Source Separation of Astrophysical Ice Mixtures

Jorge Igual, Raul Llinares, and Addisson Salazar

Universidad Politecnica de Valencia, Departamento de Comunicaciones,
Pza. Ferrandiz y Carbonell 1, CP 03801, Alcoy, Spain
{jigual, rllinares, asalazar}@dcom.upv.es

Abstract. Infrared spectroscopy provides direct information on the composition of interstellar dust grains, which play a fundamental role in the evolution of interstellar medium ISM, from cold, quiet and low density molecular clouds to warm, active and dense protostellar ones. The determination of these components is fundamental to predict under the appropriate environmental conditions their evolution, including the appearance of new molecules, radicals and complex organics. The absorption spectrum of the astrophysical ice can be considered the additive linear absorption spectra of the multiple molecules present in the ice, so a linear instantaneous ICA mixture model is appropriate. We present the ICA statement of the problem, discussing the convenience of the model and its advantages in front of supervised methods. We obtain the MAP estimate of the mixing matrix, including its non-negative entries as a prior. We present the results carried out with an ice analogs database, confirming the suitability of the ICA approach.

1 Introduction

The origin of the biogenic elements in Earth is an intriguing question [1]. It is thought that prebiotic molecules were present on primitive Earth. Different processes could be responsible for their evolution into more complex ones. One of the most famous experiments was carried out by Miller [2]. In this experiment, glycine was formed by energetic processes such as the electric discharge of a mixture of HCN and aldehydes acids. Other experiments [3] demonstrate that other molecules of interest such as carbonic acid (H_2CO_3) could be formed from simple molecules such as CO by protons implantation. The understanding of chemical evolution of elementary molecules as water, carbon monoxide, methane or carbon dioxide is one of the key questions for understanding the origin of life on Earth.

These molecules are combined in different quantities, called concentrations, in other compounds and found usually in the frozen state, named from now on ice mixtures.

One way to study this chemical evolution consists of the simulation of these processes in the laboratory. In the laboratory, we are interested in the analysis of the composition of different ices and their behavior in specific conditions that reproduce situations in which these substances are found outside of the laboratory. For example: to raise the temperature gradually or to radiate the samples with ions that are similar to the ones originated outside of the planet Earth.

The signal that identifies each molecule is the infrared absorption spectrum, where the absorption bands correspond to specific vibrational mode of the molecules, each one with different atoms and bonds. We know that the frequencies corresponding to the middle infrared spectrum (4000-400 cm^{-1}; 2.5-25 µm) span the same range as the vibrational frequencies of the adjacent atoms in molecules associated with the cosmogenically most abundant species. The components are called endmembers and the spectra signatures or digital tracks.

ICA performs a "multi-ice" unsupervised decomposition of the data in different spectra instead of a classical matched filter detection approach. The supervised approach raises other difficulties in our application; not only the requirement of an expert with the risk of misinterpretations, but furthermore, the library of endmembers could be very large because considering that the spectrum vary depending on many variables such as temperature or irradiation, we would need a description of every compound for every case. In fact, even under very controlled conditions, like in a laboratory, the spectra registered in different experiments for the same mixed ice can sometimes change. In Figure 1 we represent the CO_2 spectrum band around 2350 cm^{-1} obtained in Leiden and Polytechnic University of Valencia UPV laboratories. The mixed ice production consists of defining the mixture, preparing the gases necessary in the corresponding abundance, introducing them in the gas container using the gas law to convert pressures to abundances, and depositing on a cold substrate. The final ice can vary due to different problems, e.g., the purity of the gases, small changes in the temperature, sequential mixing of gases, or the influence of the molecular mass in the deposition rate [4].

2   Infrared Absorption Features of Astrophysical Ices

Life on Earth originated in stars which formed out of interstellar gas and dust [5]. The lifecycle of an interstellar ice mantle consists of the following stages: a.) formation and growth by surface reactions on grains, b.) the ices are subject to energetic phenomena associated to the origin of stars that lead to c.) thermal evaporation or photolysis into more complex organic solids and d.) the material is returned to the ISM or accreted to providing the necessary conditions for the formation of planetary bodies. To understand all this process, knowledge of the different components of the ices is fundamental and it is achieved through the infrared absorption signatures corresponding to each molecule. The spectrum is formed by absorption bands around some specific wavelengths determined by their atomic composition and bond structure, e.g., the 4.67 µm (2140 cm^{-1}) C≡O stretching band; their peak position and width depend on the presence or absence of some molecules that can affect their dipolar moment, e.g.., when the break-up of the hydrogen bond weakens the O-H stretching feature of fully H bonded water at 3 µm (3280 cm^{-1}) [6], in addition to temperature and particle shape. Besides, these bands usually have an area and a width related to the compound concentration in the ice.

There are different magnitudes related to the absorption. One of them is the optical depth τ, which is defined as the integral of the absorption coefficient times the density along the path. It can be calculated for every frequency v as:

$$I(v) = I_0(v)e^{-\tau_v} \qquad (1)$$

where I is the intensity which passes through the ice and I_0 is the initial infrared beam intensity. The optical depth ranges between zero, when the radiation impinging goes through the material and arrives to the detector, therefore $I = I_0$, i.e., there is no absorption at all, and infinite if the ice is able to absorb all the arriving radiation $I = 0$. As an example, in Fig. 2 we show the infrared spectra of the methane and water registered at the UPV laboratory. Obviously a background spectrum is first registered before the measurements of the ice under test.

Some artifacts can appear due to preprocessing tasks, complex physics and noise sources, e.g., changes with time and temperature, calibrations or sensor noise. For example, in Fig. 2 we can see wavelengths where the optical depth takes negative values, such as around the CH_4 methane peak around 10 μm.

Of all the preprocessing tasks, the most important one is the baseline removal, because in very attenuate absorption bands a bad approximation of the baseline can mask some compounds or produce negative values of the optical depth.
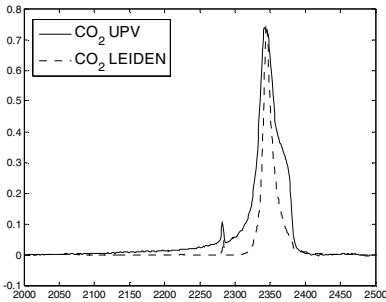
Fig. 1. CO_2 spectrum. Absorbance vs. wave-number.
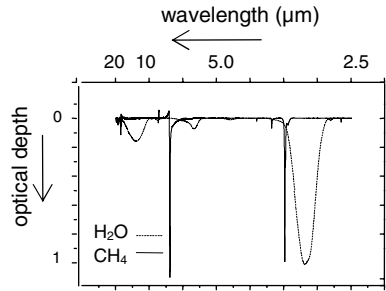
Fig. 2. Water and methane spectrum. Optical depth vs. wavelength.

3   ICA Framework

ICA has been applied in different astrophysical applications, e.g., to study celestial sources [7], to separate the cosmic microwave background, the galactic dust radiation, the synchrotron radiation and the free-free radiation components of the microwave sky radiation [8] or the mineralogical identification with the OMEGA spectra on board Mars Express [9]. The infrared spectrum has also been used in source separation problems [10], [11].

Only the sign factor indeterminacy of ICA must be considered, taking into account that it is really straightforward to detect the correct one, since the optical depth of the sharp absorption bands must be also positive. The restriction of at most one Gaussian source is not either a problem; the histogram analysis of pure molecules is far from corresponding to a Gaussian random variable.

3.1 ICA Model

The ICA formulation of the problem is: the infrared absorption spectrum x_{ij} measured for $i = 1,...,M$ ice mixtures in the spectral band $j = 1,...,P$, typically corresponding to 4000 up to 400 cm^{-1} with resolutions 1 or 2 cm^{-1}, is the linear combination of the independent absorption spectra s_{kj} of the molecules (sources) $k = 1,...,K$ present in those ices. The concentration of molecule k in ice i is the mixing matrix entry a_{ik}. The concentrations a_{ik} and absorption spectra s_{kj} are non-negative, although some preprocessing tasks such as baseline removal, noise and complex physics in the measurement process can produce a negative x_{ij}. All these processes will be resumed in a noise term n_{ij} for ice i in wavelength j. In matrix notation, for a given wavelength, we have:

$$\mathbf{x} = \mathbf{As} + \mathbf{n} \tag{2}$$

3.2 Discussion of the Model

We will review some physical considerations about the absorption spectra and endmembers in order to explain and discuss the conditions when the ICA model (2) is suitable for the problem.

Independency of the sources. The sources correspond to the spectral signatures of the endmembers. At a first glance, the hypothesis sounds rather intuitive that the absorption of different molecules has nothing to do due to the fact that vibrational frequencies of different components are not correlated. This is true, in general, but not always. As an example, at 10 K, the band width of the CO-stretch in CO/O_2 mixtures increases with O_2 concentration from 2.2 cm^{-1} up to 5.5 cm^{-1} when the amounts are equal, and then decreases again when O_2 is more abundant [6]. To model these relationships with conditional densities for a general case is very complicated. The same problem arises when the spectrum changes with temperature and radiation, so the only thing we can do in such ill-conditioned situations is for practical purposes to suppose that there are different kinds of CO, i.e., the number of sources must be increased.

Statistics of the sources. Considering the histogram of the spectra of the endmembers, most of them correspond to supergaussian signals, due to there being no absorption in most wavelengths.

Instantaneous mixture. The instantaneous model is clearly appropriate since absorption bands are not related. Nevertheless, a convolutive model could be suggested for the case aforementioned in the discussion about independency.

Linear mixture. The linear mixing model is also a usual supposition, i.e., the total absorption spectrum is the linear addition of the absorption features of every ice compound. But it is known that this is a simplified model in environments such as the planetary atmospheres, where the mixing is exponential [12].

Noise. It is basically due to measurement noise, although the preprocessing can also distort the actual signal. It is usually considered a white or coloured Gaussian noise. In our case, as the sensor noise level is relatively low (high quality measure-

ments in the laboratory), it will not be considered in our model, so the relation between sources and mixtures becomes deterministic and equation (2) simplifies.

Number of mixtures. Usually $M > K$, i.e., more ices than simple molecules are available. In fact, it will be usual in the whitening step to reduce the dimension of the problem. The list of sources usually include CO, CO_2, H_2O, NH_3 and CH_4.

Priors. Absorption spectra are non-negative. However, as we explained before, the input samples can include some few low magnitude negative values, so this prior will not be included. The concentrations are also non-negative. This positive constraint is introduced in the model with a proper simple constraint. If both of them were considered, decompositions such as non negative matrix factorization [13] could be appropriate. Finally, the supergaussianity of the sources is used in the approximation of the densities of the independent components.

3.3 A MAP Estimate

Ignoring the noise term in (2) the problem becomes:
$$\mathbf{x} = \mathbf{As} \tag{3}$$

For a given observation $\mathbf{x}$ of the optical depth in the ice mixtures at a given wavelength, the posterior probability of $\mathbf{A}$ using Bayes rule is:

$$p(\mathbf{A}/\mathbf{x}) = \frac{p(\mathbf{x}/\mathbf{A})p(\mathbf{A})}{p(\mathbf{x})} \tag{4}$$

where $p(\mathbf{A})$ is the prior probability of the mixing matrix and $p(\mathbf{x}/\mathbf{A})$ is:

$$p(\mathbf{x}/\mathbf{A}) = \left|\det \mathbf{A}^{-1}\right|\prod_{k=1}^{K} p_k(s_k) \tag{5}$$

due to (3) and to the independence of the K sources. To maximize (4), the term $p(\mathbf{x})$ is not relevant. Considering the posterior probability for the whole set of wavelengths $\mathbf{x}(j),\quad j = 1,...,P$, normalizing and taking logarithms, the maximum a posteriori estimate is obtained:

$$\mathbf{A}_{map} = \arg\max_{\mathbf{A}} \frac{1}{P}\sum_{j=1}^{P}\sum_{k=1}^{K}\log p_k(s_k(j)) - \log\left|\det \mathbf{A}\right| + \log p(\mathbf{A}) \tag{6}$$

This is the maximum likelihood estimate plus a term $\log p(\mathbf{A})$ that tries to penalize $\mathbf{A}$ entries with low probability in the prior knowledge. The demixing problem is:
$$\mathbf{y} = \mathbf{Bx} \tag{7}$$

where $\mathbf{B}$ is the recovering matrix and $\mathbf{y}$ the recovered signals. For the square case, $\mathbf{B} = \mathbf{A}^{-1}$, the equivalent function to (6) to be maximized is:

$$\mathbf{B}_{map} = \arg\max_{\mathbf{B}} \frac{1}{P}\sum_{j=1}^{P}\sum_{k=1}^{K}\log p_k(\mathbf{b}_k^T\mathbf{x}(j)) + \log\left|\det \mathbf{B}\right| + \log p(\mathbf{B}^{-1}) \tag{8}$$

At this point we introduce the prior of the mixing matrix $p(\mathbf{A})$. Because its entries must be non-negative, a simple constraint imposing that $p(a_{ij}) = 0$ for $a_{ij} < 0$ is enough. To maximize (8) we need to know the factorized pdf of the independent sources, but due to the supergaussianity of the sources, the derivative of the logarithm of the unknown pdf's can be substituted by other non linear functions. The stochastic version of the algorithm that maximizes (8) is:

$$\mathbf{B}_{new} = \mathbf{B}_{old} + \alpha(\mathbf{I} + \mathbf{f}(\mathbf{y})\mathbf{y}^T)\mathbf{B}_{old} \tag{9}$$

subject to $\mathbf{B}^{-1}$ does not have negative entries; α is the usual learning rate and $\mathbf{f}$ the non-linear component-wise function; i.e., the same updating rule as in the infomax algorithm [14] with the prior restriction. Note that (9) corresponds to the natural gradient version of the algorithm. The algorithm can be understood as finding $\mathbf{B}$ such that the distribution of $\mathbf{y}$ in (7) is as close as possible in a Kullback-Leibler divergence sense to the supposed distribution of the sources [15] without violating the constraint; e.g., if $f(y) = -2\tanh(y)$, then $\log p_i(s_i) = \beta - 2\log\cosh(s_i)$ with β a parameter fixed to make this the logarithm of a pdf. As we mentioned before, usually there will be more ices than molecules, so a first step of prewhitening will be carried out to reduce the dimension of the problem. This step also accelerates the convergence of the algorithm and obviously the same transformation must be applied to the prior.

4 Results

We tested the algorithm with the ice analogs database of Leiden [16]. This database contains the infrared spectra of laboratory analogs of interstellar ices. Different mixtures of molecules (from one up to three components selected from H_2, H_2O, NH_3, CH_4, CO, H_2CO, CH_3OH, O_2, N_2 and CO_2) at different temperatures and UV doses were produced, being the final spectrum calculated ratioing the measured and the background spectrum. The units of the data are cm^{-1} and absorbance, which is defined as optical
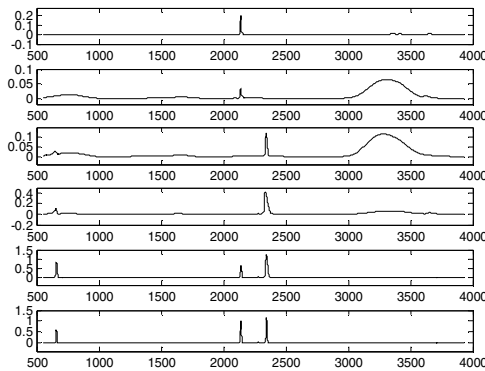
Fig. 3. Ice mixtures. Top to bottom, H_2O+CO (10:100), H_2O+CO (100:33), H_2O+CO_2 (100:14), H_2O+CO_2 (100:125), CO+CO_2 (100:70), CO+CO_2 (100:23).
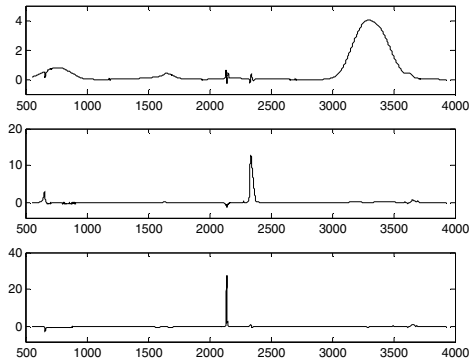
Fig. 4. Recovered Sources. Top to bottom, H2O, CO2 and CO.

length/ln10. The baseline was removed with Origin software and the useful wavelengths intervals were selected.

In Fig. 3 we show the ice mixtures, in this case ices containing H_2O, CO and CO_2. The three estimated components are shown in Fig. 4, where we can observe a CO and a CO_2 residual in the H_2O component due to the sharp peak value identifying the two components around 2150 cm^{-1} and 2350 cm^{-1} respectively.

5   Conclusions

In this paper we have presented the application of ICA to astrophysical ices. Mixture signals are the infrared absorption spectra of ices, being the sources the spectra of the different compounds present in the ices and the entries their concentrations. The study and decomposition of the components of these astrophysical ices is necessary in future studies where, depending on the composition and environmental conditions, each ice can produce a different complex compound. An additional restriction is imposed in order to obtain a constrained MAP estimate: the entries of the mixing matrix are non-negative. The effectiveness of ICA to extract the molecules present in a collection of laboratory ices has been proved.

The ICA approach has other advantages with respect to classical supervised methods: we do not need a complete library of endmembers for different situations, it is a multichannel method, concentration and endmembers can be estimated at the same time, and more challenging, it can extract signals corresponding to artifacts and unknown compounds not included in the library. Future work will try to model the dependencies between components for some molecules and to exploit the wavelength structure of the spectra.

Acknowledgements

References

1. Van Dishoeck, E.F., Blake, G.A., Draine, B.T., Lunine, J.I.: Protostars and Planets III, University of Arizona Press. Tucson, (1993) 163
2. Miller, S.L., Ann. N.Y. Acad. Sci., 69, (1957) 123
3. Strazzulla, G., Brucato, J.R., Palumbo, M.E. Satorre, M.A.: Ion implantation in ices, NIMB, 116, (1996) 289-293
4. Gerakines P., Schutte W., Greenberg J., van Dishoeck E.: The infrared band strengths of H_2O, CO and CO_2 in laboratory simulations of astrophysical ice mixtures. Astronomy and astrophysics, (1997)
5. Pendleton Y. J.: The nature and evolution of interstellar organics. ASP Conference Series, From stardust to planetesimals, Vol. 122, (1997)
6. Ehrenfreund, P., Boogert, A., Gerakines, P., Tielens, A., van Dishoeck, E.: infrared spectroscopy of interstellar apolar ice analogs. Astronomy and astrophysics, (1997)
7. Nuzillard D., Bijaoui A.:Blind source separation and analysis of multispectral astronomical images. Astronomy and astrophysics suplement series (2000) 129-138
8. Maino D. Farusi, A. Baccigalupi, C. Perrotta, F. Banday, A. J. Bedini, L. Burigana, C. De Zotti, G. Górski, K. M. and Salerno, E.. All-sky astrophysical component separation with (FastICA). Monthly Notices of the Royal Astronomical Society, (2001)
9. Forni O., Poulet F., Bibring J.P., Erard S. Component separation of OMEGA spectra with ICA. Lunar and planetary Science XXXVI, (2005)
10. Parra L., Mueller K.R., Spence C., Ziehe A., Sajda P.: Unmixing Hyperspectral Data, Advances in Neural Information Processing Systems 12, MIT Press, , (2000) 942-948
11. Bayliss J., Gualtieri J., Cromp R.: Analysing hyperspectral data with ICA. Proc. SPIE Applied Image and Pattern Recognition Wrkshp. (1997)
12. Bandfeld, J., Christensen P.: spectral data set factor analysis and end-member recovery: application to analysis of Martian atmospheric particulates. Journal of Geophysical Research, Vol. 105, No. E4, (2000) 9573-9587
13. Lee D.D., Seung, S.: Learning the parts of objects by non-negative matrix factorization. Nature, Vol. 401, (1999) 788-791
14. Bell A.J., Sejnowsky T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7 (1995) 1129-1159
15. Cardoso, J.F.: Higher-order contrasts for ICA, Neural Computation 11(1999) 157-192
16. http://www.strw.leidenuniv.nl/~schutte/database.html