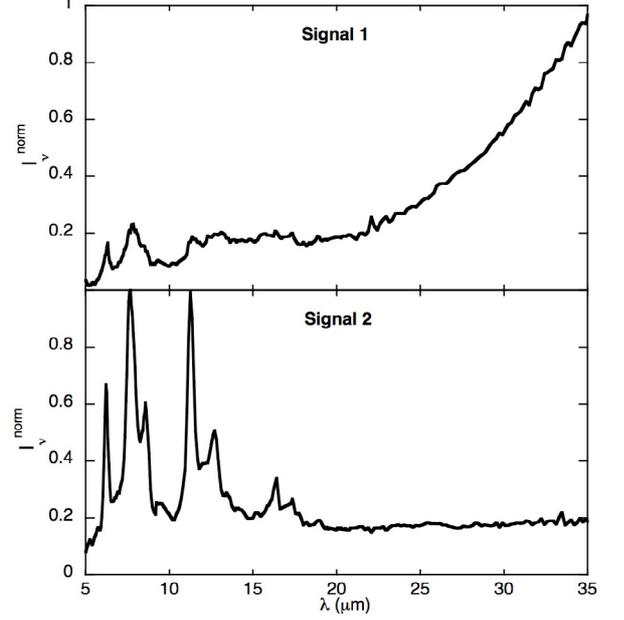


Fig. 3. BSS extracted spectra in NGC 7023-E.

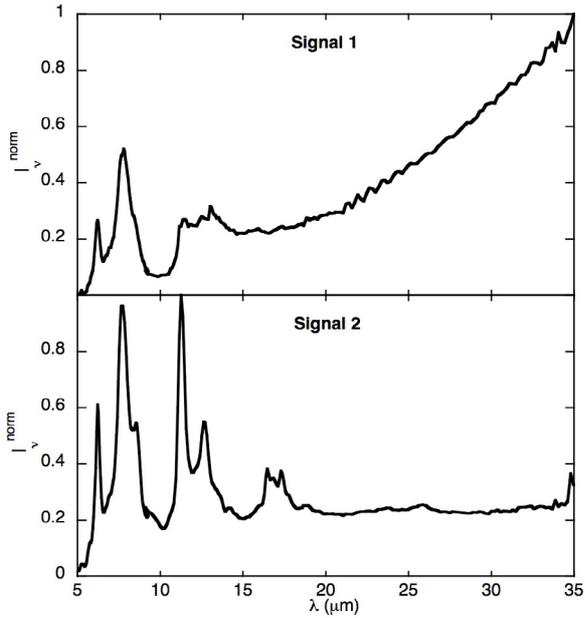


Fig. 2. BSS extracted spectra in Oph-fil.

with broad emission bands located at 6.2, 7.8, and 11.4 μm (see Table 1). On the contrary, the *Signal 2* spectra, presented on the lower parts of Figs. 1–3, are dominated by the AIB emission at 6.2, 7.6, 8.6, 11.3, 12.7 and 17.4 μm . In NGC 7023-NW we could extract three signals from the original data. These spectra are presented in Fig. 4. *Signal 1* is very similar to *Signal 1* found in the other PDRs studied above: it exhibits a broad 7.8 μm band and a steep continuum. *Signal 2* and *Signal 3* spectra are dominated by AIBs but with different relative intensities. *Signal 2* and *Signal 3* are respectively dominated by the 11.3 and 7.6 μm bands.

5.2. Spatial distributions of the extracted spectra

The observed data in matrix \mathbf{X} may be expressed as the product of two matrices, \mathbf{W} and \mathbf{H} (Eq. 5). Therefore, each observed

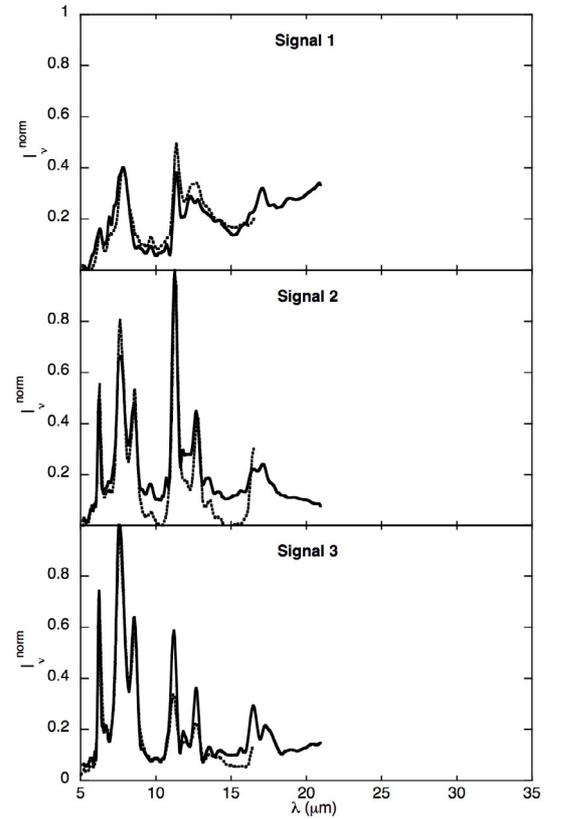


Fig. 4. BSS extracted spectra in NGC 7023-NW. In this case, the 7.7 μm band of signal 1 was artificially normalised to 0.4 to be comparable with the spectra extracted in the other objects. The dotted line spectra are those extracted from the ISOCAM observations of the entire NGC 7023 PDR by RJB.

spectrum from \mathbf{X} is defined as a linear combination of the extracted spectra in \mathbf{H} . In our case, each IRS spectrum at position (p_x, p_y) in a given cube can be written as follows:

$$\text{Obs}(p_x, p_y, \lambda) = \sum_n w(p_x, p_y)_n S_n(\lambda) \quad (10)$$

Table 1. Central wavelengths of the emission bands in the extracted spectra. The νI_ν integrated intensity ratios of the bands compared to the $7.7 \mu\text{m}$ band are written in parentheses. The bands were fitted using Lorentzian profiles.

Bands	VSGs			PAHs (neutral and cation mixture)		
	Ced 201	Oph-fil	7023-E	Ced 201	Oph-fil	7023-E
$6.2 \mu\text{m}$	6.2 (0.06)	6.21 (0.27)	6.27 (0.41)	6.24 (0.35)	6.23 (0.38)	6.24 (0.37)
$7.7 \mu\text{m}$	7.87 (1.0)	7.77 (1.0)	7.79 (1.0)	7.66 (1.0)	7.72 (1.0)	7.70 (1.0)
$8.6 \mu\text{m}$	–	8.54 (0.05)	8.51 (0.05)	8.60 (0.13)	8.57 (0.16)	8.58 (0.21)
$11.3 \mu\text{m}$	11.41 (0.19)	11.40 (0.13)	11.38 (0.13)	11.26 (0.13)	11.29 (0.26)	11.28 (0.30)
$12.7 \mu\text{m}$	–	–	–	12.59 (0.11)	12.66 (0.21)	12.62 (0.25)
Cont.	Yes	Yes	Yes	No/weak	No/weak	No/weak

where $\text{Obs}(p_x, p_y, \lambda)$ is the observed spectrum from the pixel indexed by (p_x, p_y) , $S_n(\lambda)$ is the n th extracted spectrum in H and $w(p_x, p_y)_n$ are the weights relative to spectrum n . In our case n is restricted to a maximum of 2 extracted spectra, except in the case of NGC 7023-NW where $n = 3$. The NMF algorithm extracts the sources up to arbitrary scaling factors, i.e. it provides $y_p = \eta_p S_p$, where η_p is an unknown scale factor and S_p is the p th “source”. Each output can be compared to the observations using the correlation between observations and NMF output. We centered the observations (and therefore the outputs) and considered that the extracted spectra are not correlated. Thus the correlation parameter c_p we consider here can be written:

$$c_p = E[\text{Obs}(p_x, p_y, \lambda)y_p] = \eta_p w(p_x, p_y)_p E[S_p(\lambda)^2] \quad (11)$$

where $E[.]$ stands for expectation. We can then calculate $\gamma(p_x, p_y) = c_p/c_{p'}$ the ratio of the correlations of Signal p and Signal p' to the observations:

$$\begin{aligned} \gamma(p_x, p_y) &= \frac{E[\text{Obs}(p_x, p_y, \lambda)y_p]}{E[\text{Obs}(p_x, p_y, \lambda)y_{p'}]} \\ &= \frac{w(p_x, p_y)_p}{w(p_x, p_y)_{p'}} \times \underbrace{\frac{E[S_p(\lambda)^2]\eta_p}{E[S_{p'}(\lambda)^2]\eta_{p'}}_{\kappa} \end{aligned} \quad (12)$$

where κ is a constant for a given object. γ traces the ratio of the weights $w(p_x, p_y)$ at the position (p_x, p_y) , and thus the emission ratio between sources p and p' , up to the κ scale factor, which does not depend on position (p_x, p_y) . The maps of Figs. 5–7 show the spatial distribution of the γ emission ratio between *Signal 1* and *Signal 2*, and Fig. 8 shows the c_p correlation map for each signal.

In Ced 201 the star is located in the central part of the nebula, and the *Signal 2* emission is also concentrated in this region (Fig. 5), whereas *Signal 1* dominates the periphery of the nebula (Fig. 5). In Oph-fil, the star is West of the filament. Again, *Signal 2* is dominant in regions closer to star (Fig. 6) whereas *Signal 1* emits in deeper regions (Fig. 6). In NGC 7023-E, the illuminating star is located West of the filament. The distribution maps show that the maximum *Signal 2* emission is on the edge of the filament (Fig. 7). Further East in the cloud, the emission is highly dominated by *Signal 1* (Fig. 7). Finally, in the case of NGC 7023-NW *Signal 1* is dominant behind the filament, *Signal 2* is dominant on the edge of the filament facing the

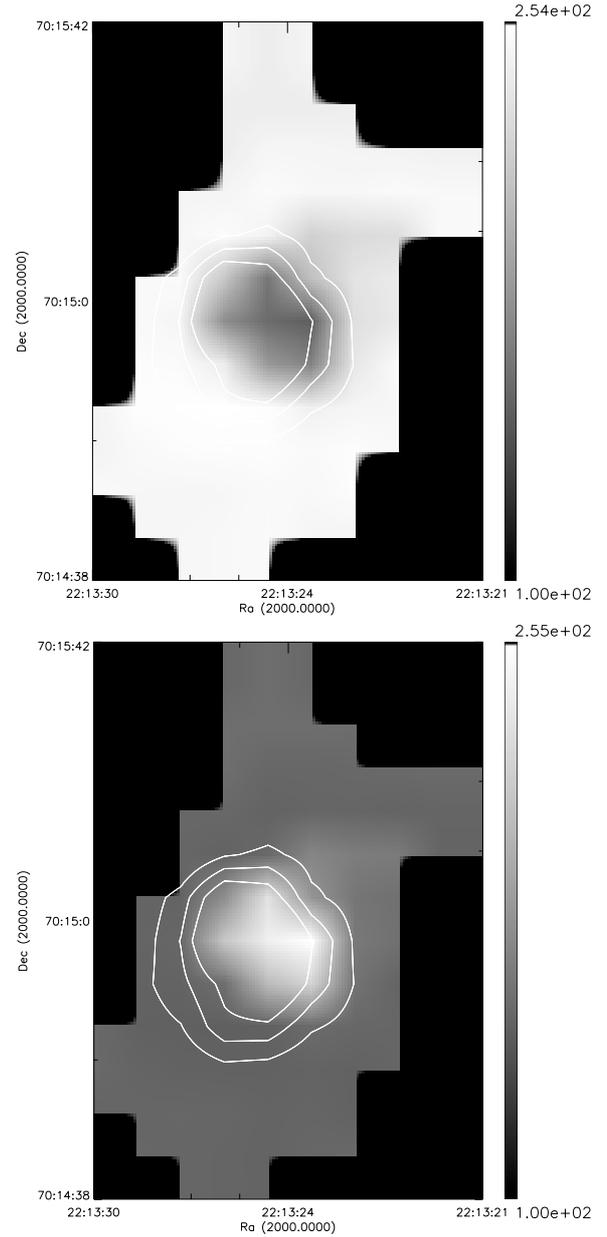


Fig. 5. Map of the emission ratios *Signal 1/Signal 2* (upper) and *Signal 2/Signal 1* (lower) in Ced 201. The $5\text{--}35 \mu\text{m}$ integrated mid-IR emission is in contours (contours 1, 2, 3 are respectively $1, 1.5, 2 \times 10^{-4} \text{ W m}^{-2} \text{ sr}^{-1}$). The illuminating star is located at the center.

star, and *Signal 3* in front of the filament, in the closest region to the star (Fig. 8).

6. Efficiency of the reconstruction of the observations using the extracted spectra

BSS methods are quite new to astrophysics, and for this reason they should be used carefully. However, several arguments confirm that in our particular case, NMF can be used efficiently.

Contrary to most approaches, there is absolutely no a priori information included in the algorithm, and no subjective physical constraints on the source spectra, so that the extracted spectra are purely based on the information contained in the data. Thus, it is quite convincing that we are able to recover a spatial structure without any information on this included in the algorithm.

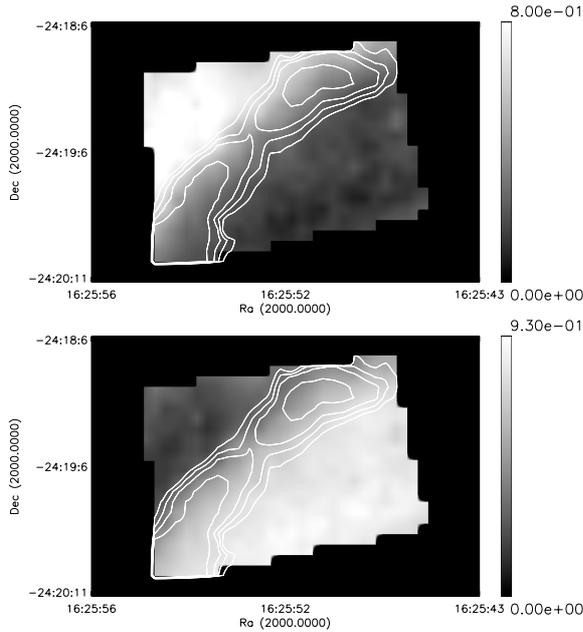


Fig. 6. Map of the emission ratios *Signal 1/Signal 2* (upper) and *Signal 2/Signal 1* (lower) in Oph-fil. The 5–35 μm integrated mid-IR emission of the filament is in contours (contours 1, 2, 3, 4 are respectively $3.5, 3.7, 3.9, 4.2 \times 10^{-4} \text{ W m}^{-2} \text{ sr}^{-1}$). The illuminating star lies South-West of the filament.

In this paper, we show that for the studied PDRs, the mid-IR emission spectrum can be fitted everywhere using linear combinations of two or three simple spectra extracted from the considered PDR. Figure 9 shows the reconstruction of the observations taken on three points of the Ced 201 PDR. The first position is at the periphery of the PDR, the second one is closer to the star and the last one is very close to the star. Using linear combinations of the extracted signals (see Sect. 5) we can reproduce the observations accurately, for the three positions. The residuals from this fit are shown in Fig. 10. The ratio of the power of the observed signal to the power of the residuals is about 30 for position 2 (see Fig. 9), which is consistent with the value of the SNR we have estimated (see Sect. 4.2). The residuals are at noise level, proving that the reconstruction from the extracted spectra of observations is efficient. It is remarkable that we are able to show that in some regions of Ced 201, nearly all the mid-IR emission is due to *Signal 1*. Also note that, even though we chose to use NMF for the reasons exposed in Sect. 4.4, *FastICA* which is a completely different method provided similar results, confirming their relevance.

7. Nature of the carriers of the extracted spectra

7.1. Continuum-dominated spectrum

In all the extractions, *Signal 1* exhibits a very clear association of continuum, and wide AIB emission (Tables 1, 2). This peculiarity was attributed to the emission of carbonaceous VSGs by Cesarsky et al. (2000) in Ced 201, from ISOCAM observations covering a wavelength range from 5 to 17 μm . The authors also proposed that these grains should be found everywhere in the interstellar medium. RJB have shown that using mathematical analysis, it is possible to identify these grains in other PDRs than Ced 201, from ISOCAM observations. They also mention that in the emission of this population, the position of the “7.7” μm band is shifted towards higher wavelengths which is also what we

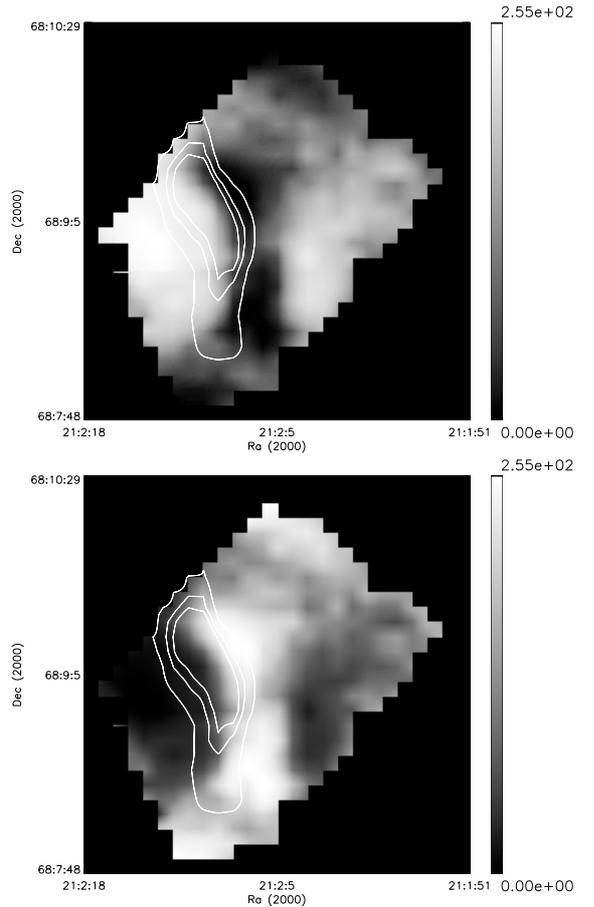


Fig. 7. Map of the emission ratios *Signal 1/Signal 2* (upper) and *Signal 2/Signal 1* (lower) in NGC 7023-E. The 5–35 μm integrated mid-IR emission of the filament is in contours (contours 1, 2, 3 are respectively $1, 1.2, 1.4 \times 10^{-4} \text{ W m}^{-2} \text{ sr}^{-1}$). The illuminating star lies West of the filament.

observe (Tables 1, 2). When compared to spectra 2 and 3, the main characteristics of the *Signal 1* bands are widths typically 50% higher and a clear red-shift of the 7.7 and 11.3 μm bands. The 8.6 μm band is weak or absent and it is not possible to extract the 12.7 μm band from the 12–15 μm plateau. The 15–18 μm region is difficult to study in details because of the presence of a strong H_2 emission at 17 μm .

The IRS-LL module enabled us to build spectral cubes up to 35 μm (note that the residual fringes observed in the 20–30 μm region are instrumental effects and are not due to the algorithm or signatures of the dust). We have shown that a single spectrum, thus a single continuum, can reproduce the continuum emission all over the Ced 201 PDR (see Sect. 6). This indicates that the shape of the continuum does not vary significantly with distance to the star, though its intensity varies. From this observation, we can conclude that these grains are most likely transiently heated. If this was not the case, their temperature, and therefore the continuum shape, should strongly change as a function of the distance from the star, as it happens for classical big grains (BGs). However we cannot completely rule out the presence of a fraction of grains at thermal equilibrium emitting in the 25–35 μm range. On this basis, and because of the strong similarity of our *Signal 1* spectrum and the ISO VSG spectra in the 5 to 16 μm region, either observed (Cesarsky et al. 2000) or extracted mathematically (RJB), we propose to attribute *Signal 1* in the 5–25 μm range to the emission of VSGs which appear to be carbonaceous,

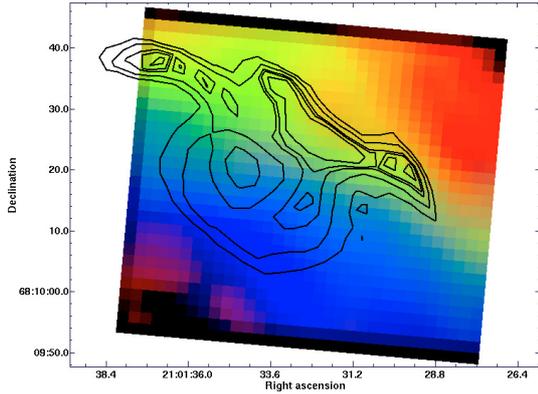


Fig. 8. Map of the correlation c_p of each extracted signal with observations, *Signal 1* in red, *Signal 2* in green and *Signal 3* in blue. The contours are the Infrared Array Camera (IRAC) emission at $8\ \mu\text{m}$ showing the filamentary structure of this PDR. The illuminating star is situated in the lower left hand corner. In this region a slight artifact correlation of *Signal 1* with observations is seen due to the presence of a continuum emanating from big grains at thermal equilibrium.

and probably corresponding to the VSGs used in the Désert et al. (1990) model. In the following section we fit the emission of the Ced 201 PDR using the model of Désert et al. (1990) in order to provide some clues on the contribution of the two populations of grains (VSGs and BGs) to the continuum.

7.2. Modelling the emission of VSGs versus BGs

The problem of calculating the infrared emission from VSGs in the ISM has been addressed by several authors (Désert et al. 1990; Siebenmorgen et al. 1992; Draine & Li 2001; Zubko et al. 2004). One question we would like to address is whether BGs contribute to the continuum below $35\ \mu\text{m}$ in the studied PDRs. Although this point would deserve a full study by itself, we provide here the first elements. To estimate the emission of BGs, we fitted the total IR emission from Ced 201 gathering the IRS, MIPS, MIPS-SED and IRAS data, using the model of Désert et al. (1990). The input parameters we have used for the model are listed in Table 3. The radiation field was calculated using the Meudon PDR code (Le Bourlot et al. 1993; Le Petit et al. 2006). The stellar spectrum is from the Kurucz library with a temperature of $10\ 500\ \text{K}$ (Kurucz 1991) and a radius of $2.26 R_\odot$. The distance to the cloud, d , was adjusted to obtain the value of $G_0 = 300$ at the interface as given by Young Owl et al. (2002). This gives $d = 0.0145\ \text{pc}$. This PDR has the strongest UV field of the three PDRs for which we have the full $5\text{--}35\ \mu\text{m}$ data, and should thus host the hottest BGs, likely to emit in the mid-IR continuum. The resulting fit (Fig. 11) shows that the emission from VSGs is dominant up to $50\ \mu\text{m}$. However, because this model is not completely adapted to the grain populations we consider, and because of the difficulties to calibrate the IRAS data, it is not possible to rule out BGs emission over $25\ \mu\text{m}$. Moreover, as mentioned by Zubko et al. (2004), the optical properties of BGs and VSGs are not well constrained, implying that their size distribution cannot be derived precisely. Still, the model provides qualitative information on the emitting population, and we can conclude from the fit that the emission in the $25\ \mu\text{m}$ IRAS band is dominated by stochastically heated VSGs.

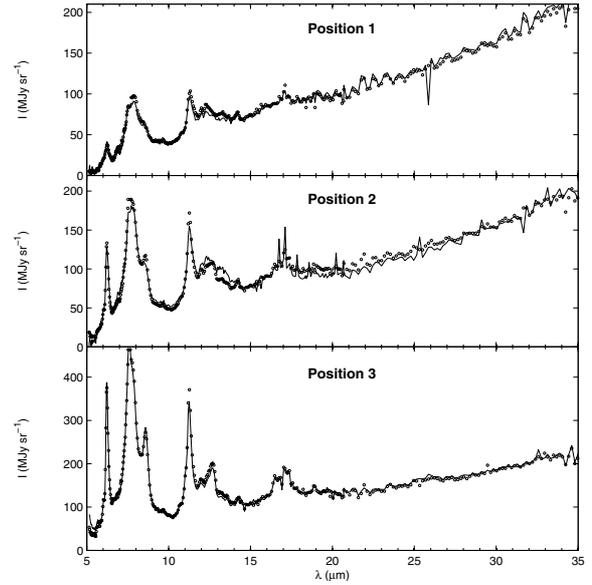


Fig. 9. Reconstruction of the observations on three pixels of the Ced 201 cube, using linear combinations of the extracted spectra. Solid lines are the observations and circles show the reconstruction. Position 1 is at the periphery $30''$ North of the central star, Position 2 is $8''$ North of the star and Position 3 is very close to the star.

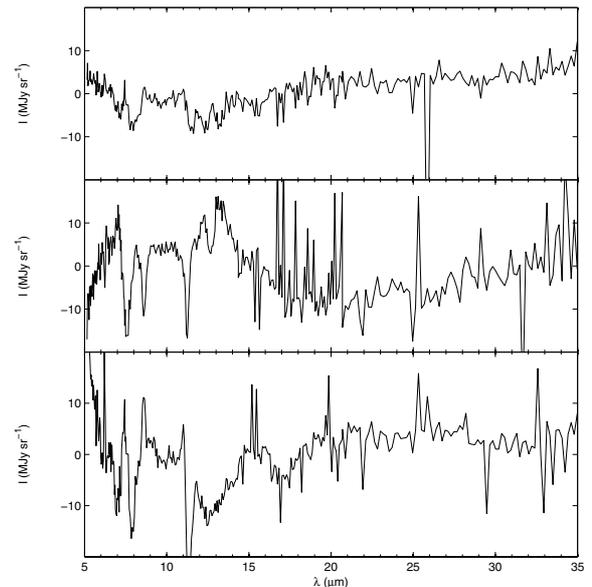


Fig. 10. Residuals from the fit presented in Fig. 9 (i.e. observation minus reconstruction).

7.3. Pure band spectra: PAHs

Signal 2 is dominated by the AIBs and is similar to the PAH-like emission of RJB. Therefore, we propose that *Signal 2* is due to free PAH emission. The three extracted PAH spectra of Figs. 1–3, are very similar. In this case, the $11\text{--}14\ \mu\text{m}$ region is dominated by the 11.3 and $12.7\ \mu\text{m}$ bands and no plateau such as the one found in the VSG spectra is detected. Here also the $15\text{--}18\ \mu\text{m}$ region is difficult to study because of the high contamination by the H_2 line at $17\ \mu\text{m}$ line. However, the spectrum extracted in the case of Ced 201 exhibits a lower $11.3\ \mu\text{m}$ band than in NGC 7023-E and Oph-fil when compared to the intensity of the “ 7.7 ” μm band (see Table 2).

Table 2. Same as Table 1 but for NGC 7023-NW.

Bands	This work			RJB		
	VSGs	PAHs ⁰	PAHs ⁺	VSGs	PAHs ⁰	PAHs ⁺
6.2 μm	6.24 (0.28)	6.22 (0.34)	6.24 (0.35)	6.29 (0.35)	6.27 (0.47)	6.28 (0.40)
7.7 μm	7.74 (1.0)	7.64 (1.0)	7.61 (1.0)	7.82 (1.0)	7.65 (1.0)	7.63 (1.0)
8.6 μm	–	8.55 (0.29)	8.57 (0.27)	–	8.58 (0.31)	8.59 (0.27)
11.3 μm	11.36 (0.10)	11.25 (0.37)	11.18 (0.17)	11.36 (0.28)	11.30 (0.42)	11.18 (0.12)
12.7 μm	–	12.66 (0.16)	12.67 (0.08)	–	12.76 (0.17)	12.63 (0.08)
Cont.	Yes	No/weak	No/weak	Yes	No/weak	No/weak

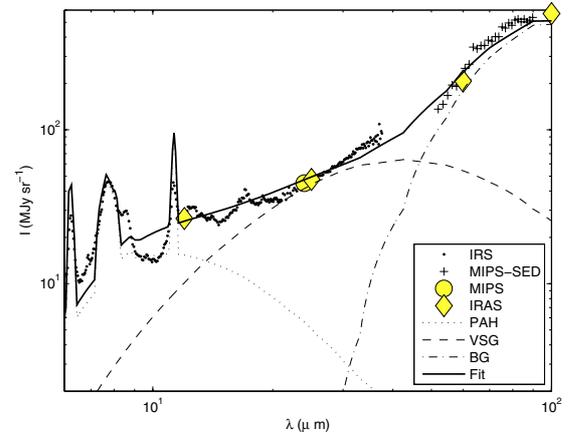
Table 3. Input parameters for the fit of the Ced 201 IR emission with the model of Désert et al. (1990).

Component	Relative mass abundance	α^*	a_{\min}	a_{\max}
PAHs	3.4×10^{-4}	3.0	0.4	1.2
VSGs	5.7×10^{-3}	2.6	1.2	10
BGs	0.994	2.0	10	500

* α is the exponent in the power law of the size distribution: $n(a) \propto a^{-\alpha}$ where a is the size in nanometers and $n(a)$ gives the numer density of grains with a radius between a and $a + da$.

This variation of the 11.3/7.7 μm ratio was interpreted by Joblin et al. (1996); Sloan et al. (1999); Bregman & Temi (2005) as an effect of variation in the relative abundance of neutral and positively ionised PAHs (respectively PAHs⁰ and PAHs⁺ hereafter), and Flagey et al. (2006) have shown that the effect of size on this ratio was weak. Thus, the PAH extracted spectra in Ced 201, Oph-fil and NGC 7023-E are mixtures of neutral and cationic PAHs. In NGC 7023-NW, RJB were able to identify the emission spectra of two populations respectively dominated by PAHs⁰ and PAHs⁺. In this object we were also able to extract two pure band spectra similar to those determined by RJB which we also attribute to PAHs⁰ and PAHs⁺ (see Fig. 4). This strengthens the validity of both results. Note however that the PAH⁺ spectrum we extract here has a higher 11.3 to 7.7 μm integrated intensity ratio than the one of RJB (see Tables 1, 2). This is because the observations we analyse here do not cover the regions close to the star. Thus our PAH⁺ population is “less” ionised than the one extracted by RJB, which we will consider as reference hereafter. On the other hand, the PAH⁰ spectrum we have extracted is close to the one of RJB (see Table 2 and Fig. 4). This extracted spectrum is also similar to the one observed in the HII region of the Horsehead nebula by Compiègne et al. (2007), where PAHs are found to be purely neutral because of the extreme abundance of electrons. From this we can conclude the following:

- the PAH⁺ spectrum extracted by RJB is dominated by PAH cations. In the following, we consider this spectrum to be representative of cation emission. This was confirmed by Flagey et al. (2006);
- the PAH⁰ spectrum can be attributed to a population containing only neutral PAHs.

**Fig. 11.** Spectral energy distribution (5–100 μm) of the Ced 201 reflection nebula. The continuous line is the emission modeled according to the dust model of Désert et al. (1990) with the parameters from Table 3.

7.4. Ionisation of PAHs in Ced 201, Oph-fil and NGC 7023-E

As mentioned in Sect. 7.3, the PAH spectra of Ced 201, Oph-fil and NGC 7023-E are mixtures of neutral and cation PAH emission. The reason why we were not able to extract the individual PAH⁰ and PAH⁺ spectra in these objects is that the spectral variations are not strong enough in these cases to provide sufficient information to the algorithm to disentangle the two populations. In order to estimate the normalised fraction of positively ionised PAHs $n_+/(n_0 + n_+)$ in these PDRs, we used the PAH⁰ and PAH⁺ spectra extracted from the data of NGC 7023-NW by RJB. By comparing the νI_ν integrated intensity ratios between the 7.7 and 11.3 μm bands of these extracted spectra (Table 2), to the pure band spectra of Ced 201, NGC 7023-E and Oph-fil (Table 1), we could estimate $n_+/(n_0 + n_+)$, using NGC 7023-NW as a reference. This ratio is equal to 0.97, 0.53, and 0.40 for Ced 201, Oph-fil and NGC 7023-E respectively (Table 5). This result shows that Ced 201 has a much higher proportion of ionised PAHs, probably because of its low density (see Table 4) implying a low recombination rate with electrons. RJB and Joblin et al. (1996) also show that a large proportion of ionised PAHs is found in NGC 7023-NW and NGC 1333-SVS3 respectively. This is likely due to the fact that these PDRs are close to the illuminating star. NGC 7023-E and Oph-fil, are dominated by neutral PAHs, probably because these PDRs are far from the star and have a higher density with respect to Ced 201 (see Table 4).

Table 4. Physical conditions of the PDRs.

Object	Star: ST, T(K)	n_{H} (cm ⁻³)	T (K)	G_0 (Habing)	Ref.
Ced 201	B9.5V, 10 400	4×10^2	200	300	Young Owl et al. (2002)
Oph-fil	B2V, 22 000	3×10^4	310	100	Habart et al. (2003)
NGC 7023-E	B5e, 15 000	7×10^3	340	87	Rapacioli et al. (2006)

Table 5. Estimated parameters in the studied PDRs: the $R_{7.7/11.3}$ band intensity ratio, the ionisation parameter, the UV field and the ionisation fraction.

Object	$R_{7.7/11.3}$	$G_0 \sqrt{T}/n_e$	G_0	$n_+/(n_0 + n_+)$
Ced 201	3.76	105000 ¹ , 75000 ²	300	0.97 ³
Oph-fil	1.89	1060 ¹ , 420 ²	100	0.53 ³
NGC 7023-E	1.63	1040 ¹ , 1650 ²	87	0.40 ³

¹ Using the results from the model described in Flagey et al. (2006). ² Using the data from Table 4. ³ Using the method described in this paper.

Flagey et al. (2006) have defined the $R_{7.7/11.3}$ parameter as the ratio between the λ_{ν} integrated intensity of the 7.7 and 11.3 μm band. We have calculated this parameter for Ced 201, NGC 7023-E and Oph-fil (Table 5). Flagey et al. (2006) have shown that the $R_{7.7/11.3}$ parameter is related to the ionisation parameter $G_0 \sqrt{T}/n_e$, where G_0 is the UV field in units of the Habing field, T is the gas temperature and n_e the electronic density. Using this result, we calculated the ionisation parameter from the PAH spectra of our PDRs with the $R_{7.7/11.3}$ ratio from Table 5. We also calculated this parameter independently using the physical parameters from Table 4. For this calculation, we used a n_e/n_{H} ratio of 1.4×10^{-4} (Snow & Witt 1995). The results of these two calculations are reported in Table 5. The estimates of the ionisation parameter found here are consistent with the ones calculated from the physical parameters. This shows that our method to quantify the ionisation state of PAHs is consistent with the ones of Flagey et al. (2006) and could be applied to any interstellar PAH spectrum.

8. Chemical evolution of the very small dust particles in the observed PDRs

The distribution maps presented in Sect. 5.2 clearly show that VSGs and PAHs⁰/PAHs⁺ are not situated in the same regions of the studied PDRs. Going towards the star in the PDRs, VSGs followed by PAHs and eventually PAHs⁺ (case of NGC 7023-NW) are successively dominant. This evolution of the dust populations across the PDRs tracking the rise in the UV flux suggests that these three populations are chemically linked. The disappearance of VSG emission, while the PAH emission increases, suggests that there is transformation of VSGs into PAHs under the action of the UV field, as proposed by Cesarsky et al. (2000) for Ced 201 and by RJB for ρ Oph-SR3 and NGC 7023-NW PDRs. Here, we report new evidence of this transformation in three regions: Ced 201, NGC 7023-E and Oph-fil PDRs. This is consistent with a scenario in which VSGs could be PAH clusters as suggested by RJB. Interestingly, recent laboratory experiments have shown that coronene clusters can be photo-evaporated into free coronene units (Bréchignac et al. 2005). PAH aggregates would likely form in cold dense regions (Boulanger et al. 1990; Bernard et al. 1993; Rapacioli et al. 2006), and their photo-chemistry would start when the UV field is strong enough. Thus, the observed variations of the

mid-IR interstellar spectra (Peeters et al. 2002; Werner et al. 2004; Bregman & Temi 2005) have to be explained as the evolution of a mixture of PAHs and VSGs under the effect of UV flux. The processing of VSGs can explain the diminution of continuum emission while approaching the irradiating star reported by Werner et al. (2004). Bregman & Temi (2005) have found that the 7.7 μm emission band shifts towards shorter wavelengths as the UV field increases, which they attributed to the evolution from anionic to cationic species. According to our scenario, this shift can be explained as the chemical evolution of the mixture, from VSGs, with a band between 7.74 and 7.87 μm , into PAHs with a band between 7.63 and 7.70 μm . In that previous work, the authors show that the $R_{7.7/11.3}$ ratio is increasing when moving away from the exciting source, tracing the ionisation state. However, after a certain distance, the ratio reaches its maximum and starts decreasing again. This was explained by the authors as a result of a greater proportion of anionic PAHs, but is more likely consistent with the presence of VSGs which exhibit a low $R_{7.7/11.3}$ band ratio (see Tables 1, 2).

9. Conclusion

Using the data from the Infrared Spectrograph onboard *Spitzer*, combined with powerful Blind Signal Separation methods, we were able to extract the spectra of two types of very small interstellar dust particles: one carrying mainly the AIB features and the other a mid-IR continuum and broad AIBs. These two populations are identified as PAHs and VSGs respectively. Concerning the pure AIB spectra, we could identify in the case of NGC 7023-NW the emission from PAH⁰ and PAH⁺ dominated populations. Using the spectra of RJB, we were able to simply estimate the ionisation fraction of PAHs in various PDRs. Cesarsky et al. (2000) proposed that VSGs should be found everywhere in the interstellar medium. However their detection is far from being obvious. Here we have shown that BSS methods enable us to extract these spectra taking advantage of both the spatial and spectral information available in the IRS spectral cubes. Thanks to the wide range of wavelengths covered, it was possible to confirm that these grains are responsible for the interstellar mid-IR continuum and could dominate the emission up to 50 μm in cool PDRs. The similarities of their spectral features with the AIBs show that they are carbonaceous. This is consistent with the predictions of Désert et al. (1986) that VSGs are mostly graphitic and not silicates. The distribution maps of PAHs

and VSGs further support this conclusion. We indeed show that there is a transformation of VSGs into PAHs, probably under the effect of the UV flux. PAHs and VSGs clearly dominate the emission in the 12 and 25 μm IRAS bands respectively. This result is consistent with the work of Fuente et al. (1992) who proposed that the 25 μm emission in reflection nebulae was due to another type of grains than the ones emitting at 12 μm . It is also consistent with the results of Verter et al. (2000), who have shown that the cloud-to-cloud variations of the mid-IR emission of molecular cirrus can only be explained by changes in the relative abundance of PAHs and VSGs. The evolution from VSGs to PAH molecules can explain the commonly observed mid-IR spectral variations. The recently available data from the MIPS photometer in the Spectral Energy Distribution mode (MIPS-SED) is currently under analysis in order to learn more about the properties of VSGs at longer wavelengths. The PAH⁰, PAH⁺ and VSGs spectra extracted here can be used as a basis to probe the composition of the very small particles in various environments, including external galaxies which could provide information on the local physical conditions.

Acknowledgements. The authors wish to acknowledge all the members of the SPECPRD team for their contribution to the success of the proposal.

References

- Abergel, A., André, P., Bacmann, A., et al. 1999, in *ESA SP-427: The Universe as Seen by ISO*, ed. P. Cox, & M. Kessler, 615
- Abergel, A., Teyssier, D., Bernard, J. P., et al. 2003, *A&A*, 410, 577
- Allamandola, L. J., Tielens, A. G. G. M., & Barker, J. R. 1985, *ApJ*, 290, L25
- Bernard, J. P., Boulanger, F., & Puget, J. L. 1993, *A&A*, 277, 609
- Boissel, P., Joblin, C., & Pernot, P. 2001, *A&A*, 373, L5
- Boulanger, F., Falgarone, E., Puget, J. L., & Helou, G. 1990, *ApJ*, 364, 136
- Bréchinac, P., Schmidt, M., Masson, A., et al. 2005, *A&A*, 442, 239
- Bregman, J., & Temi, P. 2005, *ApJ*, 621, 831
- Casey, S. C. 1991, *ApJ*, 371, 183
- Cesarsky, D., Lequeux, J., Abergel, A., et al. 1996, *A&A*, 315, L305
- Cesarsky, D., Lequeux, J., Ryter, C., & Gérin, M. 2000, *A&A*, 354, L87
- Compiègne, M., Abergel, A., Verstraete, L., et al. 2007, *A&A*, in press
- Delfosse, N., & Loubaton, P. 1995, *Signal Processing*, 45, 59
- Désert, F. X., Boulanger, F., Léger, A., Puget, J. L., & Sellgren, K. 1986, *A&A*, 159, 328
- Désert, F.-X., Boulanger, F., & Puget, J. L. 1990, *A&A*, 237, 215
- Draine, B. T., & Li, A. 2001, *ApJ*, 551, 807
- Flagey, N., Boulanger, F., Verstraete, L., et al. 2006, *A&A*, 453, 969
- Forni, O., Poulet, F., Bibring, J.-P., et al. 2005, in *36th Annual Lunar and Planetary Science Conference*, 1623
- Fuente, A., Martin-Pintado, J., Cernicharo, J., Brouillet, N., & Duvert, G. 1992, *A&A*, 260, 341
- Funaro, M., Erkki, O., & Valpola, H. 2003, *Neural Networks*, 16, 469
- Gobinet, A., Elhafid, A., Vrabie, V., Huez, R., & Nuzillard, D. 2005, in *Proceedings of the 13th European Signal processing Conference*
- Habart, E., Boulanger, F., Verstraete, L., et al. 2003, *A&A*, 397, 623
- Habing, H. J. 1968, *Bull. Astron. Inst. Netherlands*, 19, 421
- Houck, J. R., Roellig, T. L., van Cleve, J., et al. 2004, *ApJS*, 154, 18
- Hyvarinen, A. 1999, *IEEE Transactions on Neural Networks*, 10, 626
- Hyvarinen, A., Karhunen, J., & Oja, E. 2001, in *Independent Component Analysis* (Wiley)
- Joblin, C., Abergel, A., Bernard, J.-P., et al. 2005, in *IAU Symp.*, 153
- Joblin, C., Tielens, A. G. G. M., Geballe, T. R., & Wooden, D. H. 1996, *ApJ*, 460, L119
- Joblin et al. 2007, in prep.
- Kessler, M. F., Steinz, J. A., Anderegg, M. E., et al. 1996, *A&A*, 315, L27
- Kurucz, R. L. 1991, *BAAS*, 23, 1047
- Le Bourlot, J., Pineau Des Forets, G., Roueff, E., & Flower, D. R. 1993, *A&A*, 267, 233
- Le Petit, F., Nehmé, C., Le Bourlot, J., & Roueff, E. 2006, *ApJS*, 164, 506
- Lee, D. D., & Seung, H. S. 1999, *Nature*, 401, 788
- Lee, D. D., & Seung, H. S. 2001, in *NIPS*, ed. MIT press, Vol. 13, 556
- Léger, A., & Puget, J. L. 1984, *A&A*, 137, L5
- Maino, D., Farusi, A., Baccigalupi, C., et al. 2002, *MNRAS*, 334, 53
- Mathis, J. S. 1990, *ARA&A*, 28, 37
- Peeters, E., Hony, S., Van Kerckhoven, C., et al. 2002, *A&A*, 390, 1089
- Rapacioli, M., Calvo, F., Joblin, C., et al. 2006, *A&A*, 460, 519
- Rapacioli, M., Joblin, C., & Boissel, P. 2005, *A&A*, 429, 193
- Sajda, P., Du, S., Brown, T. R., et al. 2004, *IEEE Trans. Medical Imaging*, 23, 1453
- Siebenmorgen, R., Kruegel, E., & Mathis, J. S. 1992, *A&A*, 266, 501
- Sloan, G. C., Hayward, T. L., Allamandola, L. J., et al. 1999, *ApJ*, 513, L65
- Smith, J. D. T., Dale, D. A., Armus, L., et al. 2004, *ApJS*, 154, 199
- Smith, J. D. T., Draine, B. T., Dale, D. A., et al. 2007, *ApJ*, 656, 770
- Snow, T. P., & Witt, A. N. 1995, *Science*, 270, 1455
- van den Ancker, M. E., The, P. S., Tjin A Djie, H. R. E., et al. 1997, *A&A*, 324, L33
- Verter, F., Magnani, L., Dwek, E., & Rickard, L. J. 2000, *ApJ*, 536, 831
- Werner, M. W., Uchida, K. I., Sellgren, K., et al. 2004, *ApJS*, 154, 309
- Witt, A. N., Graff, S. M., Bohlin, R. C., & Stecher, T. P. 1987, *ApJ*, 321, 912
- Young Owl, R. C., Meixner, M. M., Fong, D., et al. 2002, *ApJ*, 578, 885
- Zubko, V., Dwek, E., & Arendt, R. G. 2004, *ApJS*, 152, 211