

# Coupled blind signal separation and spectroscopic database fitting of the mid infrared PAH features<sup>★</sup>

M. J. F. Rosenberg<sup>1,2</sup>, O. Berné<sup>1</sup>, C. Boersma<sup>3</sup>, L. J. Allamandola<sup>3</sup>, and A. G. G. M Tielens<sup>1</sup>

<sup>1</sup> Sterrewacht Leiden, Universiteit Leiden, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands  
e-mail: rosenberg@strw.leidenuniv.nl

<sup>2</sup> The International Space University, Parc d'Innovation, 1 rue Jean Dominique Cassini, 67400 Illkirch Graffenstaden, France

<sup>3</sup> NASA Ames Research Center, Space Science Division, Mail Stop 245-6, Moffett Field, CA 94035, USA  
e-mail: [Louis.J.Allamandola;Christiaan.Boersma]@nasa.gov

Received 16 December 2010 / Accepted 28 June 2011

## ABSTRACT

**Context.** The aromatic infrared bands (AIBs) observed in the mid infrared spectrum of galactic and extragalactic sources are attributed to polycyclic aromatic hydrocarbons (PAHs). Recently, two new approaches have been developed to analyze the variations of AIBs in terms of chemical evolution of PAH species: blind signal separation (BSS) and the NASA Ames PAH IR Spectroscopic Database fitting tool.

**Aims.** We aim to study AIBs in a photo-dissociation region (PDR) since in these regions, as the radiation environment changes, the evolution of AIBs are observed.

**Methods.** We observe the NGC 7023-north west (NW) PDR in the mid-infrared (10–19.5  $\mu\text{m}$ ) using the InfraRed Spectrometer (IRS), on board *Spitzer*, in the high-resolution, short wavelength mode. Clear variations are observed in the spectra, most notably the ratio of the 11.0 to 11.2  $\mu\text{m}$  bands, the peak position of the 11.2 and 12.0  $\mu\text{m}$  bands, and the degree of asymmetry of the 11.2  $\mu\text{m}$  band. The observed variations appear to change as a function of position within the PDR. We aim to explain these variations by a change in the abundances of the emitting components of the PDR. A blind signal separation (BSS) method, i.e. a Non-Negative Matrix Factorization algorithm is applied to separate the observed spectrum into components. Using the NASA Ames PAH IR Spectroscopic Database, these extracted signals are fit. The observed signals alone were also fit using the database and these components are compared to the BSS components.

**Results.** Three component signals were extracted from the observation using BSS. We attribute the three signals to ionized PAHs, neutral PAHs, and very small grains (VSGs). The fit of the BSS extracted spectra with the PAH database further confirms the attribution to PAH<sup>+</sup> and PAH<sup>0</sup> and provides confidence in both methods for producing reliable results.

**Conclusions.** The 11.0  $\mu\text{m}$  feature is attributed to PAH<sup>+</sup> while the 11.2  $\mu\text{m}$  band is attributed to PAH<sup>0</sup>. The VSG signal shows a characteristically asymmetric broad feature at 11.3  $\mu\text{m}$  with an extended red wing. By combining the NASA Ames PAH IR Spectroscopic Database fit with the BSS method, the independent results of each method can be confirmed and some limitations of each method are overcome.

**Key words.** astrochemistry – photon-dominated region (PDR) – ISM: lines and bands – infrared: ISM

## 1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) are carbonaceous macromolecules which were postulated to be present in the interstellar medium (ISM) in the 1980s (Léger & Puget 1984; Allamandola et al. 1985; Puget & Léger 1989; Allamandola et al. 1989) and have since undergone an intense investigation in astronomy. The state of the art and recent activity in the field of interstellar PAHs is well illustrated by the book “PAHs and the Universe”, Joblin & Tielens (2011). Astronomical PAHs are generally considered to contain roughly 50–100 C atoms and have an abundance of a few  $10^{-7}$  per H atom (Tielens 2008). Because of their nanometer size, the absorption of one far-ultraviolet (FUV) photon is sufficient to heat PAH molecules to high temperatures causing them to emit characteristic bands called aromatic infrared bands (AIBs) which peak near 3.3, 6.2,

7.7, 8.6, and 11.2  $\mu\text{m}$  (Tielens 2008 and references therein). PAHs are abundantly present in the diffuse ISM, reflection nebulae (RNe), planetary nebulae, protoplanetary disks, and galaxies. Observations of PAHs in Photo-Dissociation Regions (PDRs, Hollenbach & Tielens 1999), which are transition regions between atomic and molecular gas, but still strongly affected by the FUV photons, are of particular interest. The UV flux decreases when moving from the neutral atomic gas to the dense molecular cloud and PAH populations will also evolve as the UV flux changes (Sloan et al. 1999; Joblin et al. 1996). This effect is best studied in the mid-infrared (5–15  $\mu\text{m}$ ), where PAHs emit most strongly.

Each AIB is characteristic of a PAH vibrational mode (e.g., Puget & Léger 1989; Allamandola et al. 1985; Hony et al. 2001), the 3.3, 8.6, and 10–15  $\mu\text{m}$  features are due to the C-H stretching, in-plane and out-of-plane bending modes while the 6.2 and 7.7  $\mu\text{m}$  features are mainly due to the C-C stretching modes. These features have been observed to show strong variation in peak position, width of band (FWHM), and symmetry

<sup>★</sup> 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.

(Peeters et al. 2002). The changing ratio of the 8.6 and 11.3  $\mu\text{m}$  features was discovered first and attributed to a change in the relative abundances of neutral and ionized PAHs (Joblin et al. 1996; Sloan et al. 1999). Soon after, the 7.7/11.3  $\mu\text{m}$  ratio was observed to vary as well, which was also attributed to the charge state of the PAHs. Later, using Infrared Space Observatory (ISO), Peeters et al. (2002) catalogued the variations of the main features in the 6–9  $\mu\text{m}$  range and empirically divided them into groups based on specific spectral properties. It was found that each of these groups was representative of certain classes of objects: class A included HII regions, RNe, galaxies, and non-isolated Herbig stars. Class B included isolated Herbig stars, PNe, and two post AGB stars, while class C included only two post AGB stars. Observations of the 10–15  $\mu\text{m}$  region have been analyzed by Hony et al. (2001) in terms of the solo, duo, trio, or quartet out-of-plane (OOP) bending mode of either PAH<sup>0</sup>, PAH<sup>+</sup>, or some combination of the two. While the position and profile of these bands are quite characteristic the relative intensities do vary a lot, indicating variations in the edge structure of the aromatic molecules.

Analysis and interpretation of astronomical observations is supported by dedicated laboratory studies and quantum chemical studies. These studies are being carried out in many groups around the world, each using a different technique (see Oomens 2011 for a review of laboratory and experimental studies). In most cases, the absorption is studied at a low temperature in an inert matrix. There are also some gas-phase experiments that have been carried out at higher temperatures. These experimental studies have been extended by quantum chemical calculations using density functional theory (DFT), to species not accessible in laboratory studies. The DFT approach is used to determine the frequencies and intensities of vibrational modes. Recently, these models have been used to calculate spectra for PAHs from 54 to 130 C atoms (Bauschlicher et al. 2008, 2009). This size range is particularly relevant for comparison to observations of space-based PAHs. An extensive database has been created by the Astrophysics and Astrochemistry Laboratory at NASA Ames, which includes mid-IR to far-IR spectra of many different PAHs including large molecules, varied levels of ionization, and irregular shapes (Bauschlicher et al. 2010). Mallocci et al. (2007) used time dependent DFT methods and quantum-chemical calculations to report computed molecular properties of PAH emission for 40 molecules, available on an online database. Mulas et al. (2006a) then modeled the PAH emission, which give positions and intensities of specific PAHs in different radiation environments. The band profiles of some PAH emission were also calculated by Mulas et al. (2006b).

Recently, the improved sensitivity of *Spitzer* Space Telescope has brought a wealth of observations of AIB features. NGC 7023 is a well studied and bright IR source where PAH variations are known to occur (Cesarsky et al. 1996). Werner et al. (2004) and Sellgren et al. (2007) used *Spitzer* Infrared Spectrograph (IRS) in the short-high (SH), short-low (SL), and high-low (HL) modes to further observe the full mid-infrared range of NGC 7023. They observed all the classical AIBs above 5  $\mu\text{m}$  in addition to finding new, weak emission features at 6.7, 10.1, 15.8, 17.4, and 19.0  $\mu\text{m}$ . Berné et al. (2007) observed NGC 7023, along with three other PDRs, using *Spitzer* IRS-SL. The spectra were analyzed using a class of methods called blind signal separation (BSS), which identifies elementary spectra from spectral cubes. Using BSS on spectra of NGC 7023-NW, three component signals were recognized, PAH cations, neutral PAHs, and a third carrier which Berné et al. (2007) attributed to evaporating very small grains (VSG), following

earlier assignment by Rapacioli et al. (2005). Although it is yet unclear what the exact nature of VSGs are, it has been proposed that they could be PAH clusters (Rapacioli et al. 2005).

This paper presents a study of the PDR NGC 7023-NW, which aims to put observational constraints on the origins of the profiles and variations of the 10–15  $\mu\text{m}$  spectra. This study will complement and restrict previous results from quantum chemical calculations, ISO spectroscopy observations, *Spitzer* IRS observations, and PAH models. We analyze high resolution data from *Spitzer* IRS-SH (Werner et al. 2004), at a resolution,  $R = \lambda/\Delta\lambda = 600$ , using the NASA Ames PAH IR Spectroscopic Database and a BSS method to separate the emitting components of the PDR.

After briefly outlining the observational methods in Sect. 2, the blind signal separation method is described in Sect. 3, including the application of BSS to the data (Sect. 3.2). Section 4 presents our main results and compares these with previous studies of the region. Section 5 provides a comparison to the fit with the NASA Ames PAH IR Spectroscopic Database. Next, in Sect. 6, we discuss the implications of our results and propose strong candidates to explain the spectral variations of the 10–15  $\mu\text{m}$  region. Section 7 discusses briefly the nature of the VSG carrier and section 8 gives our concluding remarks.

## 2. Observations

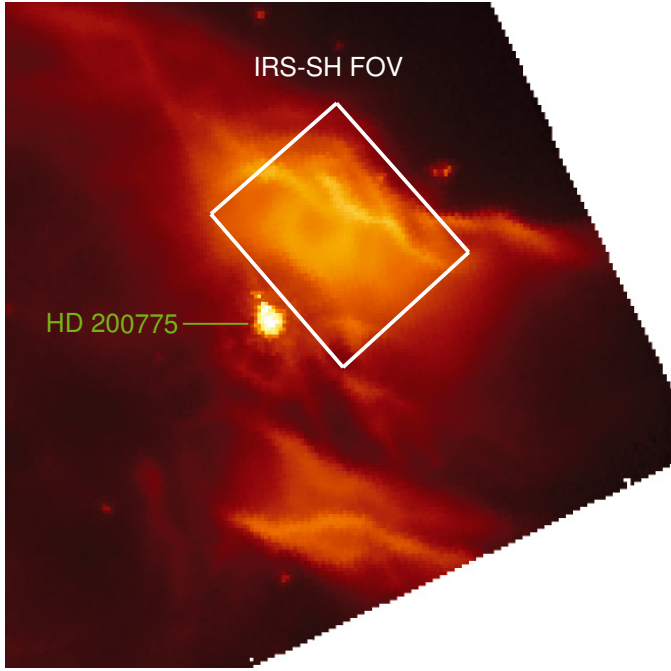
NGC 7023-NW is a PDR located 40'' to the northwest of the exciting star, HD 200775, seen in Fig. 1. HD 200775 is a magnitude 7 Herbig Be star and is located 430 pc from the sun (van den Ancker et al. 1997). There are 3 PDRs in NGC 7023 located east, south and northwest of the exciting star. The northwest PDR is the brightest of the 3 PDRs.

NGC-7023-NW was chosen for this study in view of the interesting results of the analysis of low resolution data of this region performed by Berné et al. (2007) who have shown that the AIB spectrum can be separated into 3 main components. To further this study, high-resolution short wavelength (IRS = SH, 10–19.5  $\mu\text{m}$ ) observations were obtained from the *Spitzer* Heritage Archive (SHA). The data was reduced using the CUBE Builder for IRS Spectra Maps (CUBISM), provided by NASA's *Spitzer* Space Telescope tools (Smith et al. 2007).

## 3. Methods

### 3.1. Blind signal separation

Blind signal separation (BSS) is a class of methods used in many scientific fields to separate source signals from observed linear combinations of these signals e.g., separating brainwaves or unmixing recordings in acoustics. This has only recently been applied to astronomy (Nuzillard & Bijaoui 2000; Berné et al. 2007). In the case of observing a PDR, the resultant spectrum at any spatial point is a superposition of all elements emitting in the designated wavelength range and BSS can separate these components. There are three main methods to perform BSS: independent component analysis (ICA), non-negative matrix factorization (NMF), and sparse component analysis (SCA). Based on the results of Berné et al. (2007) and the added spectral resolution of the data, the NMF method was selected to perform the analysis. If  $X$  is a matrix containing the observed spectra, assuming that each spectrum is the result of a linear combination of source signals, we can write  $X \approx WH$ , where  $H$  is the matrix of source signals, and  $W$  is the matrix of mixing coefficients. The goal of NMF is to recover  $W$  and  $H$  based on  $X$  only. The



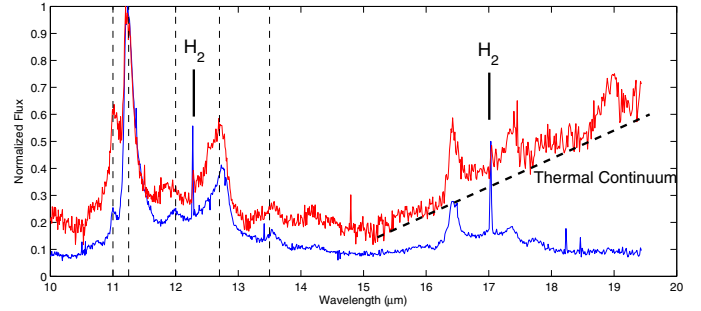
**Fig. 1.** IRAC 8  $\mu\text{m}$  image of NGC 7023 (Werner et al. 2004). Highlighted with a white box is the IRS-SH field of view around the PDR NGC 7023-NW. The star, HD 200775, is marked by the green label.

method we apply here is the same as Berné et al. (2007) where all details can be found.

### 3.2. Application to NGC 7023

Our astronomical data is comprised of 750 spectra, each taken at a different spatial location, and each spectrum including 869 wavelength points. Figure 2 shows two spectra, before they are decomposed with BSS. There is a clearly defined separation between the 11.0 and 11.2  $\mu\text{m}$  emission features as well as distinct features at 12.0, 12.7, 13.5, and 14.1  $\mu\text{m}$  with additional features at longer wavelengths. Among the features at longer wavelengths are the PAH 16.4  $\mu\text{m}$  feature, the blended PAH and  $\text{C}_{60}$  feature at 17.5  $\mu\text{m}$ , and the pure  $\text{C}_{60}$  feature at 18.9  $\mu\text{m}$ , which are further discussed in Sellgren et al. (2010). The relative intensities of the  $\text{H}_2$  lines vary with position in the nebula, representing a non-linear component in our spectra. Another non-linear aspect of the spectra is the onset of the dust grain continuum caused by heating from the source star. These non-linear components cannot be analyzed by BSS methods, since the method demands a linear combination of signals. Therefore, we have chosen to exclude the emission at wavelengths greater than 15  $\mu\text{m}$  where the continuum from dust grains is present. We have also clipped the  $\text{H}_2$  lines everywhere in the cube by hand and replaced them by a linear interpolation. The resulting spectra were then analyzed using the NMF algorithm from Lee & Seung (2001) with both divergence and Euclidian distance optimization, see Berné et al. (2007) for details.

We investigated the possibility of 3, 4, 5, and 6 component source signals. Irrespective of the particular minimization technique, when attempting to separate 4, 5 and 6 sources, there are 2 or more signals which are very similar (linear combinations of each other) and at least one signal that is pure noise. Therefore, we can conclude that there are 3 significantly different spectral components responsible for the AIBs in NGC 7023-NW. This



**Fig. 2.** Observed spectra in pixel position [15, 28] (blue) and [23, 17] (red), containing  $\text{H}_2$  lines and thermal continuum.

result confirms the findings of Berné et al. (2007) that there are only 3 source signals in this PDR.

## 4. Results of blind signal separation

### 4.1. Extracted source signals

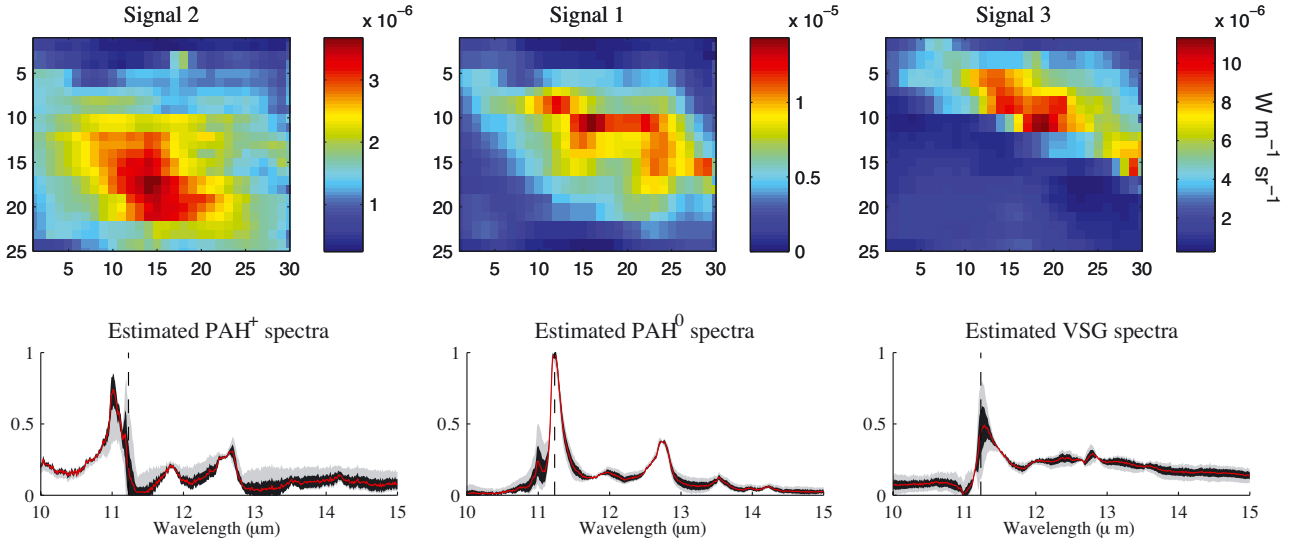
The final extracted spectra are shown in Fig. 3. To increase confidence in these results, and ensure that this solution is not a random local minimum, the same analysis was repeated 100 times using different random initializations. These 100 spectra shared the same general shape, but varied in intensity, especially in the 11.0–11.3  $\mu\text{m}$  region. The average spectra of the 100 iterations is plotted with a red line in Fig. 3, and will be used for the remainder of our analysis as the final BSS extracted signals ( $H$  matrix). We can also estimate the error at each point in the spectrum using these results (Fig. 3). The BSS method has the most difficulty separating the signals in the 11.0 to 11.3  $\mu\text{m}$  range due to the strong changing spectral gradients there. This results in large errors in this range (see Appendix A for discussion on unmixing artifacts). Since  $X \approx WH$ , we can estimate  $W$  by minimizing  $\|X - WH\|$  using a standard least squares minimization. Figure 4 compares the observations in  $X$  and the final reconstruction of these observations with  $W \times H$ . The reconstruction is in good agreement with the observations.

Using the weighting factors that come as a resultant matrix of the above reconstruction ( $W$ ), we can map the spatial distribution of each source signal separately (Fig. 3). The spatial distribution shows clear variation for the three emitting components. Signal 1 is most abundant in the middle of the PDR. Signal 2 has its highest concentration closest to the source star (located at the bottom left of this image) and Signal 3 appears to trace the edge of the PDR farthest from the star. The well defined regions where each signal is most concentrated implies a physical cause and gives further confidence that these results are not random.

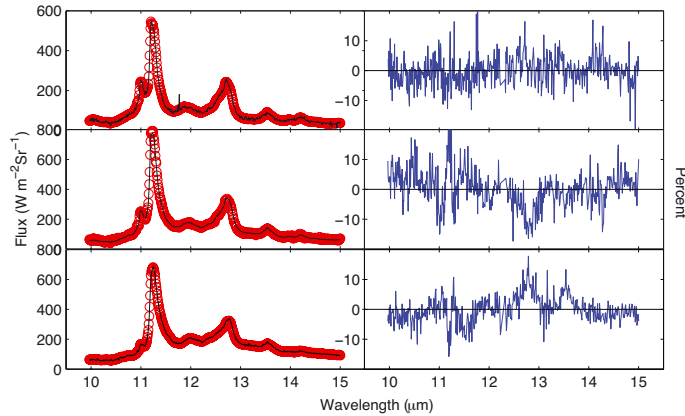
### 4.2. Carriers of the extracted spectra

In this section, we will compare our results to the results of the low resolution study of the same region (Berné et al. 2007, 2010) to gain insight about the three extracted signals. Creating spatial contours of intensity for each signal allows us to compare the spatial distribution of our signals to the distribution of the three signals from the study of Berné et al. (2007) of the 5–15  $\mu\text{m}$  low resolution spectra. The contours are created from the IRS-SH spatial distribution maps (Fig. 3) and overlaid with the spatial distributions (represented in color) of the IRS-SL results (Fig. 5). The three signals extracted here show a strong spatial correlation to the  $\text{PAH}^+$ ,  $\text{PAH}^0$ , and VSG maps of Berné et al. (2007). The





**Fig. 3.** *Bottom panels:* extracted spectra using NMF, normalized at  $11.2 \mu\text{m}$ . The vertical line represents the peak position of the estimated  $\text{PAH}^0$  spectra. The red line represents the average spectra out of 100 iterations. The grey envelope shows the minimum and maximum spectra and the black envelope shows the  $1\text{-}\sigma$  error of the 100 iterations. *Upper panels:* spatial distributions of the weighting factors obtained by Least Squares Fitting of the observed spectra in the datacube using the BSS extracted spectra shown in red in the lower panel.



**Fig. 4.** Spectra taken at three random spatial positions (*left*). The solid line represents the original observed spectra while the red overlapping circles (thick red line) represent a linear combination of the BSS extracted spectra. On the *right* is the matching residuals for each plot.

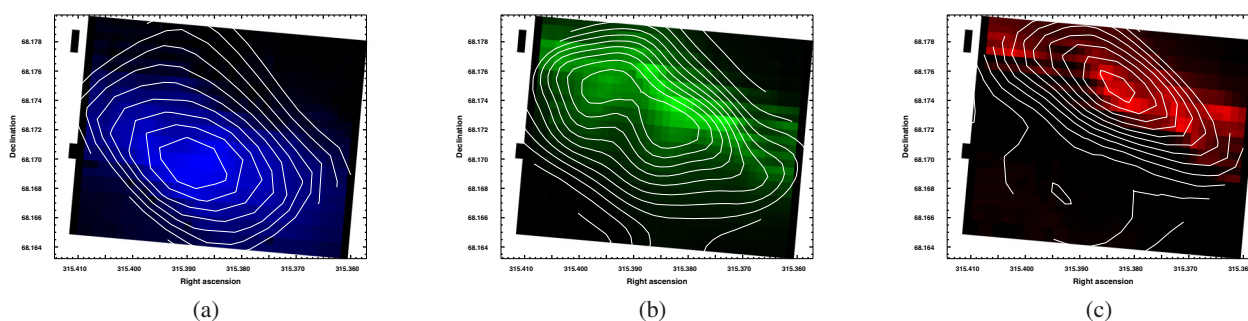
spatial distribution and the results of [Berné et al. \(2007\)](#) seem to suggest that Signal 2 traces the distribution of PAH cations, Signal 1 the neutral PAH distribution, and Signal 3 the distribution of VSGs. Although the spatial distributions of Signal 1, Signal 2, and Signal 3 correlate well with  $\text{PAH}^0$ ,  $\text{PAH}^+$ , and VSGs of [Berné et al. \(2007\)](#), there are some small discrepancies, in particular, for the  $\text{PAH}^0$  map. As discussed in [Berné et al. \(2010\)](#), the degradation of spatial or spectral resolution always implies a loss in the quality of the NMF efficiency. Since [Berné et al. \(2007\)](#) have a higher spatial resolution, while here we have a higher spectral resolution, none of the data-sets can be considered “better” and small discrepancies between the results of NMF are expected.

Figure 6 compares the low-resolution source signal spectra of [Berné et al. \(2007\)](#) to the high-resolution source spectra obtained here (Fig. 3). The low resolution extracted spectra share all the major features with the high resolution spectra, specifically the broad  $11.3 \mu\text{m}$  emission feature in the VSGs, the  $11.2 \mu\text{m}$  and  $12.7 \mu\text{m}$  emission features in the neutral PAHs, and

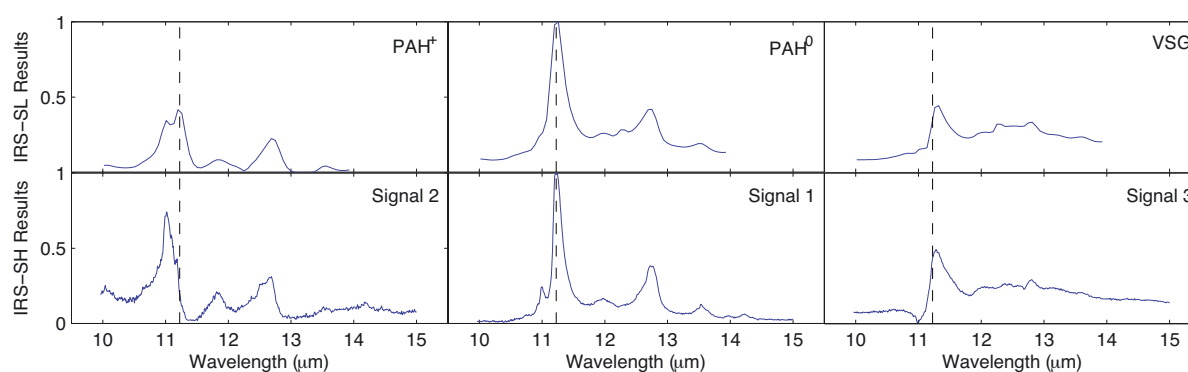
the  $11.0 \mu\text{m}$  and broad  $12.7 \mu\text{m}$  emission features in the PAH cations. Although evidence for the  $11.0 \mu\text{m}$  feature was present in the IRS-SL observations of [Berné et al. \(2010\)](#), it was not immediately attributed to PAH cations since the resolution was not high enough to fully resolve and separate the  $11.0$  and  $11.2 \mu\text{m}$  features. The IRS-SH spectra has not only validated the previous results of the IRS-SL observations but given us additional spectral detail, to infer more about each signal and how it contributes to the observed spectra. This is of particular interest in the  $11 \mu\text{m}$  region where the  $11.0 \mu\text{m}$  and  $11.2 \mu\text{m}$  bands are well resolved and isolated. We also observe a consistent broadening of the  $11.3 \mu\text{m}$  feature in the VSG spectra. One main difference in the spectra is the presence of an  $11.0 \mu\text{m}$  satellite feature in Signal 1 that is not found in the IRS-SL  $\text{PAH}^0$  signal, which we suggest is an artifact caused by the inherent limitation of unmixing and is compensated for by the sharp drop at  $11.0 \mu\text{m}$  of Signal 3. The strong correlation between both the spectra and spatial distributions allow us to confidently identify Signal 1 as a  $\text{PAH}^0$  signal, Signal 2 as a  $\text{PAH}^+$ , and Signal 3 as a genuine VSG spectrum. This also matches the physical description of the PDR with the ionized species closest to the source star.

## 5. Comparison with database fitting analysis

The BSS-method allows the separation of the observed spectra into three mathematically distinct components, without taking the actual physical or spectroscopic properties of aromatic species into account. To further explore the carriers of the three signals resulting from the BSS analysis, we turned to the NASA Ames PAH IR Spectroscopic Database ([Bauschlicher et al. 2010](#)), which contains over 600 theoretical and experimental PAH spectra. The existing extensive database of PAH spectra allows us to approach the analysis of the astronomical data from a different perspective. Specifically, it allows us to link observational properties of the infrared emission features to the molecular characteristics of the carriers. To that end, we first fit each BSS extracted spectrum with the PAH database. Second, we focus on one observed spectrum from the IRS-SH data cube and use the database to fit this spectrum. In interpreting the results



**Fig. 5.** a) The contours of Signal 2, overlaid with the distribution of PAH cations created from [Berné et al. \(2007\)](#). b) The contours of Signal 1, overlaid with the distribution of neutral PAHs created from [Berné et al. \(2007\)](#). c) The contours of Signal 3, overlaid with the distribution of VSGs created from [Berné et al. \(2007\)](#).



**Fig. 6.** The *top* spectra are the results from [Berné et al. \(2010\)](#) using low-resolution IRS data, the *bottom* row of spectra are the current results, using high-resolution IRS. The vertical dashed black line indicates the 11.2  $\mu\text{m}$  line position.

from this database analysis, it should be kept in mind that vibrational modes in the mid-IR spectral range are characteristic for molecular groups and are not very sensitive to individual molecules. Hence, the goal of this database analysis is to provide insights in trends rather than specific molecular identifications. The trends of interest here involved the effects of charge, size, molecular geometry and symmetry – including the degree of compactness of the PAH families – on the different spectral components. As a corollary, the completeness of the database is therefore of lesser concern, as long as the relevant classes are well represented.

The BSS method fits are “blind” in the sense that there is no a priori information about the nature of the signals built into the method. On the other hand, the fits of the database are based on spectra of actual aromatic molecules in specific charge states, structures, and sizes, allowing for a more direct interpretation of the results.

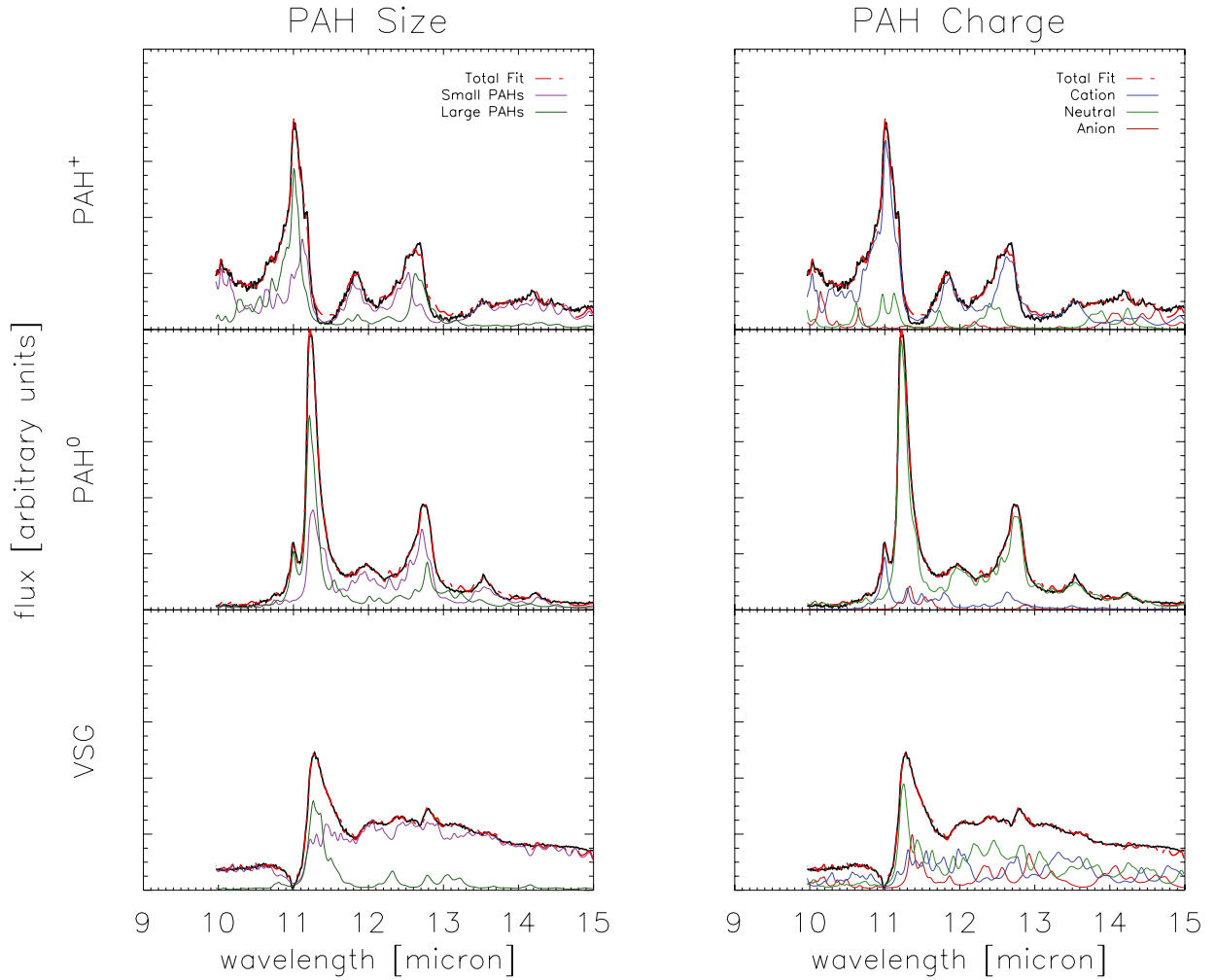
The BSS results separated 3 mathematically distinct emitting signals and based on comparison with the signals extracted from [Berné et al. \(2007\)](#), we have attributed the three signals to PAH<sup>+</sup>, PAH<sup>0</sup>, and VSGs. However, there is no aspect of the BSS method that actually identifies the signals as ionized or neutral PAH species. By fitting these signals with the database, we can obtain an independent attribution of these distinct signals as PAH classes.

### 5.1. Fit parameters

The database, at all versions, and the AmesPAHdbIDL Suite can be obtained from [www.astrochem.org/pahdb](http://www.astrochem.org/pahdb). Here, version 1.11

of the theoretical component of the database and the November 10, 2010 version of the IDL suite was used. Briefly, the spectra in the theoretical database correspond to about 600 different PAHs, ranging in size from C<sub>9</sub>H<sub>7</sub> to C<sub>130</sub>H<sub>28</sub>. Note that the database is biased towards smaller species, with PAHs containing over 50 carbon atoms making up roughly 24%. The database includes PAHs at different charge states (i.e. cations, anions and neutral species) as well as different symmetries of the same molecule. The fit included all C-H PAHs as well as polycyclic aromatic nitrogen heterocycles (PANH’s; PAHs with one or more nitrogen atoms substituted into their carbon skeleton), since [Hudgins et al. \(2005\)](#) suggested that at least 1.2% of the cosmic nitrogen is tied up in PAH molecules. The fit excludes PAHs with oxygen, magnesium, or iron, where we note that no specific spectral evidence for the existence of such species has been found yet.

Two approximations are made when fitting “observed” spectra with the database. First, the spectra in the database refer to absorption spectra at 0 K, while the observed spectra are emitted by “hot” species. Due to anharmonicity, emission bands are observed to shift to the red with increasing temperature ([Cherchneff et al. 1992](#); [Cook & Saykally 1998](#)). Systematic experimental and quantum chemical studies on a very limited set of PAHs show that this redshift depends on the mode under consideration, the molecular structure and the temperature ([Joblin et al. 1995](#); [Oomens et al. 2003](#); [Basire et al. 2010](#)). The out-of-plane bending modes are observed to shift by about 15 to 20 cm<sup>-1</sup> for the small PAHs, pyrene and coronene between 0 and 900 K and we will adopt this value for all PAHs in the database. Second, the database tools allow the user to specify a Gaussian or a Lorentzian profile with an assumed linewidth. However, the



**Fig. 7.** The fit from NASA Ames PAH IR Spectroscopic Database (Bauschlicher et al. 2010) of the three components extracted from BSS, PAH<sup>+</sup> (top), PAH<sup>0</sup> (middle), and VSG (bottom). In the left column, we compare large PAH contribution to small PAH contribution. In the right column we compare cation, neutral, and anion contribution to the fit.

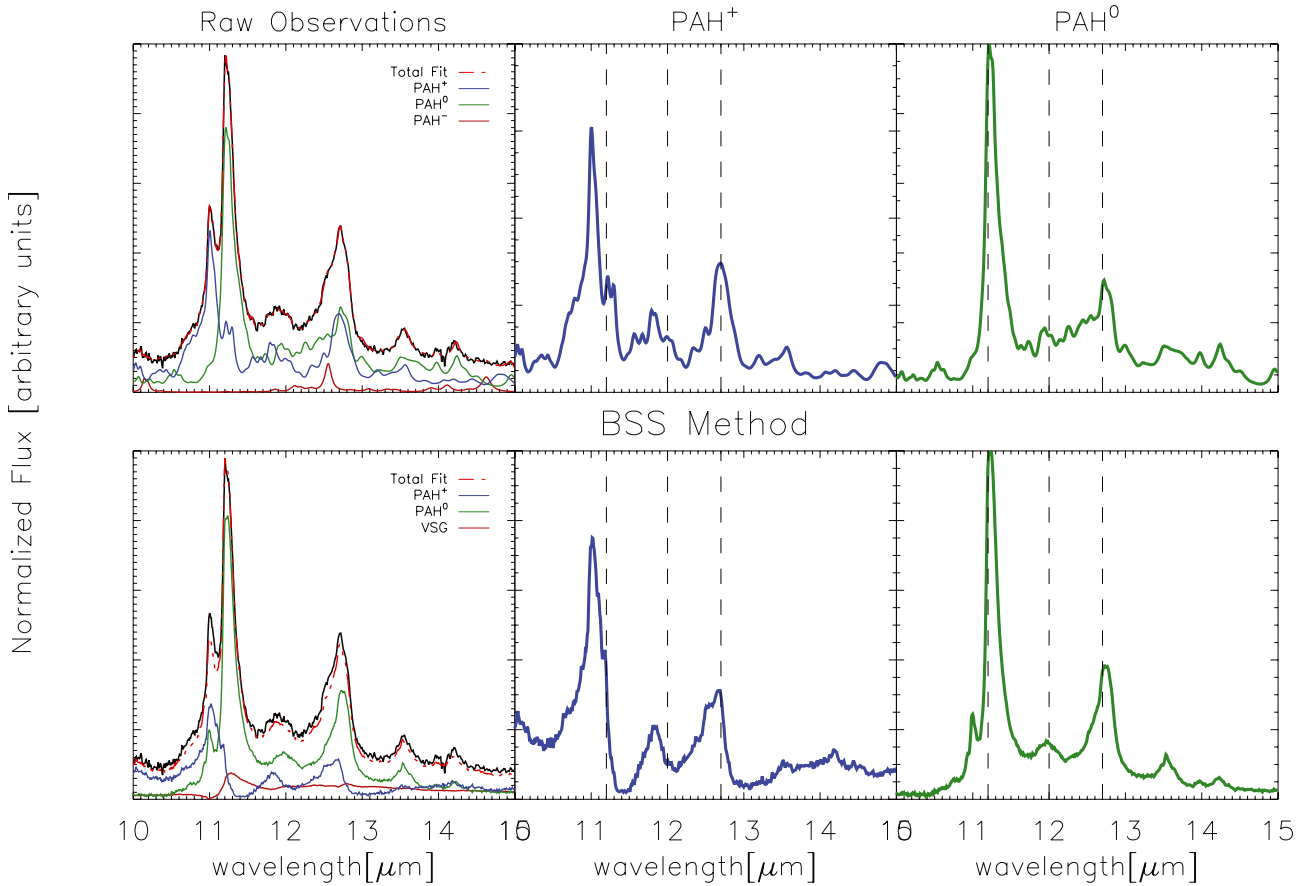
intrinsic line profile of PAH emission bands is distinctly non-Gaussian and non-Lorentzian due to the effects of anharmonicity and the accompanying red-shading (Barker et al. 1987; Pech et al. 2002). Here, we will adopt Lorentzian profiles, fully realizing that this implies that this procedure will force the fit of the observed broader and red-shaded bands to blends by emission from multiple species. We will adopt an intrinsic Lorentzian linewidth of  $6 \text{ cm}^{-1}$  for the out-of-plane bending modes and note that this is somewhat less than the measurements ( $\sim 10 \text{ cm}^{-1}$ ) at  $\sim 700 \text{ K}$  for pyrene and coronene (Joblin et al. 1995). However, our “choice” to fit the observed, inherently asymmetric line profiles of the out-of-plane bending modes with symmetric calculated profiles forces us to adopt a somewhat small intrinsic line profile. We will assess the effect of these assumptions on our fitting results later on. Lastly, we mention that the relative strength of the bands in the calculated spectrum will also depend on the internal energy (e.g., temperature) of the emitting species and hence on the absorbed photon energy and the temperature cascade. However, over the limited wavelength range considered here, this effect is very small and we will here simply adopt a single absorbed UV photon energy (6.5 eV) characteristic for the benign conditions of the PDR in NGC 7023

(Joblin & Mulas 2009). This corresponds to a peak temperature of  $\sim 900 \text{ K}$  for a 50 C-atom PAH. The database evaluates the temperature for each PAH and follows the temperature cascade consistently. Although our wavelength range is limited, we choose to use the temperature cascade to represent the most physically accurate approach.

## 5.2. Fit results

The results for the fits of the three extracted BSS component signals are presented in Fig. 7, broken down in categories of PAH size and charge. The small ( $< 50$  C atoms) and the large ( $\geq 50$  C atoms) PAHs, as well as the cation, neutral, and anion species were separated out in order to judge each subgroups contribution to the fit. A fit to the raw observed spectra at a typical position was also made (Fig. 8). The results are summarized in Table 1. It should be emphasized that the contribution of individual species was quantified in terms of emission intensity, not in terms of abundance. Figure 8 shows a comparison between the PAH<sup>+</sup> and PAH<sup>0</sup> contributions of the database fit (top), and the PAH<sup>+</sup> and PAH<sup>0</sup> contributions in the BSS decomposition (bottom). The results of the database fits highlight various trends. First, except for

## PAH Database Fit



**Fig. 8.** Comparison of the fitting results of the two methods, the NASA Ames PAH IR Spectroscopic Database (Bauschlicher et al. 2010) (top) and BSS (bottom). The left panel depicts an observed spectra which is fit with both the database and the BSS extracted spectra. The middle and right panels are the cation (middle) and neutral (right) components obtained from the respective methods. The signals are normalized to the 11.2  $\mu\text{m}$  peak.

**Table 1.** The numerical result of the fit using the PAH Database to fit the component spectra extracted using the BSS method.

Database	BSS components		
	PAH <sup>+</sup>	PAH <sup>0</sup>	VSG
Percentage			
Large PAHs	43	48	15
Small PAHs	57	52	85
PAH cations	78	15	34
PAH neutrals	16	80	21
PAH anions	6	5	45

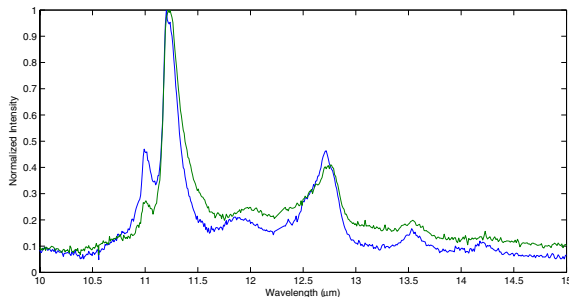
**Notes.** The percentage amounts of large ( $C \geq 50$ ), small ( $C < 50$ ), cation, neutral, and anion species are displayed for the PAH<sup>+</sup>, PAH<sup>0</sup>, and VSG component signals.

the VSG spectrum, the overall contribution of large versus small PAHs seems very even. This may arise from the three times greater number of smaller PAHs in the database. Remarkably, although they represent only 25% of the database, the main features of each spectrum are generated by the larger PAHs while the continuum and less dominant features are reproduced primarily by the smaller PAHs. Second, both the fit of the observed raw spectra shown in Fig. 8 and the various PAH charge state contributions shown in the right side of Fig. 7 clearly shows that the dominant contributors to the 11.0  $\mu\text{m}$  band are PAH<sup>+</sup> while the bulk of the 11.2  $\mu\text{m}$  feature is due to PAH<sup>0</sup> species. This

behavior, based on the spectra of hundreds of PAHs, confirms the early suggestion, based on a handful of experimental PAH spectra, that the emission between 10.8 and 11.1  $\mu\text{m}$  can be used as a tracer of PAH cations (Hudgins & Allamandola 1999). It should also be noted that the BSS extracted VSG signal has two strong absorption features at about 11.0 and 12.6  $\mu\text{m}$ . These probably arise from cross mixing as described earlier and are likely artificial (Appendix A). Therefore, the fit was also performed with a linear interpolation over these points. This fit produced almost identical results.

Common concerns regarding the PAH Database fitting methods are uniqueness and degeneracy. In the case of PAHs with a limited wavelength range, these concerns must be approached in a different way. This is further discussed and investigated in Appendix B. These studies show that if the database is tasked to fit the BSS extracted PAH<sup>+</sup> spectra with only PAH<sup>0</sup> species, this is not possible, further strengthening the attribution of the 11.0  $\mu\text{m}$  feature to PAH<sup>+</sup>. On the other hand, if the database attempts to fit the BSS extracted PAH<sup>0</sup> signal with only PAH<sup>+</sup>, a suitable fit is provided. However, without any limitations of the database fit, the reduced norm fit chooses predominately PAH<sup>+</sup> to fit the BSS extracted PAH<sup>+</sup> signal and PAH<sup>0</sup> to fit the BSS extracted PAH<sup>0</sup> signal (Table 1). This only shows that PAH cations have emission features that peak through the 10–15  $\mu\text{m}$  range and it is important to employ more than one method to apply as many astronomical constraints as possible.





**Fig. 9.** The original spectra from pixel positions (13, 15), where PAH<sup>+</sup> are highly concentrated and (20, 10), where VSGs are abundant highlighting the extremes of the observed spectral variations.

## 6. Origins of AIB variations in the 10–15 μm range

Investigating the spectra and examining the spatial distribution of the extracted emission components, we recognize four aspects that vary with spatial position: the [11.0]/[11.2] μm ratio, the red wing of the 11.2 μm feature, the precise peak position of the 11.2 μm band, and the weak features at 12.0, 12.7, and 13.5 μm. In the following section we discuss these variations in the context of Signal 1, 2, and 3, which will now be referred to as PAH<sup>0</sup>, PAH<sup>+</sup>, and VSGs, as emission feature carriers.

### 6.1. Ratio of 11.0 to 11.2 μm features

In most observations of the 10–15 μm range, the dominant 11.2 μm feature appears with a dwarfed satellite feature at 11.0 μm. This has been attributed to the solo C-H out-of-plane bending modes of PAH<sup>+</sup> (Hudgins & Allamandola 1999; van Diedenhoven et al. 2004; Hony et al. 2001; Bauschlicher et al. 2008, 2009). We observe the highest [11.0]/[11.2] μm ratio closest to the source star, which is also where the abundance of PAH<sup>+</sup> is greatest. A comparison of observed spectra from a PAH<sup>+</sup> dominated region and a VSG dominated region are shown in Fig. 9. It is important to recall that regardless of position in the PDR, the 11.2 μm PAH<sup>0</sup> feature significantly dominates the 11.0 μm PAH<sup>+</sup> feature. The separate contributions of PAH<sup>+</sup> and PAH<sup>0</sup> to the 11.0 and 11.2 μm features respectively, agrees with previous spectroscopic laboratory and quantum chemical calculation studies by e.g., Hudgins & Allamandola (1999); Hony et al. (2001); van Diedenhoven et al. (2004); Bauschlicher et al. (2008) and Cami (2011). In our decomposition analysis, the 11.0 μm band is clearly associated with the PAH<sup>+</sup> component (Signal 2) based upon the strong 6.2 and 7.7 μm bands, while the 11.2 μm band is attributed to the neutral component (Berné et al. 2007). Hence we suggest that the variation of the [11.0]/[11.2] μm ratio is due to a changing abundance of PAH<sup>+</sup> to PAH<sup>0</sup>.

### 6.2. The 11.2 μm red wing

As mentioned in the introduction, the 11.2 μm band has an observed asymmetry with a varying red wing (Roche & Aitken 1985a). This wing has been attributed to anharmonicity or to different species of PAHs with a shifted solo mode peak emission (Pech et al. 2002; van Diedenhoven et al. 2004). Observations of the 11.2 μm feature show that the shape and peak position can vary. In their analysis of the skewed variations in the 11.2 μm profile, van Diedenhoven et al. (2004) empirically divided observations of the feature into two categories: one group is characterized by a peak between 11.20 and 11.24 μm and a more

skewed red wing, while the other peaks at 11.25 μm and is much more symmetric. In agreement with this, it has also been shown that PAH anions bands fall on the red side of the 11.2 μm peak and could contribute to the red wing (Bauschlicher et al. 2008).

By separating the observed spectra into component signals, we found that the main carrier of the 11.2 μm emission feature is PAH<sup>0</sup>. However, the observed spatial variations in the profile result in a PAH<sup>0</sup> source signal that is mainly symmetric with only a weak anharmonic red wing. The BSS analysis shows that the red wing is mainly due to a changing contribution of the VSG to the observed spectra; e.g., as the VSG signal becomes more prominent towards the outer edge of the PDR, its contribution to the observed spectra also increases (cf. Fig. 6). We note that in NGC 7023, position which are near to the exciting star are characterized by a more symmetric 11.2 μm profile while positions further away have a more pronounced red wing (cf. Fig. 9).

The feature at 11.2 μm has been observed to shift peak position between 11.2 to 11.3 μm. Studying Fig. 6, there are clear emission contributions from each component species throughout the 11.0 to 11.3 μm range. We propose that the variation of abundances of VSGs and PAHs in the observed spectra causes the shifting peak position of the 11.2 μm band. If there is a stronger contribution of PAHs in a certain region, the peak is observed to be blue-shifted. If the VSGs become more abundant, the peak is redshifted. This is in disagreement with van Diedenhoven et al. (2004), where they notice a redshifted peak with a more symmetric profile.

### 6.3. The 12.0, 12.7, and 13.5 μm features

Hony et al. (2001) assigned each emission feature to a different geometry and composition of PAH, depending on how many adjacent C-H groups are attached to the ring e.g., solo, duo, trio, and quartet modes of PAH<sup>0</sup> and PAH<sup>+</sup>. The results by Hony et al. (2001) were further expanded to include compact and irregular shaped large PAHs by Bauschlicher et al. (2008, 2009), which are more astronomically relevant. Bauschlicher et al. (2009) attributed the 11.3–12.3 μm band to the “duo1” CH mode while the 12.5–13.2 μm region is attributed to “duo2” CH<sub>oop</sub> bands, the split of duo modes being caused by coupling to other bending modes. The 13.5 μm feature has been attributed to the CH<sub>oop</sub> quartet mode of large irregular PAHs (Bauschlicher et al. 2009; Hony et al. 2001) and can be used to place constraints on the edge structures of the emitting PAHs. Here we will place further astronomical constraints on the results of Hony et al. (2001) and Bauschlicher et al. (2009).

#### 6.3.1. The 12.0 μm feature

The peak position of the “12.0 μm” feature varies from 11.8 μm to 12.0 μm (Fig. 9). The 12.0 μm feature has been attributed to both the PAH<sup>0</sup> and PAH<sup>+</sup> duo modes (Hony et al. 2001). Through the separation of source signals, we have isolated the main 12.0 μm feature to the PAH<sup>0</sup>. There is a feature that shares the profile of the 12.0 μm feature in the PAH<sup>+</sup> spectrum but it is blue-shifted, peaking around 11.8 μm.

#### 6.3.2. The 12.7 μm feature

The 12.7 μm feature has been predominately attributed to the overlap of PAH<sup>+</sup> duo and trio modes, but could not be definitely attributed to either PAH<sup>0</sup> or PAH<sup>+</sup> (Hony et al. 2001; Bauschlicher et al. 2008, 2009). Examining the results of the



signal separation in Figs. 6 and 8, two unique features at 12.7  $\mu\text{m}$  are revealed, that of PAH<sup>+</sup> and PAH<sup>0</sup>. Although they share roughly the same peak position, the PAH<sup>0</sup> 12.7  $\mu\text{m}$  feature is shifted to the red and is seen from 12.5 to 13.0  $\mu\text{m}$ , while the PAH<sup>+</sup> feature is blue shifted and asymmetric located between 12.3 and 12.8  $\mu\text{m}$ . We can attribute the variable blue wing of the 12.7  $\mu\text{m}$  feature to the changing abundance of PAH<sup>+</sup>. A peak shift and prominent asymmetry is seen in Fig. 9 in Position 1, located in the most concentrated area of PAH<sup>+</sup>. This feature is seen along with the increased 11.0  $\mu\text{m}$  feature, a blue-shifted peak position of the 11.2  $\mu\text{m}$  feature, a decreased red wing of the 11.2  $\mu\text{m}$  feature, and a “12.0  $\mu\text{m}$ ” feature peaking at 11.8  $\mu\text{m}$ .

### 6.3.3. The 13.5 $\mu\text{m}$ feature

The observed spectra show a distinct 13.5  $\mu\text{m}$  feature. In a study of M 17 it was suggested that this feature is coupled to the warm dust continuum (Verstraete et al. 1996). Hony et al. (2001) further investigated this possibility and instead, attributed the 13.5  $\mu\text{m}$  feature to a quartet out-of-plane bending mode of PAH<sup>+</sup> and PAH<sup>0</sup>. Using BSS, we isolated this feature to PAH<sup>+</sup> (Signal 2) and PAH<sup>0</sup> (Signal 1), in agreement with the results of Hony et al. (2001), and likely decoupled from the warm dust continuum.

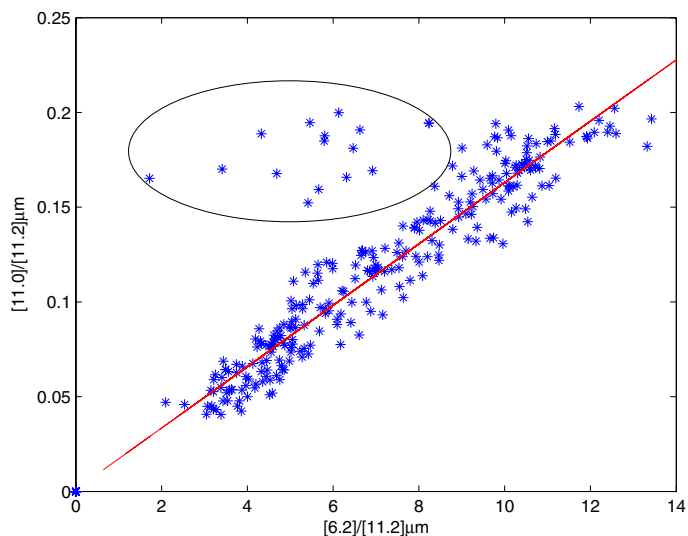
### 6.4. Systematic blue shift with ionization

We have attributed the 11.0  $\mu\text{m}$  feature to PAH<sup>+</sup>, while the 11.2  $\mu\text{m}$  feature is attributed to PAH<sup>0</sup>. In addition, the 11.8  $\mu\text{m}$  feature is attributed to PAH<sup>+</sup>, while the 12.0  $\mu\text{m}$  feature is attributed to PAH<sup>0</sup>. We also identify the broad 12.7  $\mu\text{m}$  band in both PAH<sup>+</sup> and PAH<sup>0</sup>, yet it appears to be a blend of features. The PAH<sup>+</sup> 12.7  $\mu\text{m}$  band is also bluer than the PAH<sup>0</sup> 12.7  $\mu\text{m}$  band. Specifically, the PAH<sup>+</sup> broad 12.7  $\mu\text{m}$  feature spans 12.3 to 12.8  $\mu\text{m}$  while the PAH<sup>0</sup> band stretches from 12.5 to 13.0  $\mu\text{m}$ . Comparing the PAH<sup>+</sup> and PAH<sup>0</sup> spectra, there is a systematic 0.2  $\mu\text{m}$  blue shift between the emitting bands. We do not observe this shift in the 13.5  $\mu\text{m}$  band. PAH band shifts can occur due to temperature change in the emitting region, yet according to the model proposed by Pech et al. (2002), a 0.2  $\mu\text{m}$  shift of the 11.2  $\mu\text{m}$  feature corresponds to a 650 K PAH temperature change. This PAH temperature change is too great to be observed within NGC 7023 NW, therefore it is unlikely that this band shift is due to a temperature change. Instead, we conclude that this shift is due to ionization, which modifies intrinsic emission properties of PAHs. Investigation on the exact origin of this shift is, however, beyond the scope of our paper.

### 6.5. Other possible effects on the shape of AIBs in the 10–15 $\mu\text{m}$ range

Other effects and chemical properties have been reported to alter the shape of AIBs in the 10–15  $\mu\text{m}$  range. Anharmonicity effects, as shown by e.g., Pech et al. (2002) can modify the position and the symmetry of the 6.2 and 11.2  $\mu\text{m}$  band and create the extended red wing in our observations. By means of DFT calculations, [SiPAH]<sup>+</sup>  $\pi$ -complexes were also proposed by Joalland et al. (2009) to produce a splitting of the initial 11.2  $\mu\text{m}$  PAH band into two bands at 11.0  $\mu\text{m}$  and 11.4  $\mu\text{m}$  due to the Si adsorption on the PAH edge creates and a blue-shifted 6.2  $\mu\text{m}$  band.

We argue here (see Sect. 6.2), that the asymmetry of the 11.2  $\mu\text{m}$  feature is predominately due to the contribution of VSGs. Anharmonic effects are however still observed: the PAH<sup>0</sup>



**Fig. 10.** The [6.2]/[11.2]  $\mu\text{m}$  ratio vs. the [11.0]/[11.2]  $\mu\text{m}$  ratio in NGC 7023-NW. The 11.2  $\mu\text{m}$  and 11.0  $\mu\text{m}$  measurements were made using the IRS-SH observations while the 6.2  $\mu\text{m}$  measurements were made using the IRS-SL observations. The circled data were not included in the fit (see text for details). The instrumental error is comparable to the symbol size.

signal is not fully symmetric and displays a slight red wing, suggesting that anharmonicity effects are still important, but recall that most of the red wing is due to the varying abundance of the VSG component. Since [SiPAH]<sup>+</sup> are expected to have a blue-shifted 6.2  $\mu\text{m}$  band, we inspected the SL data but found no such signature. The splitting of the 11.2 feature is seen in Signal 2 (PAH<sup>+</sup>), which is most concentrated in the regions near the star. Since the binding energy of [SiPAH]<sup>+</sup> is about 2 eV, they should be destroyed easily the highly irradiated environment near the star. Altogether, this suggests that compact PAH<sup>+</sup> are a more natural explanation for the 11.0  $\mu\text{m}$  feature, than [SiPAH]<sup>+</sup>  $\pi$ -complexes.

### 6.6. Using the 11.0 and 11.2 $\mu\text{m}$ features as tracers of ionization

With the attribution of the 11.0  $\mu\text{m}$  feature to PAH<sup>+</sup> and the 11.2  $\mu\text{m}$  to PAH<sup>0</sup>, we can investigate the possibility of using this ratio to probe the ionization fraction of PAHs in the PDR. One of the classic methods to trace the PAH ionization fraction is the [6.2]/[11.3]  $\mu\text{m}$  integrated intensity ratio (e.g., Galliano et al. 2008). There are other tracers of ionization such as the [7.7]/[11.3]  $\mu\text{m}$  and [8.6]/[11.3]  $\mu\text{m}$  ratios, but the 6.2, 7.7, and 8.6  $\mu\text{m}$  features include blended PAH<sup>+</sup> and PAH<sup>0</sup> bands. As we show here, the 11.0  $\mu\text{m}$  band is a purely cationic band and the 11.2  $\mu\text{m}$  band is purely neutral, increasing the accuracy of ionization fraction measurements. To demonstrate the reliability of the [11.0]/[11.2]  $\mu\text{m}$  ratio as an ionization indicator, we compare the [6.2]/[11.2]  $\mu\text{m}$  ratio to the [11.0]/[11.2]  $\mu\text{m}$  ratio using the IRS-SL and IRS-SH observations of the NGC 7023-NW (Fig. 10). In order to have an accurate measurement of the 11.2  $\mu\text{m}$  feature, without contamination from the 11.0  $\mu\text{m}$  satellite feature, we compare the integrated intensity of the 6.2  $\mu\text{m}$  feature from IRS-SL observations to the intensity of the 11.2  $\mu\text{m}$  feature using the high-resolution observations, since the 11.0 was not resolved and separated in the IRS-SL observations. The

maps were re-gridded using *Montage* so that each point of the SH map corresponds to the same spatial position on the SL map. Only the highest signal to noise data were used in this plot. For the  $6.2\ \mu\text{m}$  low-resolution map, we set a band integrated intensity threshold of  $10^{-6}\ \text{W m}^{-2}\ \text{sr}^{-1}$ . For the  $11.0\ \mu\text{m}$  high-resolution map we set a threshold of  $10^{-7}\ \text{W m}^{-2}\ \text{sr}^{-1}$  and the  $11.2\ \mu\text{m}$  high-resolution map has a threshold of  $10^{-6}\ \text{W m}^{-2}\ \text{sr}^{-1}$ . The  $[6.2]/[11.2]$  vs.  $[11.0]/[11.2]\ \mu\text{m}$  ratio in NGC 7023 is presented in Fig. 10. The data reveal a clear correlation, validating the use of the  $[11.0]/[11.2]\ \mu\text{m}$  ratio as a PAH ionization indicator. The outliers in the upper left corner correlate to spectra where the thermal continuum from the source star is contaminating the linear continuum subtraction. For this reason, these points were not included in the linear fit. The linear fit has a high correlation coefficient of 0.95, from which an empirical relation can be derived:

$$\frac{[11.0\ \mu\text{m}]}{[11.2\ \mu\text{m}]} = 0.016 \times \left( \frac{[6.2\ \mu\text{m}]}{[11.2\ \mu\text{m}]} \right). \quad (1)$$

## 7. Nature of the VSGs

The BSS analysis identifies an independent broad component underneath the well-known  $11.2$  and  $12.7\ \mu\text{m}$  bands. Earlier BSS studies over a wider wavelength range and the spatial distribution of Signal 3 (Fig. 5) assigns this component to emission by VSGs, proposed to be PAH clusters. Early observations of the  $11.2\ \mu\text{m}$  feature and its underlying emission support the suggestion that this broad underlying pedestal arises from a separate component (Roche & Aitken 1985b). The PAH Database analysis provides some further insight in the character of the carrier of this broad component. In this analysis, the  $11\text{--}15\ \mu\text{m}$  pedestal emission is due to a large number of individual components originating in a wide variety of molecular edge structures (solo's, duo's, and trio's), which together blend in an indistinct broad emission bump from  $11\text{--}15\ \mu\text{m}$ . For this blend, the analysis selects relatively small species from the database. However, that is a selection effect. Small PAHs have, by necessity, a preponderance of corner structures. In contrast, calculations for large PAHs have focused (for obvious reasons) on regular structures with long straight edges and consequently strong  $11.2\ \mu\text{m}$  bands and weak bands at longer wavelengths. We surmise that large irregular PAHs would equally fit the bill. The VSG component has been assigned to clusters of PAHs based upon an interpretation of the observed spatial distribution and the physical properties of clusters (Berné et al. 2007; Rapacioli et al. 2005, 2006). However, the spectroscopic properties of PAH clusters are presently unknown. While in general their spectra might be expected to resemble those of the constituent PAH molecules making up the cluster, we surmise that steric hindrance may affect the frequencies of the out-of-plane CH bending modes. We realize that there is a hidden issue here: the spectral differences in the  $11\text{--}15\ \mu\text{m}$  range – the broad and indistinct band in the VSG component versus the very distinct  $11.2$  and  $12.7$  modes of the PAHs – implies a more complicated evolutionary relationship between the VSGs and the PAHs than simple evaporation.

## 8. Conclusion

Applying a BSS method to observations from *Spitzer* Infrared Spectrograph, Short-wavelength High-resolution mode, we uncovered 3 component signals in the PDR NGC 7023-NW. We

found that each signal is most abundant in different regions of the PDR, depending on the radiation environment. We identified the three component signals as PAH cations, neutral PAHs, and VSGs. As the observed spectra suggest, the neutral PAHs dominate every region of the PDR, but are most heavily concentrated in between the PAH cations and VSGs. Both the spectra and spatial maps of each signal show high correlation to the results using *Spitzer's IRS-SL mode* (Berné et al. 2007, 2010), allowing us to use these results to verify our conclusions.

To further explore the origin of the three resolved signals, we employed the NASA Ames PAH IR Spectroscopic Database. The fit shows that the component spectra resolved by BSS could be recreated by an appropriate combination of specific classes of PAH spectra from the database. Then, we used a database to fit an observed spectrum and grouped the individual molecules into charge class, then compared the spectra of the combined charge classes to the BSS extracted  $\text{PAH}^+$  and  $\text{PAH}^0$  signals. The components were found to be very similar to the BSS extracted  $\text{PAH}^+$  and  $\text{PAH}^0$ .

Specific spectral properties are found for each population:

- We have attributed the  $11.0$  and  $11.2\ \mu\text{m}$  bands to cations and neutral species respectively.
- We conclude that the variation of the  $[11.2]/[11.0]\ \mu\text{m}$  ratio depends on the relative abundances of PAH cations to neutral PAHs.
- The extended red wing seen on the  $11.2\ \mu\text{m}$  feature is attributed to the increasing abundance of VSGs and the broad  $11.3\ \mu\text{m}$  feature that is characteristic of this component.
- The changing peak position of the  $11.2\ \mu\text{m}$  feature can also be explained by varying contributions from PAHs (blue shift) and VSGs (red shift).
- The  $12.0\ \mu\text{m}$  feature is attributed to neutral PAHs while the  $11.8\ \mu\text{m}$  feature is attributed to PAH cations, therefore, as the ionization mixture changes, the peak of this feature will shift accordingly.
- Since the  $13.5\ \mu\text{m}$  feature is present in both PAH cations and neutral PAHs, but not existent in the VSG signal, where we see the continuum, we agree with Hony et al. (2001) that the  $13.5\ \mu\text{m}$  signal is decoupled from the  $15\ \mu\text{m}$  continuum.

By using the BSS method and the PAH Database fit, we arrived at the above conclusions. Each method has unique yet complementary strengths and weaknesses. The BSS method is blind, i.e. has no intrinsic assumptions about the emitting components, however since the statistical properties of the emitting components are unknown, the unmixing is not perfectly efficient. Additionally, the BSS method separates 3 mathematically distinct signals, but gives no intuition about the molecular properties of these signals. The PAH Database allows for direct physical interpretation of the fit yet is biased towards smaller molecules and lacks spectral information for PAH clusters or other possible carriers of the VSG signal. Although both methods suffer limitations, the strengths of one compensate for the weaknesses of the other. Although the database fit may be degenerated in some cases, the interpretation of the  $\chi^2$  database fit results, in terms of classes, is in agreement with the result of the BSS for  $\text{PAH}^+$  and  $\text{PAH}^0$ . The VSG spectrum can only be obtained by BSS, and the database fit of this spectrum provide additional information on the possible chemical nature of this component. Both methods, i.e. BSS and database fitting, are powerful tools, but they *must* be used with an understanding of their limitations (described in details in Appendices A and B).

**Acknowledgements.** This work was conducted by M. Rosenberg in part fulfillment of the M.Sc. Degree in Space Studies at the International Space University (ISU), Strasbourg, France. The author also acknowledges M.Sc. scholarship support from ISU. L. J. Allamandola acknowledges support from NASA's Astrobiology and Laboratory Astrophysics Program. We acknowledge B. Joalland for fruitful discussions on the spectral properties of SiPAH complexes. This research made use of Montage, funded by the National Aeronautics and Space Administration's Earth Science Technology Office, Computation Technologies Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology. Montage is maintained by the NASA/IPAC Infrared Science Archive. C. Boersma acknowledges support by an appointment to the NASA Postdoctoral Program at the Ames Research Center, administered by Oak Ridge Associated Universities through a contract with NASA. ERC grant: studies of interstellar PAH at Leiden Observatory are supported through advanced-ERC grant 246976 from the European Research Council. The authors also thank the referee for his/her time, comments, and suggestions.

## Appendix A: Exploring the artifacts of BSS

See Figs. A.1. and A.2.

## Appendix B: Exploring the limitations of the PAH database fitting

The efficiency of NMF is subject to two main limitations: 1) the possible non-unicity of solutions 2) the inaccuracy of the unmixing in the presence of noise. These two problems are the subject of intensive theoretical research in the field of signal processing (see e.g., Donoho & Stodden 2003; and Laurberg et al. 2008). If the statistical properties of the source matrix are known, then one can determine whether NMF will function properly. In a "real life" case like here however, since the matrix is not known a priori, we cannot verify these statistical properties. We therefore rely on some empirical considerations. Procuring the component spectra (Fig. 3), the most prominent unmixing artifact is located at 11.0  $\mu\text{m}$  and reveals itself with a sharp drop in Signal 3. Some less discernible artifacts can be found at 11.2  $\mu\text{m}$  as an added peak of Signal 2 (seen in the grey envelope of Fig. 3) and at 12.7  $\mu\text{m}$  as a dip in intensity of Signal 3, roughly resembling an absorption feature.

We suggest that the unmixing artifacts seen in some signals are compensated for by an increase or decrease of intensity, at the same wavelength, in other signals. This is most clearly seen at the 11.0  $\mu\text{m}$  wavelength. The sharp drop to 0 of Signal 3 is compensated by the weak satellite feature of Signal 1. Similarly, the absorption-type feature seen at 12.7 is compensated by a slightly increased intensity of Signal 2 at 12.7  $\mu\text{m}$ . In order to better understand the unmixing efficiency, two tests were conceived:

1. We have artificially mixed the database PAH<sup>0</sup> and PAH<sup>+</sup> (Fig. 8) signals with the NMF extracted VSG signal of Berné et al. (2007) (Fig. 6). We mix these signals with a  $3 \times 100$  random matrix, but we make sure that the coefficients correspond to realistic relative abundances (e.g., there are no regions with 100% PAH<sup>+</sup>). We add synthetic white gaussian noise to the spectra to obtain a signal to noise ratio of about 20 (similar to our observations). In this way, we created 100 simulated spectra. We then apply NMF to this set of spectra. An example of the extracted signals is shown in Fig. A.1. There is a clear absorption band in the VSG spectrum, very similar to what is observed in the VSG spectrum extracted from the observations (Fig. 3). Lowering the noise does not improve these results. This implies that this

artifact is caused by improper unmixing, and that this is an inherent problem of the NMF algorithm: the *source* spectra we are trying to extract most likely do not verify completely the statistical properties that are needed for NMF to function efficiently (Donoho & Stodden 2003). We conclude that the absorption feature we observe in the VSG extracted spectrum (Fig. 3), is caused by the improper unmixing of NMF.

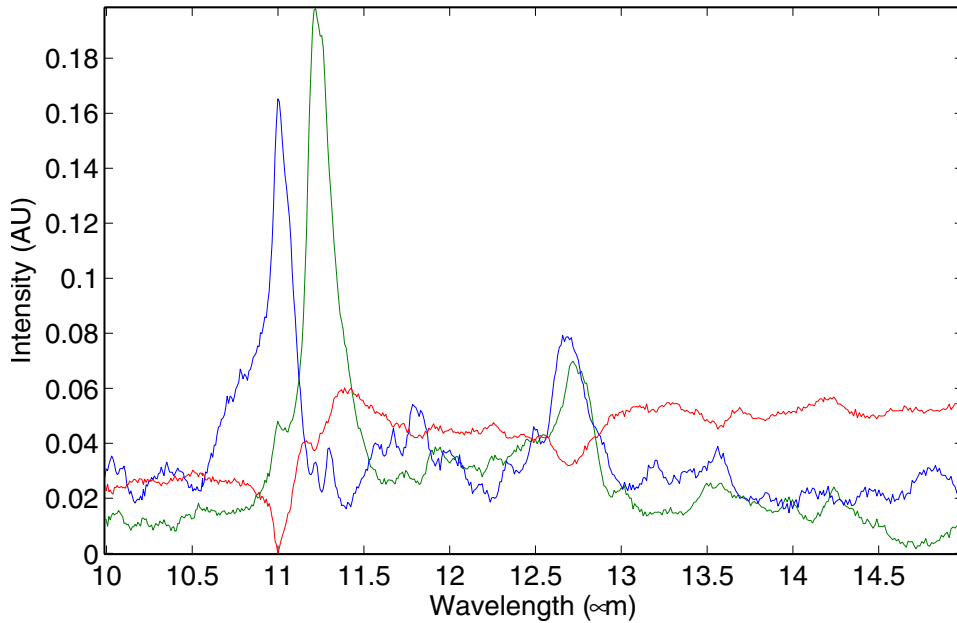
2. In order to test why Berné et al. (2007) did not see such an artifact in their VSG spectrum, the same BSS NMF algorithm was used to separate the IRS-SL data cube components but with only the 10–14  $\mu\text{m}$  wavelength range included. The result of this BSS is shown in Fig. A.2, where the 11.0  $\mu\text{m}$  absorption artifact is clearly visible. The fact that this artifact becomes present when only including the 10–14  $\mu\text{m}$  spectral range implies a lack of information and constraints in this particular wavelength range. Rapacioli et al. (2005) and Berné et al. (2007) show that the main features of the VSG spectrum are 1) a broad 7.7  $\mu\text{m}$  band, 2) a continuum, dominant from 15 to 20  $\mu\text{m}$ . Without these constraints in the reduced spectral range of IRS-SH, BSS has a difficulties extracting enough information for this signal. We conclude that the artifacts especially in the VSG spectrum are mostly due to this reduced spectra range.

To investigate the level of degeneracy of the database fit, we have performed several tests. Here, we present four fits with explore this important issue. All fits follow the same values described in Sect. 5: a 6  $\text{cm}^{-1}$  FWHM, a 15  $\text{cm}^{-1}$  band redshift, a Lorentzian profile, a 6 eV excitation energy with cascade emission, and molecules including only Carbon, Hydrogen, and Nitrogen. The fits are described below.

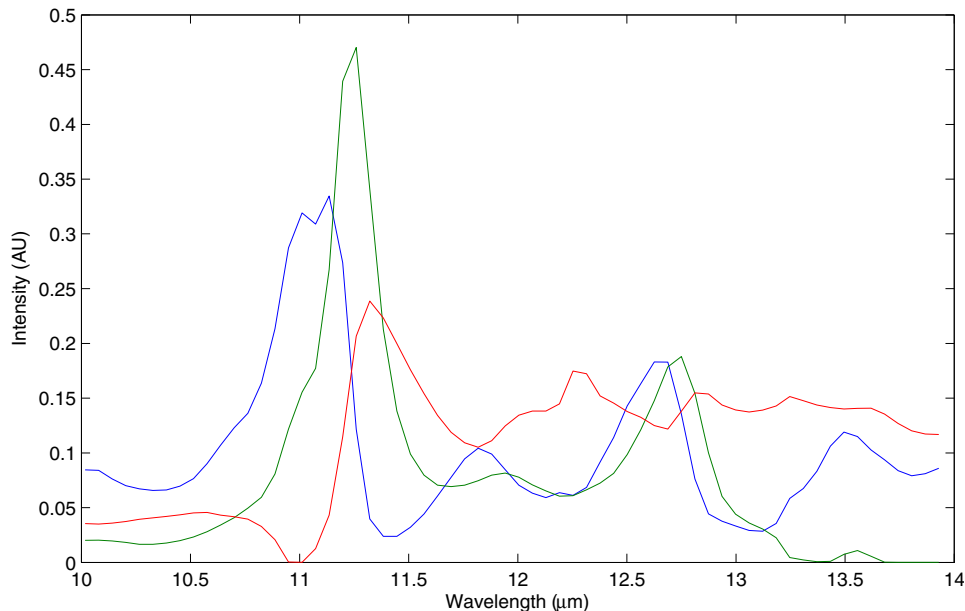
1. The BSS extracted PAH<sup>+</sup> spectra is fit with the database restricting it to *only* ionized species (top left of Fig. B.1).
2. The BSS extracted PAH<sup>0</sup> spectra is fit with the database restricting it to *only* neutral species (top right of Fig. B.1).
3. The BSS extracted PAH<sup>+</sup> spectra is fit with the database restricting it to *only* neutral species (bottom left of Fig. B.1).
4. The BSS extracted PAH<sup>0</sup> spectra is fit with the database restricting it to *only* ionized species (bottom right of Fig. B.1).

While the database fit of the BSS PAH<sup>+</sup> signal with PAH<sup>+</sup> cations is excellent, when it is limited to neutral species (bottom left Fig. B.1), it is clear that the PAH<sup>+</sup> spectrum cannot be recreated. However, the database fit of the BSS PAH<sup>0</sup> species with cations does provide a suitable fit, except for one key region: the 11.2  $\mu\text{m}$  feature. The best fit of the database produces a two-pronged feature with neither prong peaking exactly at the 11.2  $\mu\text{m}$  band. However, it is possible that by varying the FWHM and band shift, which are somewhat arbitrary values, these bands could blend to recreate the 11.2  $\mu\text{m}$  feature. This is due to the fact that various PAH cations peak throughout the mid-IR spectrum. Specifically, PAH cations have emission features that peak throughout the 10.8–11.3  $\mu\text{m}$  range. This only highlights again the need for both approaches to apply as many astronomical constraints to the emitting species as possible, including wider spectral coverage when available. The BSS method allows us to spatially separate the three emitting components and therefore we create a physical picture in which Signal 2, located nearest to the star, represented the PAH<sup>+</sup>. In this case, the database supports this claim since we cannot fit this spectra with only neutral PAH species. We then move to Signal 1, which has a much different emission spectrum than the PAH<sup>+</sup> and is located in the middle





**Fig. A.1.** The result of BSS using the NMF method on artificially mixed combinations of the PAH Database PAH<sup>+</sup> and PAH<sup>0</sup> signal and the VSG signal from Berné et al. (2007). BSS extracted PAH<sup>+</sup> in blue, extracted PAH<sup>0</sup> in green, and extracted VSG in red.



**Fig. A.2.** The result of BSS using the NMF method on the IRS-SL data from Berné et al. (2007) restricting the wavelength range to 10–14  $\mu\text{m}$ . BSS extracted PAH<sup>+</sup> in blue, extracted PAH<sup>0</sup> in green, and extracted VSG in red.

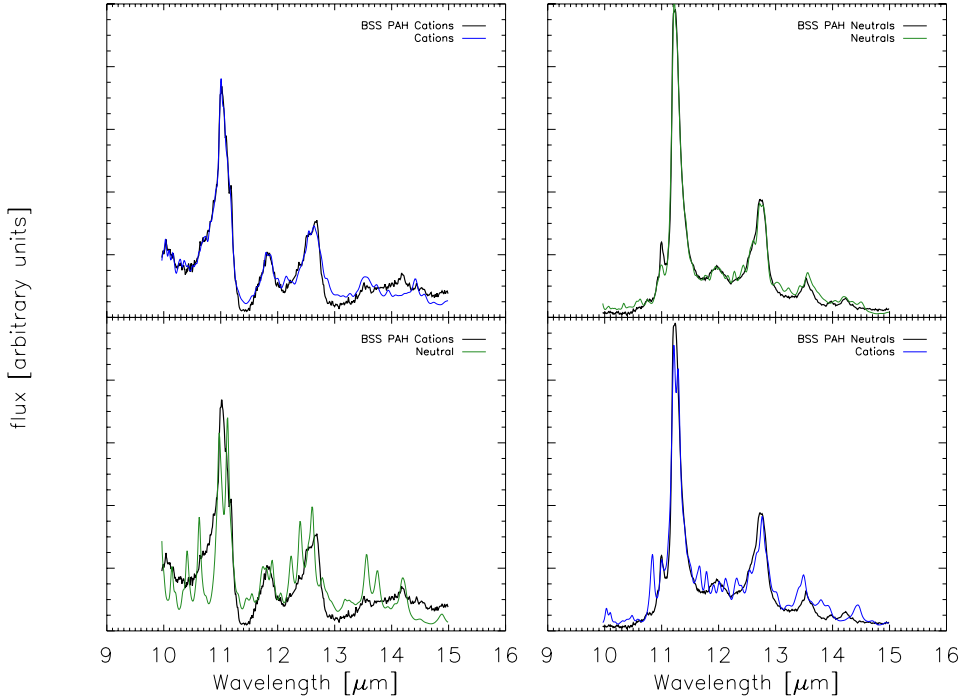
of the PDR with a much higher abundance than the PAH<sup>+</sup>. The database shows that this can be fit with either only neutral or only cation PAH species, but the physical intuition of BSS indicates that this is most likely a neutral PAH spectrum.

Essentially, the database is a collection of Lorentzians that when restricted to a narrow band pass in which aromatic bands fall, and added together with varying intensity, can likely fit any type of aromatic-rich materials. Especially in the case of cations, the fit is degenerate (Fig. B.1). Yet when the fit is performed with the full range of the database, the highest  $\chi^2$  solution is strikingly similar to the BSS extracted solution (Fig. 8). Although our tests show that the fits are poorly constrained for individual PAHs, this is not the case when considering PAH sub populations, which trace properties such as size and charge. Thus, the astronomical mid-IR bands studied here are assigned to PAH molecules as a class, not to individual PAH molecules.

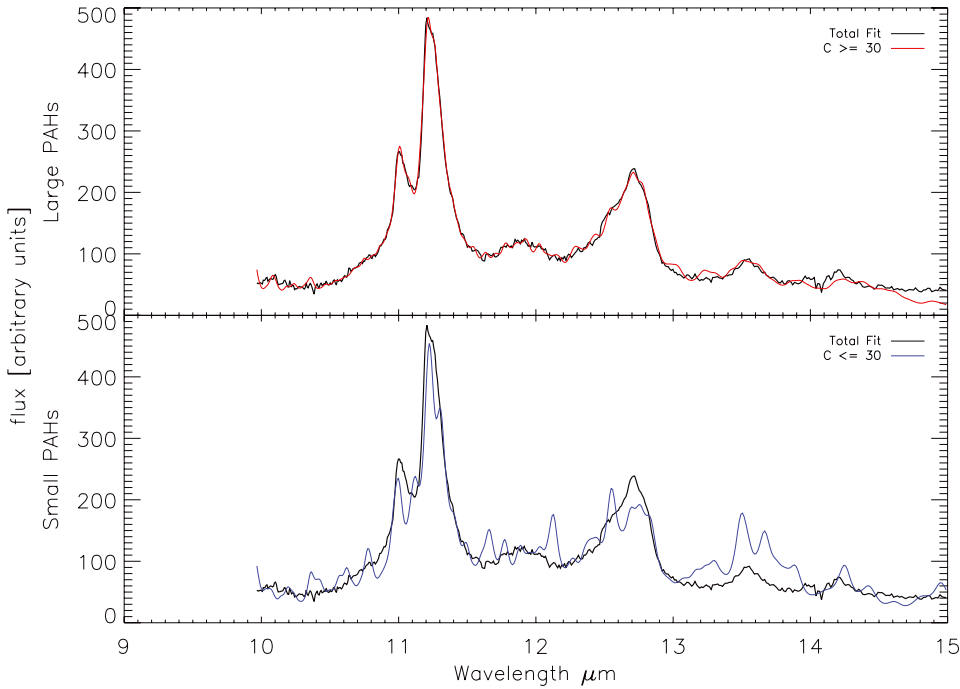
The literature is rich in examples where it is shown that the mid-IR PAH emission features can be categorized by charge, size, shape and composition (Peeters et al. 2004; Bauschlicher et al. 2008, 2009; Boersma et al. 2010). When a single species is removed from the pool of species that can be fit, another species from the same subgroup will simply replace it. Therefore, a mid-IR fit made with the database is expected not to provide information about the presence of an individual PAH molecule, but about the subclasses of PAHs involved and the response of that subclass to the local astronomical environment.

Aside from charge, the database also probes the subclass of size. Although we have a limited wavelength range, the main features are best reproduced by larger PAH molecules. Similarly to the case of charge, we can treat PAH size as another sub population with a set of characteristic properties. In Fig. B.2, we use the database to fit the observed spectra from Fig. 8. Here, besides





**Fig. B.1.** PAH Database fit results of fitting the BSS extracted  $\text{PAH}^+$  (left) and  $\text{PAH}^0$  (right) signals using only cations (blue) or only neutral species (green). The black line represents the original BSS extracted spectra while the fit is represented in color.



**Fig. B.2.** PAH Database fit results of fitting an observed spectrum limiting the species to large PAHs ( $C \geq 30$ ) (top) and small PAHs ( $C \leq 30$ ) (bottom). The black line represents the original observed spectrum, where the fit is represented in color.

size the only constraint applied is including only C, H, and N atoms. The parameters are equal to those described in Sect. 5.1. The top panel displays a fit of the observed spectrum using only PAHs with 30 or more C atoms. The bottom panel is a fit of the observed spectrum using PAHs with 30 or less C atoms. There are 255 PAH species available to fit  $C \leq 30$ , whereas there are 237 PAH species available to fit  $C \geq 30$ , once the other constraints are taken into account. Even though there are less PAH species included in the fit of large PAHs, it is a superior fit to the small PAHs and has a lower reduced norm value. From this limited wavelength range, we can already deduce that large PAHs

are responsible for the main bands, however to have a stronger constraint on size, an analysis using the full 3–15  $\mu\text{m}$  range in necessary.

## References

- Allamandola, L. J., Tielens, A. G. G. M., & Barker, J. R. 1985, *ApJ*, 290, L25  
Allamandola, L. J., Tielens, A. G. G. M., & Barker, J. R. 1989, *ApJS*, 71, 733  
Barker, J. R., Allamandola, L. J., & Tielens, A. G. G. M. 1987, *ApJ*, 315, L61  
Basire, M., Parneix, P., & Calvo, F. 2010, *J. Phys. Chem. A*, 114, 3139  
Bauschlicher, Jr., C. W., Peeters, E., & Allamandola, L. J. 2008, *ApJ*, 678, 316  
Bauschlicher, C. W., Peeters, E., & Allamandola, L. J. 2009, *ApJ*, 697, 311

- Bauschlicher, C. W., Boersma, C., Ricca, A., et al. 2010, *ApJS*, 189, 341
- Berné, O., Joblin, C., Deville, Y., et al. 2007, *A&A*, 469, 575
- Berné, O., Tielens, A., Pilleri, P., & Joblin, C. 2010, in 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 1
- Boersma, C., Bauschlicher, C. W., Allamandola, L. J., et al. 2010, *A&A*, 511, A32
- Cami, J. 2011, *EAS Publ. Ser.*, 46, 117
- Cesarsky, D., Lequeux, J., Abergel, A., et al. 1996, *A&A*, 315, L305
- Cherchneff, I., Barker, J. R., & Tielens, A. G. G. M. 1992, *ApJ*, 401, 269
- Cook, D. J., & Saykally, R. J. 1998, *ApJ*, 493, 793
- Donoho, D., & Stodden, V. 2003, *Advances in Neural Information Processing Systems (NIPS)*
- Galliano, F., Madden, S. C., Tielens, A. G. G. M., Peeters, E., & Jones, A. P. 2008, *ApJ*, 679, 310
- Hollenbach, D. J., & Tielens, A. G. G. M. 1999, *Rev. Mod. Phys.*, 71, 173
- Hony, S., Van Kerckhoven, C., Peeters, E., et al. 2001, *A&A*, 370, 1030
- Hudgins, D. M., & Allamandola, L. J. 1999, *ApJ*, 516, L41
- Hudgins, D. M., Bauschlicher, Jr., C. W., & Allamandola, L. J. 2005, *ApJ*, 632, 316
- Joalland, B., Simon, A., Marsden, C. J., & Joblin, C. 2009, *A&A*, 494, 969
- Joblin, C., & Mulas, G. 2009, *EAS Publ. Ser.*, 35, 133
- Joblin, C., & Tielens, A. G. G. M. (eds) 2011, *PAHs and the Universe: A Symposium to Celebrate the 25th Anniversary of the PAH Hypothesis (EDP Sciences)*, *EAS Publ. Ser.*, vol. 46
- Joblin, C., Boissel, P., Leger, A., D'Hendecourt, L., & Defourneau, D. 1995, *A&A*, 299, 835
- Joblin, C., Tielens, A. G. G. M., Geballe, T. R., & Wooden, D. H. 1996, *ApJ*, 460, L119
- Laurberg, H., Christensen, M. G., Plumbley, M. D., Hansen, L. K., & Jensen, S. H. 2008, *Comput. Intell. Neurosci.*, 2008, 764206
- Lee, D. D., & Seung, H. S. 2001, *Neural Information Processing Systems (MIT Press)*, 13, 556
- Léger, A., & Puget, J. L. 1984, *A&A*, 137, L5
- Mallocki, G., Joblin, C., & Mulas, G. 2007, *Chem. Phys.*, 332, 353
- Mulas, G., Mallocki, G., Joblin, C., & Toublanc, D. 2006a, *A&A*, 460, 93
- Mulas, G., Mallocki, G., Joblin, C., & Toublanc, D. 2006b, *A&A*, 456, 161
- Nuzillard, D., & Bijaoui, A. 2000, *A&AS*, 147, 129
- Oomens, J. 2011, *EAS Publ. Ser.*, 46, 61
- Oomens, J., Tielens, A. G. G. M., Sartakov, B. G., von Helden, G., & Meijer, G. 2003, *ApJ*, 591, 968
- Pech, C., Joblin, C., & Boissel, P. 2002, *A&A*, 388, 639
- Peeters, E., Hony, S., Van Kerckhoven, C., et al. 2002, *A&A*, 390, 1089
- Peeters, E., Mattioda, A. L., Hudgins, D. M., & Allamandola, L. J. 2004, *ApJ*, 617, L65
- Puget, J. L., & Léger, A. 1989, *ARA&A*, 27, 161
- Rapacioli, M., Joblin, C., & Boissel, P. 2005, *A&A*, 429, 193
- Rapacioli, M., Calvo, F., Joblin, C., et al. 2006, *A&A*, 460, 519
- Roche, P. F., & Aitken, D. K. 1985a, *MNRAS*, 215, 425
- Roche, P. F., & Aitken, D. K. 1985b, *MNRAS*, 215, 425
- Sellgren, K., Uchida, K. I., & Werner, M. W. 2007, *ApJ*, 659, 1338
- Sellgren, K., Werner, M. W., Ingalls, J. G., et al. 2010, *ApJ*, 722, L54
- Sloan, G. C., Hayward, T. L., Allamandola, L. J., et al. 1999, *ApJ*, 513, L65
- Smith, J. D. T., Armus, L., Dale, D. A., et al. 2007, *PASP*, 119, 1133
- Tielens, A. G. G. M. 2008, *ARA&A*, 46, 289
- van den Ancker, M. E., The, P. S., Tjin A Djie, H. R. E., et al. 1997, *A&A*, 324, L33
- van Diedenhoven, B., Peeters, E., Van Kerckhoven, C., et al. 2004, *ApJ*, 611, 928
- Verstraete, L., Puget, J. L., Falgarone, E., et al. 1996, *A&A*, 315, L337
- Werner, M. W., Uchida, K. I., Sellgren, K., et al. 2004, *ApJS*, 154, 309